Advances in formal Slavic linguistics 2022

Edited by

Berit Gehrke

Denisa Lenertová
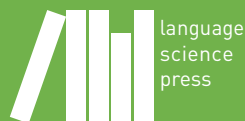
Roland Meyer

Daria Seres

Luka Szucsich

Joanna Zaleska

language
science
press

Open Slavic Linguistics

Editors: Berit Gehrke, Denisa Lenertová, Roland Meyer, Radek Šimík, Luka Szucsich & Joanna Zaleska

In this series:

1. Lenertová, Denisa, Roland Meyer, Radek Šimík & Luka Szucsich (Eds.). Advances in formal Slavic linguistics 2016.

2. Wiland, Bartosz. The spell-out algorithm and lexicalization patterns: Slavic verbs and complementizers.

3. Marušič, Franc, Petra Mišmaš & Rok Žaucer (eds.). Advances in formal Slavic linguistics 2017.

4. Blümel, Andreas, Jovana Gajić, Ljudmila Geist, Uwe Junghanns & Hagen Pitsch (eds.). Advances in formal Slavic linguistics 2018.

5. Dočekal, Mojmír & Marcin Wągiel (eds.). Formal approaches to number in Slavic and beyond.

6. Wągiel, Marcin. Subatomic quantification.

7. Kolaković, Zrinka, Edyta Jurkiewicz-Rohrbacher, Björn Hansen, Dušica Filipović Đurđević & Nataša Fritz. Clitics in the wild: Empirical studies on the microvariation of the pronominal, reflexive and verbal clitics in Bosnian, Croatian and Serbian.

8. Biskup, Petr, Marcel Börner, Olav Mueller-Reichau & Iuliia Shcherbina (eds.). Advances in formal Slavic linguistics 2021.

9. Jędrzejowski, Łukasz, Uwe Junghanns, Kerstin Schwabe & Carla Umbach (eds.). Syntax, semantics, and the lexicon: Papers by and in honor of Ilse Zimmermann.

10. Gehrke, Berit, Denisa Lenertová, Roland Meyer, Daria Seres, Luka Szucsich & Joanna Zaleska (eds.). Advances in Formal Slavic Linguistics 2022.

Advances in formal Slavic linguistics 2022

Edited by

Berit Gehrke

Denisa Lenertová

Roland Meyer

Daria Seres

Luka Szucsich

Joanna Zaleska

FREIE
UNIVERSITÄT
BERLIN

Contents

Contents

Preface

Advances in Formal Slavic Linguistics 2022 brings together a collection of 22 articles presenting the latest developments in formal approaches to Slavic linguistics. Covering a broad spectrum of topics across all branches of linguistics, the contributions explore various phenomena in BCMS, Slovenian, Russian, Ukrainian, Polish, Czech, Upper Sorbian, Bulgarian, and Macedonian, along with their regional dialects.

The authors use a wide range of methodologies, including fieldwork, acoustic analysis, corpus studies, psycholinguistic experiments, judgment studies, and theoretical modelling. The studies examine topics such as clitics, nominalizations, l-participles, the dual, verbal prefixes, assibilation, verbal and adjectival morphology, lexical stress, vowel reduction, focus particles, aspect, multiple wh-fronting, definiteness, polar questions, negation words, and argument structure. Analytical frameworks include Metrical Phonology, Distributed Morphology, Nanosyntax, Minimalist syntax, and formal semantics.

Early versions of these articles were presented at the conference on Formal Description of Slavic Languages 15, held in Berlin on 5–7 October 2022, with a special session on formal approaches to Ukrainian. Each article underwent a thorough, double-blind peer review process. We would like to extend our sincere thanks to the reviewers, as well as to the community proofreaders. We are also grateful to Mihaela Chirpanlieva for her assistance with Bulgarian transliteration and to the Language Science Press editorial team, Sebastian Nordhoff and Felix Kopecky, as well as Radek Šimík, for their invaluable help in producing this volume.

We hope this book provides a valuable resource for linguists interested in the complexities of Slavic languages and their implications for linguistic theory.

Berlin, May 22, 2025

Berit Gehrke
Denisa Lenertová
Roland Meyer
Daria Seres
Luka Szucsich
Joanna Zaleska

Chapter 1

From scope freezing to, well, everything: Investigations into the syntax of Instrumentals in Ukrainian

Svitlana Antonyuk

University of Graz

I present arguments in favor of a particular view of argument hierarchies in alternations that results from taking a deductive approach to scope freezing (Antonyuk 2015, 2020, 2022, submitted) and especially from relying on the Scope Freezing Diagnostic (SFD) as a source of insight in this domain. The primary focus of the paper is on several causative constructions in Ukrainian that feature an Instrumental case-marked argument. It is argued that in all constructions under consideration Instrumentals are merged in a structurally more prominent position than either the Accusative or the Dative argument, resulting in the V Instrumental >> Accusative >> Dative relative argument ordering at Merge. Next, I follow Lavine (2022, 2023) in arguing against the Undifferentiated Initiator view (Ramchand 2008; Bruening 2013; Legate 2014; Wood 2017, i.a.) and provide a new argument in support of Lavine's structural differentiation between the Instrumental Agents of passives and Instrumental Inanimate Causers.

1 Introduction

While research into argument structure alternations (ASAs) has made great advances, yielding important results over more than six decades, it has also been characterized by the vastly different, indeed, often diametrically opposing conclusions that have been reached based on largely the same types of evidence. Limiting our attention to the Slavic languages for the purposes of this discussion, we notice that the debate around the Ditransitive Alternation (exemplified in (1) and

(2) for English and Ukrainian) has witnessed accounts that posit a derivational relation between the two frames as well as those that posit independent projection of the two structures. Within these two large groups, further differences are observed based on the posited Merge position of the internal arguments, with accounts differing on, for example, whether the Dative argument is merged in the Specifier of V (Greenberg & Franks 1991, Franks 1995, Richardson 2007), the complement of V (Bailyn 1995, 2010, 2012; Antonyuk 2015, 2020; Titov 2017), the Specifier of an Applicative head (Dyakonova 2009), as well as accounts that posit Datives base-merged at different heights (Boneh & Nash 2017).

(1) The Ditransitive Alternation
  a. Mike gave a toy to his cat.   Prepositional Dative (PPD)
  b. Mike gave his cat a toy.   Double Object Construction (DOC)

(2) a. Myxajlo   po-daruvav (jakus') igrašku svojij kišci.   PPD
        Myxajlo.ɴᴏᴍ ᴘᴏ-gift.ᴘsᴛ   some   toy.ᴀᴄᴄ his   cat.ᴅᴀᴛ
        'Myxajlo gifted (some)/a toy to his cat.'
     b. Myxajlo   po-daruvav svojij kišci   (jakus') igrašku.   DOC
        Myxajlo.ɴᴏᴍ ᴘᴏ-gift.ᴘsᴛ his   cat.ᴅᴀᴛ some   toy.ᴀᴄᴄ
        'Myxajlo gifted his cat (some)/a toy.'   (Ukrainian)

A rather underinvestigated property of the Ditransitive Alternation, familiar since Larson (1990), is the "frozen" surface scope relation between the internal arguments in the DOC, which contrasts with scope fluidity/scope ambiguity of the PPD frame, cf. (3a)–(3b).

(3) a. Mike gave some toy to every cat.
        Surface scope: For some toy x, for every cat y, Mike gave x to y (e.g., this week).
        Inverse Scope: For every cat x, for some toy y, Mike gave x y (different cat–toy pairings possible).
     b. Mike gave some cat every toy.   (frozen surface scope)
        Surface scope: For some cat x, for every toy y, x received y from Mike (i.e., 'one happy cat' scenario).
        Unavailable inverse scope: For every toy x, for some cat y, x was given to y (different toy-cat pairings possible).

As noted in Larson (1990), Schneider-Zioga (1988) observes the arguably identical pattern of scope freezing in the 'with'-variant of the *Spray-Load* Alternation

(4), thus the scope ambiguity-scope freezing patterns found in the two alterna-tions have come to be known as their characteristic, albeit fairly understudied, property.[1]

(4)   The *Spray-Load* Alternation

    a. Mike planted the flowers in the garden.   locative frame

    b. Mike planted the garden with flowers.   the 'with'-frame

(5)   a. Mike planted some sort of flowers in every garden.
      Surface scope: For some sort of flowers x, for every garden y, Mike planted x in y.
      Inverse scope: For every garden x, for some sort of flowers y, Mike planted x with y.

    b. Mike planted some garden with every sort of flowers.

                                  (frozen surface scope)

      Surface scope: For some garden x, for every sort of flowers y, Mike planted x with y.
      Unavailable inverse scope: For every sort of flowers x, for some garden y, Mike planted x in y.

While the above scope patterns and the scope freezing in particular have come into focus in Bruening (2001), the phenomenon has effectively been explained away: Bruening takes scope freezing to be epiphenomenal, that is, a mere conse-quence of the purportedly Superiority-obeying nature of the covert movement operation Quantifier Raising (QR), which, combined with Richards's (1997) Tuck-ing in, is taken to result in situations where the structurally more prominent QP takes obligatory wide scope upon QR due to the lower one obligatorily tucking in and thus scoping below it.[2] Assuming the treatment in Antonyuk (2015, 2020, 2022, submitted), where scope freezing is viewed as an empirical phenomenon in its own right and thus in need of a principled theoretical explanation, the present paper will take without justification the need for the deductive approach

---

[1]Scope freezing is the only property of the ditransitive alternation (DA) where the mirror-image behavior of the two frames breaks down with respect to the Barss & Lasnik's (1986) diagnostics. Yet, as noted in Harley & Miyagawa (2017), it continues to be underinvestigated and poorly understood.

[2]Bruening's (2001) treatment of scope freezing is critically evaluated and ultimately rejected in Larson et al. (2019) for English and Antonyuk & Mykhaylyk (2022) for Ukrainian. See Abels & Grabska (2022) for a recent account that adopts (a modified version of) Bruening (2001) to model scope relations in Polish ditransitives and Hallman (2018, 2024) for the causative ditransitive alternation in Syrian Arabic.

to scope freezing underlying the above works and will proceed to rely on the SCOPE FREEZING DIAGNOSTIC (Antonyuk 2015) yielded by this approach in order to probe underlying argument structure relations.[3] As will hopefully become apparent, this approach to scope freezing and especially the diagnostic contributed by it is highly valuable as far as diagnostics developed for Slavic languages go, able to provide non-trivial insights into the syntax of argument structure alternations and remaining remarkably consistent where other diagnostics fail to be so. Among the findings reported here is that across a number of causative constructions involving an Instrumental NP, the latter is consistently implicated by the Scope Freezing Diagnostic (SFD) to be merged in a structurally superior position relative to its VP-internal co-argument(s). Furthermore, I show that an Instrumental Inanimate Causer NP is distinguished from the (morphologically indistinguishable) Instrumental Agent NP in passives in terms of scope behavior, which supports Lavine's (2022, 2023) broader argument for the need to distinguish between agent/animate causers, which are arguments of Voice, and a variety of inanimate causers, argued to merge lower in the structure.

The remainder of the paper is structured as follows. Section 2 presents empirical data from Ukrainian that demonstrate the Scope Freezing Generalization and introduces the Scope Freezing Diagnostic that is based on it. Section 3 is concerned with mapping out the verb phrase by applying the SFD to a variety of constructions involving Accusative, Dative and Instrumental arguments. It shows, i.a., that the relative ordering of internal arguments at Merge differs significantly from that often considered "standard" in the field (Section 3.1). The rest of Section 3 then goes deeper into the syntax of Instrumentals by contrasting their scope behavior with that of the Instrumentally case-marked subjects of passives, providing new evidence against the Undifferentiated Initiator view and thus offering support to Lavine's (2022, 2023) position (Section 3.2.1) according to which Inanimate Causer Instrumentals are merged lower than Agents. Section 3.2.2 argues that the Merge position of Inanimate Causer Instrumentals is

---

[3]An anonymous reviewer challenges the terminology, pressing about what makes this approach deductive rather than inductive. Of course, the reviewer is correct to press on it, as the present approach does in fact combine elements of both. Inductively, this approach follows the data to form generalizations to form broader theory of scope freezing. The approach is however deductive in that, relying on the general minimalist set of assumptions and the newly formed theory (or, barring that, a set of strong generalizations about scope freezing), it allows us to test a range of properties under discussion, from properties of scope freezing itself, to argument hierarchy in argument structure alternations and syntactic properties of particular constructions to, e.g., information structural phenomena manifested in such constructions, such as specificity, discourse neutrality, etc. (see esp. Antonyuk 2022). To the extent that any such insights are derived from the theory of scope freezing advocated for here, they are all derived by deductive reasoning. I am grateful to the reviewer for the chance to clarify this point.

at the same time much higher than usually assumed, specifically, Spec,*v*Cause (rather than the often assumed Complement of V position). Section 4 offers my conclusions.

2 Scope freezing in Ukrainian: The empirical domain

While taking such an approach to scope freezing (i.e., one deducing what scope freezing is from patterns of alternating word order-QP scope distribution) makes good scientific sense in general, it is arguably especially important as far as Slavic languages are concerned, for, as shown in Antonyuk (2015) and elaborated in Antonyuk (2022, submitted), unlike English, where there appear to be only two constructions that exhibit frozen surface scope, the phenomenon of frozen surface scope is not only found in (East) Slavic as well, but it is found in a considerably broader range of constructions.[4] What all these constructions have in common is the pattern, schematized in (6), whereby one linearization of internal arguments is scopally ambiguous while the opposite order of arguments in the postverbal field is surface scope frozen, allowing only the QP scope interpretation that corresponds to overt c-command relations.

(6)   a. $[_{TP}$ EA $[$ <EA> V+v $[_{VP}$ QP2 <V> QP1 $]]]$ scope ambiguous
      b. $[_{TP}$ EA $[$ <EA> V+v $[_{XP}$ QP1 <V> $[_{VP}$ QP2 <V> QP1 $]]]]$
                                                                 scope frozen*,[5]

           *where QP1 ≠ PP$_{\text{LOCATIVE/DIRECTIONAL}}$

According to (6), the conditions under which scope freezing arises are quite specific and limited, that is, scope freezing arises whenever the structurally lower internal argument raises overtly across its structurally superordinate co-argument to a c-commanding position in the postverbal field. The word order alternations below provide some of the empirical basis for (6).[6]

---

[4]The phenomena discussed here are observed in many Slavic languages across the subgroups. See Antonyuk (2015, 2020, 2022) and Boneh & Nash (2017) on Russian. See Antonyuk & Mykhaylyk (2022) on the interaction of quantification and Object Shift in Ukrainian. See also Marvin & Stegovec (2012) for a brief discussion of scope freezing in Slovenian DOCs and Abels & Grabska (2022) for a detailed discussion of scope distribution in Polish ditransitives.

[5]When QP1 = PP$_{\text{LOCATIVE/DIRECTIONAL}}$ the result of argument permutation is surface scope bias, not surface scope freezing. See Antonyuk (2020, submitted) for a relevant discussion.

[6]See Dyakonova (2009) and Bailyn (2010, 2012) on Russian ditransitives and the discussion of the relation between the two linearizations. Both authors agree they instantiate the two frames of the Ditransitive Alternation.

(7) The Ditransitive Alternation

    a.   Myxajlo   po-daruvav jakus'   igrašku kožnij   kišci.
       Myxajlo.NOM PO-gift.PST some.ACC toy.ACC every.DAT cat.DAT
       'Myxajlo gifted some toy to every cat.'
       Surface scope: For some toy x, for every cat y, Myxajlo gifted x to y
       (i.e., the same toy).
       Inverse scope: For every cat x, for some toy y, Myxajlo gifted x y (i.e.,
       different cat–toy pairs).

    b.   Myxajlo   po-daruvav jakijs'   kišci   kožnu   igrašku.
       Myxajlo.NOM PO-gift.PST some.DAT cat.DAT every.ACC toy.ACC
       'Myxajlo gifted some cat every toy.'
       Surface scope: For some cat x, for every toy y, Myxajlo gifted x y.
       Unavailable inverse scope: For every toy x, for some cat y, Myxajlo
       gifted x to y.

(8) The *Spray-Load* Alternation

    a.   Myxajlo   za-lyv   [jakyjs'   vyd   pal'noho] [v
       Myxajlo.NOM za-fill.PST some.ACC type.ACC gas.GEN   into
       kožen   bak].
       every.ACC tank.ACC
       'Myxajlo filled some type of gas into every tank.'
       Surface scope: For some type of gas x, for every tank y, Myxajlo filled
       x into y.
       Inverse scope: For every tank x, for some type of gas y, Myxajlo filled
       x with y (i.e., possibly different type of gas for each tank).

    b.   Myxajlo   za-lyv   [jakyjs'   bak]   [kožnym vydom
       Myxajlo.NOM za-fill.PST some.ACC tank.ACC   every.INS type.INS
       pal'noho].
       gas.GEN
       'Myxajlo filled some tank with every type of gas.'
       Surface scope: For some tank x, for every type of gas y, Myxajlo filled
       x with y.
       Unavailable inverse scope: For every type of gas x, for some tank y,
       Myxajlo filled x into y.

(9)   Reflexive Causatives

   a. Likar   infikuvav-sja jakojus' xvoroboju vid kožnoho
      doctor.NOM infect.PST-REFL some.INS illness.INS from every.GEN
      pacijenta.
      patient.GEN
      'The doctor got infected from every patient with some illness.'
      Surface scope: For some illness x, for every patient y, the doctor got
      infected with x from y.
      Inverse scope: For every patient x, for some illness y, the doctor got
      inflected by x with y.

   b. Likar   infikuvav-sja vid jakohos' pacijenta   kožnoju
      doctor.NOM infect.PST-REFL from some.GEN patient.GEN every.INS
      xvoroboju.
      illness.INS
      'The doctor got infected with some illness from every patient.'
      Surface scope: For some patient x, for every illness y, the doctor got
      infected by x with y.
      Unavailable inverse scope: For every illness x, for some patient y, the
      doctor got infected with x by y.

(10)   Causative verbs

   a. Likar   infikuvav jakojus' xvoroboju kožnoho pacijenta.
      doctor.NOM infect.PST some.INS illness.INS every.ACC patient.ACC
      'The doctor infected every patient with some illness.'
      Surface scope: For some illness x, for every patient y, the doctor
      infected y with x.
      Inverse scope: For every patient x, for some illness y, the doctor
      infected x with y.

   b. Likar   infikuvav jakohos' pacijenta   kožnoju xvoroboju.
      doctor.NOM infect.PST some.ACC patient.ACC every.INS illness.INS
      'The doctor infected some patient with every illness.'
      Surface scope: For some patient x, for every illness y, the doctor
      infected x with y.
      Unavailable inverse scope: For every illness x, for some patient y, the
      doctor infected y with x.

The key observation to note here is, of course, that the change in scope interpretations in each pair (i.e., the appearance of scope freezing in the (b) examples)

tracks the change in overt word order, which constitutes the foundation for the Scope Freezing Generalization in (11), according to which scope freezing is an inherently derivational, in fact, *a derivation-by-movement* phenomenon.[7]

(11) The Scope Freezing Generalization (SFG) (Antonyuk 2015)
   Scope freezing obtains when one QP raises overtly across another to a c-commanding´ position as a result of a single instance of movement within the *v*P/VP.

That scope freezing must be limited to the domain of *v*P/VP (i.e., below VoiceP in more modern terminology) becomes obvious once the above examples are compared to those involving a subject QP: Here we observe that the interaction between an object QP and a subject QP never amounts to a frozen scope relation.[8]

(12) Simple SVO transitives: No scope freezing
   Jakas'   divčynka na-hoduvala   kožnu   kišku.
   some.NOM girl.NOM NA-feed.PST.F every.ACC cat.ACC
   'Some girl fed every cat.'
   Surface scope: For some girl *x*, for every cat *y*, *x* fed *y*.
   Inverse scope: For every cat *x*, for some girl *y*, *x* was fed by *y*.

(13) OSV (locally scrambled) clauses: No scope freezing
   Jakus'   kišku   kožna   divčynka na-hoduvala.
   some.ACC cat.ACC every.NOM.F girl.NOM.F NA-feed.PST.F
   'Some cat, every girl fed.'
   Surface scope: For some cat *x*, for every girl *y*, *x* was fed by *y*.
   Inverse scope: For every girl *x*, for some cat *y*, *x* fed *y*.

(14) OVS clauses: No scope freezing
   Jakus'   kišku   na-hoduvala   kožna   divčynka.
   some.ACC cat.ACC NA-feed.PST.F every.NOM.F girl.NOM.F
   'Some cat was fed by every girl.'
   Surface scope: For some cat *x*, for every girl *y*, *x* was fed by *y*.
   Inverse scope: For every girl *x*, for some cat *y*, *x* fed *y*.

---

[7]The importance of this conclusion should be immediately clear: It implicates a much greater degree of derivationality inherent in the derivation of numerous structures and argument structure alternations than is currently assumed.

[8]Building on the original observation in Larson (1990), Bruening (2001) provides independent evidence in favor of scope freezing being limited to object QPs, never an object and a subject QP. Thus, conclusions based on the richer East Slavic empirical data are fully aligned with the conclusions based on the English data. As argued especially in Antonyuk (submitted), since the scope freezing found in Ukrainian and Russian is arguably identical to the English case, the conclusions reached on the basis of the former should be viewed as naturally extending to the latter.

While the absence of scope freezing in OVS is important, showing that object QP–subject QP pairs do not result in scope freezing under any circumstances, much will depend on one's analysis of OVS, which is a hotly contested issue in Slavic linguistics.[9] Whatever derivation of OVS one may assume, however, the basic empirical observation is that subject QP >> object QP pairs as well as object QP >> subject QP pairs do not result in scope freezing. Coupled with the observation that direct objects readily participate in frozen surface scope configurations, the conclusion this invites is that VoiceP, the layer that on most assumptions introduces the external argument, is outside the domain within which scope freezing can be established. On the assumption that OVS clauses are derived by overtly raising the object phrase to a position preceding and c-commanding the subject, the SVO–OVS examples are particularly telling in this respect when compared with, e.g., the ditransitive alternation in (7) or, e.g., (10), where the two orders are arguably also derived by an overt instance of movement raising the lower object to a position preceding and c-commanding the higher one. Whereas in the former cases the overt instance of movement still yields scope ambiguity, in the latter case what results in surface scope freezing.

Another argument in favor of the domain of scope freezing to exclude the subject layer comes from nominalizations (see esp. Alexiadou 2009, 2017; Borer 2014; see also Chomsky 1970; Grimshaw 1990; Harley 2009; Marantz 1997, i.a.). Specifically, the nominalizations in (15) and (16), which exclude the external argument, still preserve the argument structure of the corresponding verbal layer of full sentences, with the scope relations between the internal arguments being preserved as well. Thus, once again, we see scope freezing reappear in one of the two possible linearizations of internal arguments in the absence of (an overt) subject:

(15)   a.   za-lyv-annja   jakohos' vydu   pal'noho v   kožen
            ZA-pour-*annja*.NOM some.GEN type.GEN gas.GEN   into every.ACC
            bak
            tank.ACC
            'the pouring of some type of gas into every tank'   (cf. (8a))
            Surface scope: For some type of gas *x*, for every tank *y*, there is a filling of *x* into *y*.
            Inverse scope: For every tank *x*, for some type of gas *y*, there is a filling of *x* with *y* (i.e., possibly different type of gas for each tank).

---

[9]For the sake of clarity, I assume the analysis of OVS proposed in Antonyuk (2021).

     b. za-lyv-annja   jakohos' baku   kožnym vydom pal'noho
       ZA-pour-*annja*.NOM some.GEN tank.GEN every.INS type.INS gas.GEN
       'the pouring of some tank with every type of gas'   (cf. (8b))
       Surface scope: For some tank x, for every type of gas y, there is a
       filling of x with y.
       Unavailable inverse scope: For every type of gas x, for some tank y,
       there is a filling of x into y.

(16) a. infikuv-annja   jakojus' xvoroboju kožnoho   pacijenta
       infect-*annja*.NOM some.INS illness.INS every.ACC patient.ACC
       'infecting of every patient with some illness'   (cf. (10a))
       (Lit.: the infecting with some illness of every patient)
       Surface scope: For some illness x, for every patient y, there is infecting
       of y with x.
       Inverse scope: For every patient x, for some illness y, there is infecting
       of x with y.
     b. infikuv-annja   jakohos' pacijenta   kožnoju xvoroboju
       infect-*annja*.NOM some.ACC patient.ACC every.INS illness.INS
       'the infecting of some patient with every illness'   (cf. (10b))
       Surface scope: For some patient x, for every illness y, there is infecting
       of x with y.
       Unavailable inverse scope: For every illness x, for some patient y,
       there is infecting of y with x.

Of course, COMPLEX EVENT NOMINALS (CENs), such as the examples in (15) and (16), probably do not really "exclude" the subject layer, as the data above may superficially suggest: the subject is implicit, as evidenced by the ability of such nominalizations to be modified by agentive modifiers such as *navmysne* 'deliberate', as in *navmysne infikuvannja* 'deliberate infecting'. Thus, especially within DM, on various instantiations of the "Phrasal Layering" analysis of Complex Event Nominals, the verbal structure, including the layer that introduces the external argument, is included in nominalizations, with the nominalizer head little *n* attaching on top of fully projected verbal structure (Alexiadou 2001, 2017; Bruening 2013; Borer 2014; Iordăchioaia 2020; McGinnis 2020, i.a.).[10]

    This means that the examples in (15) and (16) do not really provide bullet-proof evidence that the subject is "excluded" is any real sense, merely that it appears to

---

[10]See also Wood (2023) for a detailed theoretical overview as well as an alternative proposal based on Icelandic data that CENs can be formed in syntax without nominalizing full verbal structure.

be inert and not participating in whatever processes result in the establishment of the surface scope freezing relation. Minimally, the account of scope freezing adopted here (spelled out in more detail in later sections) is committed to the following conclusion: the preservation of the scope freezing relation means that the layer of structure arguably (and crucially) implicated in scope freezing, ApplP (located between *v*P and VoiceP), must be included in the verbal structure nominalized by *n* (cf. Wood 2023).

To summarize, we see that the SFG points to a derivational nature of scope freezing, resulting from an overt instance of movement I will henceforth refer to as ARGUMENT INVERSION (following Antonyuk & Mykhaylyk 2022). Taking place in the postverbal field, Argument Inversion (AI) constitutes a local instance of overt syntactic movement, and, as examples such as the anaphor binding data below suggest, AI involves A-movement:[11]

(17)   a.   Dolja   po-daruvala nas$_i$ odyn odnomu$_i$.
           Fate.NOM PO-gift.PST   us.ACC each other.DAT
           'Fate gifted us to each other.'

       b.   Dolja   po-daruvala nam$_i$ odyn odnoho$_i$.
           fate.NOM PO-gift.PST   us.DAT each other.ACC
           'Fate gifted us each other.'

The anaphor binding data provides clear evidence that A-movement is involved. Considering A-movement leads to new binding relations, however, it can be argued that the directionality of the derivation in fact goes in the opposite direction. That is, rather than assuming that (17a) represents the base order and (17b) is derived from it by overt movement, it has been argued that (17b) represents the base order, and (17a) obtains from overtly raising the direct object across the Dative-marked anaphor, thus establishing a new binding relation.

While this type of argument has indeed been made many times, let me point out that it works much better with cases involving binding than it does with cases involving scope relations. This is so since the binding data in (17) is symmetric, i.e., both (17a) and (17b) can be explained on either type of account. With QP scope, things are very different. On the widely held assumption that the Dative argument in ditransitives is merged in a structurally higher position than the Accusative argument, the scope freezing familiar from the DAT >> ACC ditransitives must be seen as an inherent property of the construction or, e.g., on Bruening's (2001) Superiority account, as an entirely epiphenomenal property. As pointed out in Larson et al. (2019), Bruening's account thus predicts that scope

---

[11]The original examples are due to Asarina (2005), cited in Bailyn (2012).

freezing should be found in English in many more constructions, namely all of those where the base relation between the two QPs can be plausibly analyzed (as in Bruening 2001) as involving asymmetric c-command in the base structure, – contrary to fact. Since this prediction is thoroughly falsified, we are left with a conclusion that scope freezing is, perhaps, just an inherent property of certain constructions, at least, as far as English is concerned. That is, we are left with no way of predicting where else we might find scope freezing. And perhaps one could accept it on the grounds that this property is somehow "exceptional", i.e., found in these two ditransitive constructions only.

Drawing now on our insights from Ukrainian/East Slavic, it becomes clear that our inability to model scope freezing, to predict where else in the language we may find it, is, in fact, a problem, as the phenomenon is found in a significant number of constructions, thus posing a challenge for syntactic theory. On the view adopted here, on the other hand, not only do we have an extensive list of environments where scope freezing has already been found to obtain, the Scope Freezing Generalization allows us to predict potential further environments where scope can be expected to "freeze" in this way as well. While this result is significant, the SFG also provides us with a powerful diagnostic tool for probing argument structure relations and allows for other non-trivial insights into the derivation of the extended verbal domain in East Slavic. Thus, I conclude that the below schematization accurately describes the conditions on scope freezing for the language under discussion, Ukrainian:

(18) Surface scope freezing, schematized:
$[_{TP}$ ExtA $[_{VoiceP}$ <ExtA> V+v $[_{XP}$ QP1 $[_{VP}$ QP2 <V> <QP1>]]]]

In what follows then, I will be relying on the Scope Freezing Diagnostic (SFD), which allows us to probe relative argument structure relations at Merge (Antonyuk 2015):

(19) The Scope Freezing Diagnostic
Frozen surface scope implicates a derived structure resulting from Argument Inversion.

The methodological goal of this paper is to demonstrate the remarkable internal consistency of the diagnostic and the non-trivial insights into the syntax of argument structure alternations that can be gained by applying it. I will not attempt to demonstrate all the insights already derived from the application of the SFD to the East Slavic (Ukrainian and Russian) data for reasons of space, directing the reader to the original research papers where these results are presented. Here I

will only summarize and briefly exemplify some of the findings so we can build on them in this paper.[12]

3 Mapping out the verb phrase

3.1 The relative ordering of internal arguments at Merge

As should already be clear from our preliminary discussion, the SFD is a simple diagnostic, one which can (at most) point to a derived structure among the alternating frames under investigation (if such exists), as well as point to the relative ordering of arguments at Merge. The SFD cannot tell us much about the exact Merge position of arguments; what it does give us is a heuristic according to which the linearization that is surface-scope frozen must be the derived one, and moreover that it must be derived by raising the structurally lower of the two internal arguments overtly across the structurally higher one in the postverbal field, thus gaining c-command over it.[13,14] In other words, what the SFD gives us is a mere ordering of arguments at Merge. For the constructions discussed earlier, these relative ordering statements are as follows:

---

[12]See Antonyuk (2015, 2020, 2022, submitted) and Antonyuk & Mykhaylyk (2022) for the original data and findings. See Abels & Grabska (2022) for an experimental confirmation of the empirical claims in Antonyuk's work based on Polish QP scope data as well as for a critical engagement with the account of scope freezing assumed here.

[13]Antonyuk (submitted) argues that, according to the SFD (and contra Antonyuk 2015), the *Spray-Load* Alternation is, in fact, *not* a case of a derivational relation between the two alternating frames. Crucially, she argues that the 'with'-variant (i.e., the Instrumental case-marked frame in Slavic) is, nevertheless, derived, as suggested by the fact that it exhibits scope freezing, just not from the locative frame, but from the Instrumental frame via Argument Inversion of the Accusative argument across NP_{INS}. That is, scope freezing in the 'with'-variant is indeed a marker of its derived status, but what it is derived from is not what is traditionally viewed as its "alternating" frame. This, of course, is good news, as the two frames differ not only in the morphosyntactic marking on their two internal arguments, but also in their theta roles, making a derivational account problematic in this case.

[14]Significantly, argument structure alternations, in Slavic and elsewhere, appear to never exhibit scope freezing on both possible orders/alternating frames. On accounts which posit independent projection of the two alternating frames, this should in principle be a logical possibility. On Antonyuk's (2015) treatment of scope freezing, assumed here, such a situation is impossible in principle, for obvious reasons. The only conceivable exception to this would be if what were (mistakenly) considered to be an alternation consisted of two structures, each of which would be derivationally related to another structure that is not viewed as part of the alternation. The closest case to this hypothetical situation would in fact be the *Spray-Load* Alternation, where (as described in the previous footnote) the frozen frame (V $NP_{ACC} \gg NP_{INS}$) is derivationally related to another structure, (i.e., V $NP_{INS} \gg NP_{ACC}$), rather than to the locative frame. Apart from this scenario, I will venture a prediction that such a situation should be impossible.

Table 1: Relative ordering statements

	Frozen order	Base
The Ditransitive Alternation:	DAT >> ACC	ACC >> DAT
The *Spray-Load* Alternation:	ACC >> INS	INS >> ACC
Reflexive Causatives:	PP$_{FROM}$ >> INS	INS >> PP$_{FROM}$
Causative verbs:	ACC >> INS	INS >> ACC

Now, to some working on argument structure alternations and verbal argument structure more generally these insights may be surprising, as they go against a lot of what has arguably been assumed to be settled in the literature. My goal in this paper is not to provide conclusive evidence in favor of a particular structure for a particular alternation, but rather to demonstrate the general insights afforded by the SFD, especially in what concerns the relative Merge position of Instrumental arguments. A larger point, hinted at here and developed in detail elsewhere (Antonyuk 2025) is that, taken together, these insights suggest a rather interesting alternative view of how the derivation of the verb phrase in Discourse Configurational Slavic languages may proceed and what confounding factors have thus far prevented us from seeing this picture.

With respect to the Instrumentals specifically, the suggestion that they might be merged higher than the Accusative argument is certainly surprising, both because it is common to assume a low Merge/complement of V position for the Instrumental NP (as is standard for the Oblique case-marked arguments), as well as because the alternative would also suggest the Accusative NP must be merged low (in any case, lower than NP$_{INS}$). As already suggested in Table 1, the indication that the Instrumental NP is merged higher than either the Accusative-marked co-argument or the PP co-argument remains consistent even once we significantly expand the range of the constructions under investigation.

Consider the following data (which include examples presented earlier, for convenience):

(20) Causative verbs

  a. Likar   infikuvav jakojus' xvoroboju kožnoho pacijenta.
     doctor.NOM infect.PST some.INS illness.INS every.ACC patient.ACC
     'The doctor infected every patient with some illness.'
     Surface scope: For some illness x, for every patient y, the doctor infected y with x.
     Inverse scope: For every patient x, for some illness y, the doctor infected x with y.

b. Likar   infikuvav jakohos' pacijenta kožnoju xvoroboju.
doctor.NOM infect.PST some.ACC patient.ACC every.INS illness.INS
'The doctor infected some patient with every illness'
Surface scope: For some patient x, for every illness y, the doctor
infected x with y.
Unavailable inverse scope: For every illness x, for some patient y, the
doctor infected y with x.

⇒ BASE ORDER: V NP$_{\text{INS}}$ >> NP$_{\text{ACC}}$

According to the SFD, the V NP$_{\text{INS}}$ >> NP$_{\text{ACC}}$ linearization represents the relative
order of arguments at Merge. The same conclusion is reached for the 'with'-
variant of the *Spray-Load* Alternation:

(21)   Instrumental/'with'-frame of the Spray-Load Alternation

a. Myxailo   za-lyv   jakyjs' bak   kožnym vydom
Myxailo.NOM ZA-fill.PST some.ACC tank.ACC every.INS type.INS
pal'noho.
gas.GEN
Lit.: 'Myxailo filled some tank with every type of gas.'
Surface scope (∃>∀): for some tank x, for every type of gas y, Myxailo
filled x with y.
Unavailable inverse scope (∀>∃): for every type of gas x, for some tank
y, Myxailo filled x into y.

b. Myxailo   za-lyv   jakymos' vydom pal'noho kožen
Myxailo.NOM ZA-fill.PST some.INS type.INS gas.GEN every.ACC
bak.
tank.ACC
Lit.: 'Myxailo filled with some type of gas every tank.'
Surface scope (∃>∀): For some type of gas x, for every tank y, Myxailo
filled x into y.
Inverse scope (∀>∃): For every tank x, for some type of gas y, Myxailo
filled x with y.

⇒ NP$_{\text{INS}}$ >> NP$_{\text{ACC}}$ is the order at Merge within the Instrumental/'with'
frame.

The passive-like *-no/-to* construction, taken up in detail in Lavine (2022, 2023)
and illustrated in (22), is useful for our purposes in that it involves two internal
arguments, hence it can be subject to the SFD as well.

(22) The *-no*/*-to* construction (Lavine & Freidin 2002)

Cerkvu   bulo   spaleno   blyskavkoju.
church.ACC was.NON-AGR burned.down.NON-AGR lightning.INS
'The church was burned down by lightning.'

What we see is that, as before, the two arguments are permutable, and overtly permuting/crossing the arguments in the postverbal field results in scope freezing of the $NP_{ACC} \gg NP_{INS}$ order.[15]

(23) a. Jakus'   cerkvu   bulo   spaleno   kožnoju
          some.ACC church.ACC was.NON-AGR burned.down.NON-AGR every.INS

          blyskavkoju / z   blyskavok.
          lightning.INS from lightning.GEN.PL
          'Some church was burned down by every lightning.'
          Surface scope: For some church *x*, for every lightning *y*, *x* was burned down by *y*.
          Unavailable inverse scope: For every lightning *x*, for some church *y*, *x* burned down *y*.

      b. Jakojus' blyskavkoju bulo   spaleno
          some.INS lightning.INS was.NON-AGR burned.down.NON-AGR

          kožnu   cerkvu.
          every.ACC church.ACC
          'Some church was burned down by every lightning.'
          Surface scope: For some lightning *x*, for every church *y*, *x* burned down *y*.
          Inverse scope: For every church *x*, for some lightning *y*, *x* was burned down by *y*.

      ⇒ BASE ORDER: V $NP_{INS} \gg NP_{ACC}$

Something worth pointing out is that the surface scope interpretation for (23a) is infelicitous in that it entails a situation where one and the same church was burned down by every one of the lightning strikes under consideration. Thus, if there were five lightning strikes during the night, every single one must have

---

[15]I assume that the surface word order in such structures is derived by ultimately raising either of the arguments to Spec,TP. In the scopally frozen sentence, Argument Inversion first raises the lower argument, NP_{ACC}, to a position to the left of its co-argument, NP_{INS}, still in the postverbal field, which results in scope freezing. The thus inverted NP_{ACC} then undergoes raising into Spec,TP. See Antonyuk & Mykhaylyk (2022) for evidence that scope freezing, once established by AI in the postverbal field, cannot be disturbed/"unfrozen" by further syntactic movement.

hit the church and contributed to its burning down for the sentence to be true
on its surface scope interpretation. Yet, despite its real-world incongruent sur-
face scope interpretation, which should have really facilitated the inverse scope
reading, if it were available, to come through, the latter is nevertheless excluded
for this sentence. This helps demonstrate one of the core insights about scope
freezing: it is an all-or-nothing, categorical phenomenon, as opposed to surface
scope bias, familiar from other contexts, which can be manipulated by the choice
of lexical items, syntactic contexts, information-structural properties of the sen-
tence, etc. This property is what arguably makes the Scope Freezing Diagnostic
an incredibly reliable, internally consistent diagnostic (see esp. Antonyuk 2022).

   The exact same conclusion about the higher Merge position of the Instrumen-
tal argument (relative to the Accusative) can be reached by applying the SFD to
the so-called NON-AGREEING ACCUSATIVES (Lavine & Freidin 2002; Lavine 2022,
2023, i.a.). For clarity, the derivational path for the derived ((24a) and (25a)) is
assumed to be largely identical to that described for the *-no/-to* constructions in
footnote 15:

(24) a. Soldata   po-ranylo   kuleju.
          soldier.ACC PO-wound.NON-AGR bullet.INS
          'A soldier was wounded by a bullet.'

                                    (modeled on Lavine & Freidin 2002)

      b. Kuleju   po-ranylo   soldata.
          bullet.INS PO-wound.NON-AGR soldier.ACC
          'A soldier was wounded by a bullet.'

(25) a. Jakohos' soldata   po-ranylo   kožnoju kuleju.
          some.ACC soldier.ACC PO-wound.NON-AGR every.INS bullet.INS
          'Some soldier was wounded by every bullet.'
          Surface scope: For some soldier x, for every bullet y, x was wounded
          with y.
          Unavailable inverse scope: For every bullet x, for some soldier y, x
          wounded y.

      b. Jakojus' kuleju   po-ranylo   kožnoho soldata.
          some.INS bullet.INS PO-wound.NON-AGR every.ACC soldier.ACC
          'Every soldier was wounded by a bullet.'
          Surface scope: For some bullet x, for every soldier y, x wounded y.
          Inverse scope: For every soldier x, for some bullet y, x was wounded
          with y.

   ⇒ BASE ORDER: V NP$_{\text{INS}}$ >> NP$_{\text{ACC}}$

The non-agreeing accusatives involve two internal arguments, marked for Dative and Accusative case, allowing us to test whether the relative ordering V NP$_{ACC}$ >> NP$_{DAT}$ obtained for the DOC will be replicated. As shown in (27), this is indeed the case:

(26) a. Xlopcevi vidrizalo   palec'   na ruci.
           boy.DAT severed.NON-AGR finger.ACC on hand
           'The boy's finger was severed.'

      b. Palec'   na ruci vidrizalo   xlopcevi.
           finger.ACC on hand severed.NON-AGR boy.DAT
           'A finger on hand was severed from a guy's hand.'

(27) a. Jakomus' xlopcevi vidrizalo   kožen   palec'   (na ruci).
           some.DAT guy.DAT severed.NON-AGR every.ACC finger.ACC on hand
           'Some boy got every one of his fingers severed.'
           Surface scope: For some boy *x*, for every finger *y*, *x* had *y* severed from *x*'s hand.
           Unavailable inverse scope: For every finger *x*, for some boy *y*, *x* was severed from *y*'s hand.

      b. Jakyjs'   palec'   (na ruci) vidrizalo   kožnomu xlopcevi.
           some.ACC finger.ACC on hand severed.NON-AGR every.DAT boy.DAT
           'Some finger was severed from every boy's hand.'
           Unavailable surface scope: For some finger *x*, for every boy *y*, *x* was severed from *y*'s hand.
           Inverse scope: For every boy *x*, for some finger *y*, *x*'s *y* was severed.
     ⇒ BASE ORDER: V NP$_{ACC}$ >> NP$_{DAT}$

Now, it is interesting that up to now we have seen two of the relevant "Merge orderings", that is, V NP$_{ACC}$ >> NP$_{DAT}$ and V NP$_{INS}$ >> NP$_{ACC,}$ established and replicated over several types of constructions here and elsewhere. By transitivity, NP$_{INS}$ should precede NP$_{DAT}$ as well: V NP$_{INS}$ >> NP$_{DAT}$. Since the example in (26) can be spelled out more fully, to include an implicit Instrumental Inanimate Causer argument, this new relative ordering, due to transitivity, can be put to the test by applying the SFD to the Instrumental and the Dative-marked quantificational arguments. Doing so confirms that V NP$_{INS}$ >> NP$_{DAT}$ is, indeed, the correct base structure ordering:

(28) Xlopcevi vidrizalo   palec'   elektryčnoju pyloju.
      boy.DAT severed.NON-AGR finger.ACC electric.INS   saw.INS
      'The boy's finger got severed by an electric saw.'

(29) a. Jakomus' xlopcevi vidrizalo   palec'   kožnym
          some.DAT guy.DAT severed.NON-AGR finger.ACC every.INS
          instrumentom.
          instrument.INS
          'Some boy got a finger severed by every instrument.'   (frozen)

      b. Jakymos' instrumentom vidrizalo   palec'   kožnomu
          some.INS instrument.INS severed.NON-AGR finger.ACC every.DAT
          xlopcevi.
          boy.DAT
          'With some instrument (or other), every boy's finger was severed.'
                                                                (ambiguous)

      ⇒ BASE ORDER: V NP_{INS} >> NP_{DAT}

We have thus arrived at the relative "Merge ordering" for 3-argument NPs, summarized in (30). The SFD makes abundantly clear that the NP_{INS} argument cannot be equated with other obliques in occupying the lowermost, complement position (cf. Pesetsky 1995, i.a.). In fact, according to the SFD, NP_{INS} is merged the highest of the three internal arguments.

(30) V NP_{ACC} >> NP_{DAT}
      V NP_{INS} >> NP_{ACC}
      V NP_{INS} >> NP_{DAT}

Thus, whatever the actual Merge positions (something the SFD does not and cannot provide an answer to), the relative structure representation we arrive at looks like the following:

(31) V NP_{INS} >> NP_{ACC} >> NP_{DAT}

3.2 The syntax of Instrumentals in Ukrainian

While the above relative ordering is all the SFD can give us, it is plenty, of course. Supplementing with other types of evidence (see Antonyuk & Mykhaylyk 2022; Antonyuk submitted for arguments based on Ukrainian data; see also Bailyn 2010, 2012; Antonyuk 2015, 2020, 2022; Titov 2017 for related data from Russian; Kovačević 2020 for Serbo-Croatian, i.a.), we get a fairly clear idea of what the actual Merge positions must be. Most relevantly for our purposes, the SFD evidence provided here complements the argumentation and the analysis of Inanimate Causers and the Split Voice structure (see Pylkkänen 2002, 2008). In fact, we can adduce additional supporting evidence, also due to QP scope data, that strongly

supports Lavine's (2022, 2023) argumentation for the differentiated structural representation of the Instrumental case-marked Agents of passives and the Inanimate Causer Instrumentals.

3.2.1 On passives vs passive-like: Against the "Undifferentiated Initiator" view

Lavine (2022, 2023) argues against the "Undifferentiated Initiator" idea advanced in Ramchand (2008); Bruening (2013); Legate (2014) and Wood (2017). Examining a range of causative constructions in the crosslinguistic perspective, Lavine argues that the "Initiator" argument in the constructions under investigation is an Oblique Causer (Natural Force) argument that originates in the VP. This Natural Force Instrumental crucially licenses the presence of vCauseP in the structure, which licenses Accusative case on the direct object in the absence of an external argument. While I will come back to the question of the precise Merge position for the Instrumental Inanimate Causer, as this is the question where the analysis proposed here differs from Lavine's, let us for now focus on the latter point, namely the argument against the Undifferentiated Initiator position of Ramchand (2008) and subsequent work. Here, Lavine's conclusions, which he reaches based on independent types of evidence, receive strong support from the quantifier scope data presented here.[16] Perhaps the strongest argument for limiting the domain of application of the Scope Freezing Generalization to vP (i.e., crucially the layer below VoiceP) has been the absence of scope freezing in doubly quantified transitive SVO sentences (i.e., subject QP >> V >> object QP structures) as well as any other constructions involving a subject QP. Crucially, passive sentences do not exhibit scope freezing either (cf. (32) and the related OVS structure in (14)).

(32) Jakas'   kiška   bula nahodovana kožnoju   divčynkoju.
     some.NOM.F cat.NOM.F was.F fed.PASS.F   every.INS girl.INS
     'Some cat was fed by every girl.'
     Surface scope: For some cat x, for every girl y, x was fed by y.
     Inverse scope: For every girl x, for some cat y, x fed y.

---

[16]Lavine (2023) takes on the theoretical question of whether Agents, Instruments and Natural Forces are all realizations of the "macro-Initiator role" and provides evidence that crosslinguistically, causer arguments are differentially realized in the syntax. Some of the strongest evidence to this effect comes from the differences in the distribution of the Inanimate (Instrumental) Causer arguments in Ukrainian, Polish, and Icelandic that sets them apart from the Agents of passives in these languages.

Lavine's analysis of Transitive Impersonals (which in Ukrainian happen to be morphologically indistinguishable from the Instrumental Agent of passives) makes a strong prediction regarding QP scope. It predicts that a passive structure and a superficially similar impersonal construction will differ with respect to the availability of scope freezing. This is so since on Lavine's analysis the NP$_{INS}$ in an impersonal construction is merged lower than the merge position of an Instrumental Agent NP (assumed to be an argument of Voice). Hence, the former, but not the latter, is predicted to show scope freezing. Moreover, based on the data we have already seen, we can predict that transitive impersonals will show scope freezing in the ACC >> INS order of arguments. Both predictions are correct (cf. (34a) and (35a)):[17]

(33) a. Passive
          Cerkv-a   bul-a   spalen-a   {#blyskavk-oju /
          church.F-NOM was-AGR.F burned.down.PASS-AGR.F   lightning-INS
          okupant-amy}.
          invaders-INS
          (Intended:) 'The church was burned down by {the lightning / the invaders}.'

      b. Trans. impers.
          Cerkv-u   bul-o   spalen-o   {blyskavk-oju /
          church.F-ACC was-NON-AGR burned.down-NON-AGR   lightning-INS
          #okupant-amy}.
          invaders-INS
          (Intended:) 'The church was burned down by {the lightning / the invaders}.'

(34) a. Passive
          Jakas'   cerkv-a   bul-a   spalen-a
          some.NOM.SG.F church.F-NOM was-AGR.F burned.down.PASS-AGR.F
          kožnym   okupantom.
          every.INS invader.INS
          'Some church was burned down by every invader.' (scope ambiguous)

---

[17]As detailed in Lavine (2022, 2023), a passive and a transitive impersonal differ with respect to the type of Instrumental NP they require: thus, a passive structure is infelicitous with an Inanimate Causer argument while a transitive impersonal is ungrammatical with an animate/ Agent argument. This fact provides one of the arguments in favor of a differentiated approach to Instrumental NPs in Ukrainian.

b. Jakymos' okupantom bul-a   spalen-a
some.INS.SG invader.INS was-AGR.F burned.down.PASS-AGR.F
kožn-a   cerkv-a.
every-F.NOM church.F-NOM
'Every church was burned down by some invader.' (scope ambiguous)

(35) a. Trans. impers.
Jak-us'   cerkv-u   bul-o   spalen-o
some-ACC.SG.F church.F-ACC was-NON-AGR burned.down-NON-AGR
kožn-oju blyskavk-oju.
every-INS lightning-INS
'Some church was burned by every lightning.'   (scope frozen)

b. Jak-ojus'   blyskavk-oju bul-o   spalen-o
some-INS.SG lightning-INS was-NON-AGR burned.down-NON-AGR
kožn-u   cerkv-u.
every-ACC.F church.F-ACC
'Every church was burned down by some lightning.'

(scope ambiguous)

Of course, we want to ask whether differences regarding scope freezing avail-
ability in the two structures are indeed due to different Merge positions for the
Instrumental Inanimate Causer argument and the Agent argument of a passive
sentence. Needless to say, if we could attribute the lack of scope freezing in pas-
sives to some other factor, the above argument would disappear or be signifi-
cantly weakened. Consider the facts again. Both the subjects of active SVO sen-
tences and the subjects of passives exhibit lack of scope freezing, as do all other
constructions in which one of the two interacting QPs is the subject QP. In other
words, while there are indeed non-trivial structural, semantic, and information-
structural differences between, e.g., OSV, OVS, and passives, conspicuously, the
absence of scope freezing is a property they all share. Furthermore, all internal
arguments irrespective of case marking and structural height (i.e., ACC/DAT/
INSTR) invariably participate in the scope ambiguity-scope freezing distribution
patterns under AI, which is to say that the lack of scope freezing cannot be at-
tributed to some confound that is due to the presence of any of these arguments.[18]

---

[18]As mentioned earlier, the only exception here is the locative/directional PPs, which consis-
tently show surface scope bias, but not the categorical scope freezing. See Antonyuk (2020,
submitted) for detailed discussions and the argument that the existence of this seeming ex-
ception to the Scope Freezing Generalization is not a problem for the SFG, but a source of
additional insight in the search for an adequate, sufficiently restrictive account of scope freez-
ing as a grammatical phenomenon.

Thus, in a passive structure (which, again, differs in terms of scope freezing from the superficially similar transitive impersonal), the difference can hardly be due to anything other than the status of the Agent QP.

One may also wonder whether subjects fail to participate in scope freezing relations due to any of the semantic properties that a subject QP may be associated with by virtue of its syntactic and semantic prominence (e.g., givenness, specificity, topicality, etc). I believe this is unlikely: research on the interaction of specificity and quantification in Ukrainian shows that specificity and QP scope diverge, i.e., a(n object) QP can be specific and take either high or low scope; more generally, further syntactic movement and specificity of either of the two or both objects can neither perturb existing scope freezing nor establish a new scope freezing relation once a QP raises above low temporal/manner adverbs, generally taken to mark the *v*P edge (in current terms: VoiceP edge).[19] This strongly suggests that the status of a subject QP as specific would likewise not interfere with its ability to participate in a scope freezing relation. From what I can tell, the same concerns givenness and topicality, though of course the status of the structurally higher QP as topical/given/specific can lead to some surface scope bias, meaning that the wide scope for the topical/given/specific subject QP may be the preferred interpretation in such a case.

On the Undifferentiated Initiator view all causer arguments, including the non-volitional ones, are arguments of Voice (Kallulli 2006, Ramchand 2008, Bruening 2013, Legate 2014, Wood 2017, i.a.). Thus, on these accounts, we expect that the Instrumental case-marked agents of passives and the Instrumental inanimate/ non-volitional causers of Transitive Impersonal constructions would be syntactically identical, precisely because of their being generated in the same position, Spec,VoiceP. As we have just observed, their behavior is certainly not identical as far as QP scope relations are concerned. Recall that on the account adopted here, the scope freezing relation obtains as a result of a single instance of movement of the structurally lower internal argument QP to a position above the structurally higher one, but, crucially, below the Merge position of the external argument. Thus, in the impersonal *-no/-to* constructions, scope freezing obtains when the lower NP$_{ACC}$ raises overtly above NP$_{INS}$. If we assume the generalization regarding scope freezing is correct, then on the Undifferentiated Initiator view, which has all initiators merged as arguments of Voice, *all* structures involving any type of initiator argument are then predicted to lack scope freezing, precisely because, as we have shown, scope freezing emerges in a layer of structure that crucially excludes VoiceP.

---

[19]See Antonyuk & Mykhaylyk (2022) for details.

Needless to say, of course, much will ultimately depend on the account of passives one adopts, and this is certainly a research area characterized by continued debate as well as significant differences in the general approach of individual researchers. Let us focus here on the structural, configurational accounts in which the agent phrase of a passive is base-generated in the same position as in the active, in accordance with Baker's (1988) UTAH (see esp. Collins 2005; cf. Bruening 2013; Hallman 2021, i.a.). On Collins' (2005) influential account, a participial phase containing the VP fronts around the agent in its Merge position, thus inverting the hierarchical relation between the agent and the object phrase. On an account of passives roughly such as this one, the lack of scope freezing in passives follows naturally. This is so since the agent is merged in its usual position, Spec,VoiceP (assumed to be Spec,vP in Collins 2005), and we have seen extensive evidence that external arguments never participate in the surface scope freezing relation, thus we expect the same in passives. On the present account, which follows Lavine (2022, 2023), what sets the non-volitional/inanimate causer Instrumentals apart from the agent Instrumentals of passives is precisely the structural difference in Merge positions, with the former being merged in a structurally lower position than the latter, hence the correctly predicted differences in scope behavior.[20]

Finally, note that I have provided no independent evidence that the Agent of passives is generated in the same position as the Agent of active transitives, though, arguably, the scope data (i.e., the same scope-taking ability of passives as other structures involving the Agent QP) do provide indirect evidence to this effect. Crucially, however, to argue against the Undifferentiated Initiator view, it is enough to show that the Instrumentals in Transitive Impersonals are generated in a position other than Spec,VoiceP. I conclude that the scope difference between a passive (34) and a transitive impersonal (35) is indeed due to a difference in the structural position of the Instrumental argument in the two types of constructions under consideration. Thus, we gain a novel argument against

---

[20]A question remains, of course, why the nature of a QP's Merge position should be relevant for establishing a scope freezing relation. While I cannot do this question justice here, in my related work (Antonyuk submitted; Antonyuk 2025) I argue that scope freezing is a first-phase syntax phenomenon, which arises when the lower QP raises into the specifier of ApplP (which I argue is located in Slavic between VoiceP and vP). The overt movement of the lower argument into the specifier of ApplP is argued to define and lead to the Spell-out of the first phase, the domain in which the scope freezing relation is set. Note that irrespective of whether this account is correct, what we are crucially relying on here is the empirically grounded generalization that the domain of scope freezing excludes the subject QP. Thus, the external argument, taken to be generated in Spec,VoiceP, is always going to be literally too high, being outside the domain in which this scope relation can be established.

the Undifferentiated Initiator view (Ramchand 2008; Legate 2014; Wood 2017, i.a.) and in favor of making a distinction between a higher verbal layer, VoiceP, which introduces the external argument and the lower verbal layer, *v*P, which introduces causative semantics and an inanimate causer argument in the absence of Voice (Lavine 2022, 2023).

3.2.2 On the VP-external Merge position of argument Instrumentals

Going back to the previous point, i.e., the exact nature of the Inanimate Causer's Merge position, as mentioned earlier, and as should be clear from the SFD insights already reviewed, I argue that the Merge position of Inanimate Causer Instrumentals cannot be VP-internal. Lavine (2022, 2023), on the other hand, posits a low VP-internal position in the complement of V, which is traditionally reserved for Oblique arguments.

As a reminder, the argument hierarchy deduced from applying the SFD to a variety of constructions is the following: V NP$_{INS}$ >> NP$_{ACC}$ >> NP$_{DAT}$. Thus, while NP$_{DAT}$ is shown to pattern with the Obliques (i.e., arguably sharing the same Merge position with directional and locative PPs), the Inanimate Causer Instrumentals are in fact shown to have the highest Merge position of the three arguments. Hence, to the extent the SFD is accepted as being correct, the complement of V position for NP$_{INS}$ becomes ruled out. Spec,V, on the other hand, might be somewhat more likely. Specifically, in dyadic constructions involving two internal arguments, NP$_{ACC}$ and NP$_{INS}$, the former is consistently shown by the SFD to be merged lower than NP$_{INS}$, hence it is conceivable that the Merge position for the two arguments could simply be the reverse of what is commonly assumed.[21] This conclusion appears equally unavailable, however, once we consider again examples involving three internal arguments (ex. (28) and (29) repeated here for convenience as (36) and (37)).[22]

(36) Xlopcevi vidrizalo   palec'   elektryčnoju pyloju.
      boy.DAT   severed.NON-AGR finger.ACC electric.INS   saw.INS
      'The boy's finger got severed by an electric saw.'

---

[21]This is precisely what Landau (2010) proposes as far as the position of the Accusatives in object experiencer constructions (i.e., an Accusative as a concealed low Oblique), which in Ukrainian involve Accusative and Instrumental argument NPs.

[22]The # sign next to (38a) is meant to indicate that this example (just like (23a)) is infelicitous on its surface scope interpretation in that it describes a situation where the same finger is severed with every instrument in some contextually salient set of instruments. Our world knowledge tells us this situation is impossible or improbable, but in terms of logico-semantic properties of the sentence, this interpretation indeed obtains and is in fact the only interpretation available for this sentence.

(37) a. Jakomus' xlopcevi vidrizalo   palec'   kožnym
       some.DAT guy.DAT severed.NON-AGR finger.ACC every.INS
       instrumentom.
       instrument.INS
       'Some boy got a finger severed by every instrument.'   (frozen)

   b. Jakymos' instrumentom vidrizalo   palec'   kožnomu
       some.INS instrument.INS severed.NON-AGR finger.ACC every.DAT
       xlopcevi.
       boy.DAT
       'With some instrument (or other), every boy's finger was severed.'
                                                                (ambiguous)

   ⇒ BASE ORDER: V NP$_{\text{INS}}$ >> NP$_{\text{DAT}}$

(38) a. # Xlopcevi vidrizalo   jakyjs' palec'   kožnym
         boy.DAT severed.NON-AGR some   finger.ACC every.INS
         instrumentom.
         instrument.INS
         'Some boy got a finger severed by every instrument.'   (frozen)

   b. Xlopcevi vidrizalo   jakymos' instrumentom kožen
       boy.DAT severed.NON-AGR some.INS finger.INS   every.ACC
       palec'.
       finger.ACC
       'Some boy got every finger severed by some instrument.' (ambiguous)
   ⇒ BASE ORDER: V NP$_{\text{INS}}$ >> NP$_{\text{ACC}}$

What we see here is that when NP$_{\text{DAT}}$, NP$_{\text{ACC}}$ and NP$_{\text{INS}}$ are all arguments of the same verb, the SFD points to the same relative ordering of arguments, i.e., V NP$_{\text{INS}}$ >> NP$_{\text{ACC}}$ >> NP$_{\text{DAT}}$. The above examples, while not producing any new evidence beyond that discussed earlier in the paper, nevertheless drive home the point that there is simply no place left in the VP for the Inanimate Causer argument. Hence, my proposal (in accordance with Lavine 2022, 2023 and Pylkkänen 2002, 2008), developed in detail in related work, is that the data considered in this paper are to be interpreted as evidence in favor of the Un-bundled Voice projection in Ukrainian, the lower of which, call it *v*CauseP, fol-lowing Lavine (2023), hosts the Inanimate Causer Argument while the higher one, VoiceP (Kratzer 1996), hosts the Agent external argument. Without going into the specifics of the larger proposal here, I will point out that semantically, it makes good sense for a Causer argument to be merged higher than the other VP-internal arguments (NP$_{\text{ACC}}$, NP$_{\text{DAT}}$, PP); syntactically, apart from the purely

theoretical reasons for favoring NP$_{INS}$ as being merged in Spec,vCauseP, we now have solid empirical evidence that supports precisely this view.[23]

4 Conclusions

In this paper I have provided a novel argument against the Undifferentiated Initiator view (Ramchand 2008; Bruening 2013; Legate 2014 and Wood 2017), which posits distinct structural Merge positions for the Instrumental Agent of passives and the morphologically indistinguishable Instrumental Inanimate Causer argument of a range of "Transitive Impersonal" constructions in Ukrainian, thus providing independent support for the main theoretical claim in Lavine (2022, 2023). The novel evidence presented in this paper shows that the Instrumental Agents of passives and the Instrumental Inanimate Causers (e.g., in Ukrainian non-agreeing accusative constructions) behave differently with respect to QP scope, with the latter, but not the former, being able to participate in establishing a surface scope freezing relation. In this regard passives behave very much in line with a range of transitive constructions involving a subject QP: as shown in the paper, subject QPs categorically resist participating in scope freezing (Larson 1990; Bruening 2001). Thus, while no novel evidence bearing on the Merge position of the subject of passives is provided here, the paper does provide indirect support for configurational, structural accounts of passives such as Collins (2005), by showing strong similarities with respect to non-participation in scope freezing on the part of subject QPs in SVO, OSV, SOV, OVS and passive sentences. In other words, subject QPs are outside the domain in which scope freezing can be established, and the subjects of passives in this respect behave similarly to all other external arguments of transitive sentences.

I have also provided evidence in favor of a higher Merge position than is commonly assumed for a variety of Instrumental case-marked arguments. I suggest that their Inanimate Causer semantics ensures their being merged above all other internal arguments. Specifically, I argue that the Inanimate Causer Instrumentals must be merged outside the core VP (cf. Lavine 2022, 2023 for a low complement of V position for all inanimate cause Instrumentals), the most likely candidate for the Merge position being vCauseP (i.e., the lower Voice projection in Pylkkänen's 2002, 2008 Unbundled Voice proposal), which introduces causative semantics in the absence of Voice, thus very much being in the spirit of Lavine's work (see also esp. Harley 2013).

---

[23]As mentioned earlier, Lavine (2023) treats the inanimate causer arguments in Transitive Impersonal constructions as the Natural Force argument, which is arguably well aligned with the standard view of Instrumental arguments as being merged low in the VP (cf. Szucsich 2007; Schäfer 2008; Junghanns et al. 2017; Wood 2017 i.a. for other theoretical solutions).

Finally, it should be pointed out that the above arguments were all made based on the insights afforded by the Scope Freezing Diagnostic (Antonyuk 2015, 2020, 2022, submitted), which is a testament to its diagnostic utility. Here as elsewhere, the SFD is shown to be a remarkably consistent diagnostic tool, one that is especially well suited for the Discourse Configurational Languages such as Ukrainian and the rest of Slavic where overt syntactic movement correlates with semantic interpretation as well as discourse-related properties of the sentence.

Abbreviations

ACC	accusative	NON-AGR	non-agreeing form
AGR	agreement	PASS	passive
DAT	dative	PL	plural
F	feminine	PST	past
GEN	genitive	REFL	reflexive
INS	instrumental	SG	singular
NOM	nominative		

Acknowledgments

References

Abels, Klaus & Dagmara Grabska. 2022. On the distribution of scope ambiguities in Polish. *Glossa: a journal of general linguistics* 7(1). DOI: 10.16995/glossa.8170.

Alexiadou, Artemis. 2001. *Functional structure in nominals*. Amsterdam: John Benjamins. DOI: 10.1075/la.42.

Alexiadou, Artemis. 2009. On the role of syntactic locality in morphological processes: The case of (Greek) derived nominals. In Anastasia Giannakidou & Monika Rathert (eds.), *Quantification, definiteness and nominalization*, 253–280. Oxford: Oxford University Press. DOI: 10.1093/oso/9780199541089.003.0011.

Alexiadou, Artemis. 2017. On the complex relationship between deverbal compounds and argument supporting nominals. In Maria Bloch-Trojnar & Anna Malicka-Kleparska (eds.), *Aspect and valency in nominals*, 53–82. Berlin/Boston: De Gruyter Mouton. DOI: 10.1515/9781501505430-003.

Antonyuk, Svitlana. 2015. *Quantifier scope and scope freezing in Russian.* Stony Brook: Stony Brook University. (Doctoral dissertation).

Antonyuk, Svitlana. 2020. The puzzle of Russian ditransitives. In Anna Pineda & Jaume Mateu (eds.), *Dative constructions in Romance and beyond*, 43–74. Berlin: Language Science Press. DOI: 10.5281/zenodo.3776533.

Antonyuk, Svitlana. 2021. Russian OVS: Towards a better understanding of the construction's properties and significance. *Journal of Slavic Linguistics* 29(FASL 28 issue). http://ojs.ung.si/index.php/JSL/article/view/156.

Antonyuk, Svitlana. 2022. Base-generated or derived? Here's how to tell structures apart in Russian. *Annual Review of the Faculty of Philosophy, Novi Sad* 47(3). 111–133. DOI: 10.19090/gff.2022.3.111-133.

Antonyuk, Svitlana. 2025. *On quantification and phase theory: A phase-based approach to the scope freezing phenomenon.* Ms. University of Graz.

Antonyuk, Svitlana. Submitted. On methodology: The scope freezing diagnostic as a probe of argument structure relations. In Svitlana Antonyuk, Ángel Jiménez-Fernández & Isabel Oltra-Massuet (eds.), *The syntax of argument structure alternations across frameworks, Special Collection issue at Glossa: a journal of general linguistics.*

Antonyuk, Svitlana & Roksolana Mykhaylyk. 2022. Scope freezing and object shift in Ukrainian: Does superiority matter? Remark. *Syntax* 25. 122–146. DOI: 10.1111/synt.12229.

Asarina, Alsa. 2005. *Russian Binding Theory: Two improved movement approaches.* Ms. MIT.

Bailyn, John Frederick. 1995. *A configurational approach to Russian "free" word order.* Ithaca: Cornell University. (Doctoral dissertation).

Bailyn, John Frederick. 2010. What's inside VP? New (and old) evidence from Russian. In Wayles Browne (ed.), *Formal approaches to Slavic linguistics (FASL) #18: The Cornell meeting 2009*, 21–37. Ann Arbor: Michigan Slavic Publications.

Bailyn, John Frederick. 2012. *The syntax of Russian.* Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511984686.

Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing.* Chicago: Chicago University Press. https://dspace.mit.edu/handle/1721.1/15069.

Barss, Andrew & Howard Lasnik. 1986. A note on anaphora and double objects. *Linguistic Inquiry* 17(2). 347–354. https://www.jstor.org/stable/4178491.

Boneh, Nora & Léa Nash. 2017. The syntax and semantics of dative DPs in Russian ditransitives. *Natural Language & Linguistic Theory* 35. 899–953. DOI: 10.1007/s11049-017-9360-5.

Borer, Hagit. 2014. Derived nominals and the domain of content. *Lingua* 141. 71–96. DOI: 10.1016/j.lingua.2013.10.007.

Bruening, Benjamin. 2001. QR obeys superiority: Frozen scope and ACD. *Linguistic Inquiry* 32(2). 233–273. DOI: 10.1162/00243890152001762.

Bruening, Benjamin. 2013. *By*-phrases in passives and nominals. *Syntax* 16. 1–41. DOI: 10.1111/j.1467-9612.2012.00171.x.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick Jacobs & Peter Rosenbaum (eds.), *Readings in English transformational grammar*, 184–221. Waltham: Ginn & Company.

Collins, Chris. 2005. A smuggling approach to the passive in English. *Syntax* 8. 81–120. DOI: 10.1111/j.1467-9612.2005.00076.x.

Dyakonova, Marina. 2009. *A phase-based approach to Russian free word order* (LOT Dissertation Series 230). Utrecht: LOT.

Franks, Steven. 1995. *Parameters of Slavic morphosyntax*. Oxford: Oxford Academic. DOI: 10.1093/oso/9780195089707.001.0001.

Greenberg, Gerald & Steven Franks. 1991. A parametric approach to dative subjects and the second dative in Slavic. *The Slavic and East European Journal* 35(1). 71–97. DOI: 10.2307/309034.

Grimshaw, Jane B. 1990. *Argument structure* (Linguistic Inquiry Monographs 18). Cambridge: MIT Press.

Hallman, Peter. 2018. Double-object constructions in Syrian Arabic. *Syntax* 21. 238–274. DOI: 10.1111/synt.12157.

Hallman, Peter. 2021. On passive and perfect participles. In Kleanthes Grohmann, Akemi Matsuya & Eva-Maria Remberger (eds.), *Passives cross-linguistically: Theoretical and exprimental approaches*, 64–97. Amsterdam: Brill. DOI: 10.1163/9789004433427_004.

Hallman, Peter. 2024. Argument structure hierarchies and alternations in causative and double object constructions. *Glossa: a journal of general linguistics* 9(1). DOI: 10.16995/glossa.9110.

Harley, Heidi. 2009. The morphology of nominalizations and the syntax of *v*P. In Anastasia Giannakidou & Monika Rathert (eds.), *Quantification, definiteness, and nominalization*, 321–343. Oxford: Oxford University Press. DOI: 10.1093/oso/9780199541089.003.0013.

Harley, Heidi. 2013. External arguments and the mirror principle: On the distinctness of Voice and *v*. *Lingua* 125. 34–57. DOI: 10.1016/j.lingua.2012.09.010.

Harley, Heidi & Shigeru Miyagawa. 2017. Syntax of ditransitives. In Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics.* Oxford: Oxford University Press. DOI: 10.1093/acrefore/9780199384655.013.186.

Iordǎchioaia, Gianina. 2020. Event structure and argument realization in English zero-derived nominals with particles. *Nordlyd* 44(1). 35–51. DOI: 10.7557/12.5205.

Junghanns, Uwe, Denisa Lenertová & Dorothee Fehrmann. 2017. Parametric variation of Slavic accusative impersonals. In Olav Mueller-Reichau & Marcel Guhl (eds.), *Aspects of Slavic linguistics: Formal grammar, lexicon and communication*, 140–165. Berlin: Walter de Gruyter. DOI: 10.1515/9783110517873-007.

Kallulli, Dalina. 2006. A unified analysis of passives, anticausatives, and reflexives. In Olivier Bonami & Patricia Cabredo Hofherr (eds.), *Empirical issues in formal syntax and semantics*, vol. 6, 201–226. Paris: CNRS. http://www.cssp.cnrs.fr/eiss6/kallulli-eiss6.pdf.

Kovačević, Predrag. 2020. Not all object experiencers are the same: The role of *se* and the argument structure of Serbian psych verbs anticausatives. *Bucharest Working Papers in Linguistics* 22(2). 77–96. DOI: 10.31178/BWPL.22.2.4.

Kratzer, Angelika. 1996. Severing the external argument from its verb. In Johan Rooryck & Laurie Zaring (eds.), *Phrase structure and the lexicon*, 109–137. Dordrecht: Kluwer. DOI: 10.1007/978-94-015-8617-7_5.

Landau, Idan. 2010. *The locative syntax of experiencers.* Cambridge: MIT Press.

Larson, Richard K. 1990. Double objects revisited: Reply to Jackendoff. *Linguistic Inquiry* 21. 589–632. https://www.jstor.org/stable/4178697.

Larson, Richard K., Svitlana Antonyuk & Lei Liu. 2019. Superiority and scope freezing. *Linguistic Inquiry* 50(1). 233–252. DOI: 10.1162/ling_a_00300.

Lavine, James E. 2022. Verbal layers and initiators in Ukrainian. Invited talk given at Formal Description of Slavic Languages (FDSL) 15. Berlin.

Lavine, James E. 2023. *Verbal layers and initiators in Ukrainian.* Ms. Bucknell University, Lewisburg, PA.

Lavine, James E. & Robert Freidin. 2002. The subject of defective T(ense) in Slavic. *Journal of Slavic Linguistics* 10(1/2). 251–287. http://www.jstor.org/stable/24599685.

Legate, Julie Anne. 2014. *Voice and v: Lessons from Acehnese.* Cambridge: MIT Press. https://www.jstor.org/stable/j.ctt9qf6h9.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4(2). 201–225. https://repository.upenn.edu/handle/20.500.14332/45440.

Marvin, Tatjana & Adrian Stegovec. 2012. On the syntax of ditransitive sentences in Slovenian. *Acta Linguistica Hungarica* 59(1–2). 177–203. DOI: 10.1556/aling. 59.2012.1-2.8.

McGinnis, Martha. 2020. Cross-linguistic contrasts in the structure of causatives in clausal nominalizations. In Bronwyn Moore Bjorkman & Daniel Currie Hall (eds.), *Contrast and representations in syntax*, 138–178. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198817925.003.0006.

Pesetsky, David. 1995. *Zero syntax: Experiencers and cascades.* Cambridge: MIT Press.

Pylkkänen, Liina. 2002. *Introducing arguments.* Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Pylkkänen, Liina. 2008. *Introducing arguments* (Linguistic Inquiry Monographs 49). Cambridge: MIT Press.

Ramchand, Gillian Catriona. 2008. *Verb meaning and the lexicon: A first-phase syntax.* Cambridge: Cambridge University Press.

Richards, Norvin. 1997. *What moves where when in which language?* Cambridge: MIT. (Doctoral dissertation).

Richardson, Kylie. 2007. *Case and aspect in Slavic.* Oxford: Oxford University Press.

Schäfer, Florian. 2008. *The syntax of (anti-)causatives: External arguments in change-of-state contexts* (Linguistik Aktuell/Linguistics Today 126). Amsterdam: John Benjamins. DOI: 10.1075/la.126.

Schneider-Zioga, Patricia. 1988. Double objects and small clauses. Ms. University of Southern California. Los Angeles.

Szucsich, Luka. 2007. Nothing wrong with finite T: Non-agreeing accusative impersonal sentences. In Magda Golędzinowska, Ulyana Savchenko & Richard Compton (eds.), *Formal approaches to Slavic linguistics 15*, 417–435. Ann Arbor: Michigan Slavic Publications.

Titov, Elena. 2017. The canonical order of Russian objects. *Linguistic Inquiry* 48(3). 427–457. DOI: 10.1162/ling_a_00249.

Wood, Jim. 2017. The Accusative-Subject Generalization. *Syntax* 20. 249–291. DOI: 10.1111/synt.12138.

Wood, Jim. 2023. *Icelandic nominalizations and allosemy.* Oxford: Oxford University Press.

Chapter 2

Slavic creation/consumption predicates in light of Talmy's typology

Alessandro Bigolin

Centre de Lingüística Teòrica, Universitat Autònoma de Barcelona

The chapter is concerned with the licensing of creation/consumption predicates in Slavic languages, in light of Talmy's (2000) typology. I present the results of a pilot study suggesting that Slavic languages behave as verb-framed languages in the domain of creation/consumption predicates, despite these languages being commonly regarded as a type of satellite-framed languages (Talmy 2000) referred to as "weak satellite-framed languages" (Acedo-Matellán 2010; 2016). Assuming a neo-constructionist view on argument structure, I propose a morphosyntactic account of Talmy's typology, according to which the verb-framed vs. satellite-framed distinction depends on a specific Phonological Form requirement. In verb-framed languages, the null functional head involved in verbal predication must, by assumption, incorporate its complement as an externalization condition. I propose that so-called weak satellite-framed languages, to which Slavic languages have been argued to belong, are fundamentally verb-framed languages, and that the availability of satellite-framed resultative constructions in these languages is granted by the lexical presence of result morphemes that can incorporate into *v* via prefixation.

1 Introduction

This chapter is concerned with the licensing of creation/consumption predicates in Slavic languages, in light of Talmy's (2000) typology. In creation/consumption predicates, the direct object is understood as being "created" or "consumed" during the event denoted by the predicate (Hale & Keyser 2002; Volpe 2004; Harley 2005; Mateu 2012, among others). For instance, *a hole* in (1a) is formed while the digging process takes place, and *the apple* in (1b) is consumed during the eating process.

(1)   a. He dug a hole in his garden.   (Washio 1997: 46)

    b. John ate the apple.   (Folli & Harley 2005: 103)

The typological distinction proposed by Talmy (2000) divides languages into two broad classes, depending on how the Path (or "change") core component of re-sultative events of change of state and location is expressed.[1] In one class of languages, Path is typically encoded in a satellite (e.g., a particle, PP or AP) distinct from the main verb, which in turn may express a co-event.[2] The co-event usually provides information about the manner in which the main resultative event unfolds, or about the cause which triggers it.[3] In the other class of languages, Path is always encoded in the main verb, so that information about a co-event is either not expressed or provided via adjuncts. Languages of the former type are thus referred to as "satellite-framed", while languages of the latter type are referred to as "verb-framed". The examples from English (a satellite-framed language) and Spanish (a verb-framed language) in (2) and (3) illustrate the two patterns, for events of change of location and events of change of state respectively.

(2)   Satellite-framed pattern (English):

    a. The bottle [floated]$_{\text{CO-EVENT}}$ [into the cave]$_{\text{PATH}}$.   (Talmy 2000: 227)

    b. She [shot]$_{\text{CO-EVENT}}$ him [dead]$_{\text{PATH}}$.   (Goldberg 1995: 136)

(3)   Verb-framed pattern (Spanish):

    a. La botella [entró]$_{\text{PATH}}$ ([flotando]$_{\text{CO-EVENT}}$) a la cueva.
       the bottle enter.PST.AGR float.GER   to the cave
       'The bottle moved into the cave (floating).'   (Talmy 2000: 227)

    b. Lo   [mató]$_{\text{PATH}}$ ([de un disparo]$_{\text{CO-EVENT}}$).
       ACC.M.SG kill.PST.AGR of a shot
       'He/she killed him with a shot.'   (CORPES XXI[4])

Slavic languages, along with Latin, have been classified as "weak satellite-framed" (Acedo-Matellán 2010; 2016) since, although they allow the expression

---

[1]I consider an event "resultative" if it involves a scalar change along a scale that denotes a property or a path (Rappaport Hovav 2014).

[2]PPs are explicitly excluded in Talmy's (2000) notion of "satellite", defined as "[...] the grammatical category of any constituent other than a noun phrase or prepositional-phrase complement that is in a sister relation to the verb root" (Talmy 2000: 120). Following Mateu (2002), Beavers et al. (2010), Acedo-Matellán & Mateu (2013), Acedo-Matellán (2016), among others, I adopt a broader definition of satellite, which includes non-adjunct result PPs like the one in (2a).

[3]See Talmy (2000) for an exhaustive classification of possible conceptual interpretations attributable to co-events.

[4]*Corpus del Español del Siglo XXI*, Real Academia Española (2024).

of Path in a satellite, this must form a prosodic word with the verb. For instance, the object *svoju ručku* 'her pen' in (4) is understood to be brought into a state where all its ink is used up by means of the prefixal satellite *iz-* 'out', while the verb *pis-* 'write' specifies the co-event that causes the transition undergone by the referent of the direct object (Spencer & Zaretskaya 1998; Mateu 2008).

(4) Ona   [iz]_{PATH}-[pis]_{CO-EVENT}-a-l-a svoju ručku.
     she.NOM out-write-TH-PST-AGR   POSS pen.ACC
     'Her pen has run out of ink.' (Lit. 'She has written her pen out (of ink).')
                                                    (Russian; Spencer & Zaretskaya 1998: 17)

The satellite-framed/verb-framed distinction is also found in the domain of predicates denoting events of creation/consumption (Mateu 2003; 2012). In a similar way to (2), satellite-framed languages allow the expression of a co-event in the verb in creation/consumption predicates, giving rise to creation/consumption predicates of the type in (5a) (hereafter, "complex creation/consumption predicates"). The predicate in (5a) can be paraphrased as "make a hole in the coat by brushing", whereby it is clear that the main verb of the predicate is understood as specifying a co-event of the main event of creation. Verb-framed languages instead consistently express the event that leads to the creation/consumption of the direct object by means of the main verb, which may be either a light verb (e.g., *make*, as in the Spanish example in (5b)) or a verb whose meaning is likely to imply the creation/consumption of the object, which in turn is interpreted as a hyponym of the verb (as in (1); see Hale & Keyser 1997; 2002). The specification of a possible co-event, as in the verb-framed change-of-location/state examples in (3), is relegated to an optional adjunct.

(5) a. Brush a hole in one's coat.   (Levin & Rapoport 1988: 279)

    b. Hizo   un agujero en su   abrigo (al   cepillar=lo).
       make.PST.AGR a   hole   in POSS coat   at.the brush.INF=ACC.M.SG
       'She made a hole in her coat, by brushing it.'

A non-trivial difference between creation/consumption predicates and change-of-location/state predicates is that the argument structure of creation/consumption predicates has been argued to lack a Path component (see Rappaport Hovav & Levin 1998; Rappaport Hovav 2008; Rappaport Hovav & Levin 2010 for works adopting a lexicalist approach; see Hale & Keyser 1993; 2002; Mateu 2002; Harley 2005; Folli & Harley 2005; 2008; 2020; Ramchand 2008; Acedo-Matellán 2016, among others, for works adopting a neo-constructionist, syntactic approach). Accordingly, in light of contrasts like (5), Mateu (2012) concludes that a proper

descriptive account of the cross-linguistic variation associated with Talmy's typology should not be understood in terms of a requirement about the expression of Path (either in the main verb or in a verb's satellite), but rather in terms of whether or not a language allows the expression of a co-event in the main verb.

A prediction of this line of reasoning is that weak satellite-framed languages such as Slavic languages should allow complex creation/consumption predicates of the type in (5a), since these languages more generally display constructions where the main verb expresses a co-event (as exemplified in (4)). In this chapter, I present the results of a pilot study investigating the availability of different types of creation/consumption predicates in several Slavic languages, comparing them with data from bona fide satellite-framed languages and verb-framed languages. I provide evidence suggesting that Slavic languages behave as verb-framed languages in the domain of creation/consumption predicates, as they must resort to run-of-the-mill verb-framed strategies to express such predicates and rule out constructions such as complex creation/consumption predicates (Section 2). Assuming a neo-constructionist approach to argument structure (Mateu 2002; Borer 2005; Mateu & Acedo-Matellán 2012, among others), I propose a morphophonological account of Talmy's typology, which is argued to follow from a Phonological Form (PF) requirement, in verb-framed languages, on the null syntactic head *v* involved in verbal predication. I suggest that Slavic languages, and weak satellite-framed languages in general, should be considered as fundamentally verb-framed languages, predicting the unavailability of complex creation/consumption predicates in this class of languages (Section 3). Next, I explore the prediction – following from the present account – that a complex creation/consumption reading is available in Slavic languages for predicates that are perfectivized via so-called "internal" verbal prefixes (Svenonius 2004; Borik 2006; Gehrke 2008, among others), which have been argued to express an abstract result in a resultative construction (Gehrke 2008; Acedo-Matellán 2016; Kwapiszewski 2022, among others) (Section 4). Finally, I address some potential counterexamples from Latin (another weak satellite-framed language; Acedo-Matellán 2016) to the prediction that weak satellite-framed languages lack complex creation/consumption predicates of the type found in satellite-framed languages. I argue that Latin lacked such predicates in the same way as Slavic languages do, *pace* Acedo-Matellán (2016) and consistently with the predictions of the present account (Section 5). I draw general conclusions in Section 6.

2 Creation/consumption predicates in Slavic languages

In order to investigate the availability of complex creation/consumption predicates in Slavic languages, I carried out a pilot study to check, with the help of

native speakers, whether it was possible to directly translate different creation/consumption predicates that are licensed in satellite-framed English into several Slavic languages. I further examined whether it was possible to directly translate the English examples into four additional bona fide satellite-framed languages and five verb-framed languages, respectively. Effort was invested in gathering evidence from different language families, contributing to the diversity of languages represented in the collected data. For the class of satellite-framed languages, data were collected from Dutch, German, Chinese, and Hungarian. Regarding verb-framed languages, data were collected from Italian, Catalan, Spanish, Basque, and Greek. Finally, for the class of Slavic languages, data were collected from Russian and Ukrainian (East Slavic languages), Polish and Slovak (West Slavic languages), and Serbian and Croatian (South Slavic languages).[5]

2.1 The English data

The English examples range from constructions involving verbs whose meaning can be taken to imply the creation/consumption of the direct object, therefore using a verb-framed strategy, to constructions that can be taken to involve the expression of a manner co-event in the main verb, and which are expected to be ungrammatical in verb-framed languages.[6] The list of the selected examples, starting with verb-framed constructions, is provided in (6) to (24).[7]

(6) John sang a song. (Truswell 2007: 1361)

---

[5]Serbian and Croatian are considered individually alongside the other languages examined, notwithstanding classifications that see them as distinct varieties of a single language (e.g., Serbo-Croatian, or BCMS).

[6]The selection of the data was primarily based on examples from relevant literature pertaining to hyponymous objects, effected objects, and Talmy's typology. Additionally, some examples were taken from corpora or made up and subsequently checked with native speakers. Following Mateu (2002), I have included the examples in (18) and (19) as representatives of the class of complex consumption predicates, where the consumption of the direct object constitutes the main event denoted by the predicate, while the verb denotes a co-event. See Kuno (1973) and Condamines (2013) for possible examples of this type in verb-framed Japanese and French, respectively (I thank an anonymous reviewer for bringing my attention to the data analyzed in these works, which deserve further investigation).

[7]The examples have been arranged in the present order based on my own intuitions, as a native speaker of one of the verb-framed languages tested, about the degree of "manner" provided by the verb in each of them. Determining the degree of manner provided by the verb in each of the sentences in (6) to (24) is a complex process that takes place at the conceptual level. To the best of my knowledge, there is currently no objective method to quantitatively measure the degree of manner provided by the verb in a specific construction, leaving the intuition-based approach as the only viable option.

(7)   They danced a Sligo jig.   (Gallego 2012: 98)

(8)   Ariel ate the mango.   (Ramchand 2008: 52)

(9)   He dug a hole in the ground.   (COCA[8])

(10)   She wove the tablecloth.   (adapted from Folli & Harley 2020: 452)

(11)   Marco painted a sky.   (Folli & Harley 2020: 438)

(12)   Maria carved a doll.   (Folli & Harley 2020: 439)

(13)   She burned a hole in her coat.   (made up)

(14)   He scratched a hole in the ground.   (COCA)

(15)   She punctured a wound in her finger.   (made up)

(16)   She cut a wound in her foot.   (made up)

(17)   She bit a hole in the bag.   (COCA)

(18)   The adventurer walked the trail.
               (Mateu 2002: 297, adapted from Tenny 1994: 17)

(19)   The adventurer swam the channel.
               (Mateu 2002: 297, adapted from Tenny 1994: 17)

(20)   Deanne kicked a hole in the wall.   (COCA)

(21)   She magicked a cursor.   (COCA)

(22)   She brushed a hole in her coat.
        (Mateu & Rigau 2002: 213, adapted from Levin & Rapoport 1988)

(23)   John smiled his thanks.
           (Mateu 2012: 255, adapted from Levin & Rapoport 1988)

(24)   Elna frowned her discomfort. (Acedo-Matellán & Kwapiszewski 2021: 35)

---

[8]Corpus of Contemporary American English (Davies 2008–).

All the examples in (6) to (24) are taken to lack a Path component in their argument structure. While this is proposed by much work adopting both the lexicalist approach and the neo-constructionist approach (as pointed out in Section 1), such work is mostly concerned with the argument structure of (verb-framed) predicates in which the meaning of the verb can be taken to imply the creation/consumption of the object. Following Mateu (2012); Acedo-Matellán (2016); Folli & Harley (2020), among others, I extend such an analysis to satellite-framed predicates of creation/consumption in which the verb is taken to express a co-event. At first sight, predicates of this type might be argued to involve the argument structure of resultative predicates since most of them typically require a locative PP, which is instead omissible in predicates of creation/consumption that involve a verb-framed strategy. See, in this respect, the contrast between (1a), assumed to be verb-framed, and (5a), repeated in (25a) and (25b), respectively.[9]

(25)   a.   He dug a hole (in his garden).
       b.   Brush a hole *(in one's coat).

Based on the contrast in (25), the current assumption that satellite-framed predicates denoting events of creation/consumption do not involve a Path component in their argument structure might be questioned. Specifically, an anonymous reviewer suggests that the PP could be expressing a null Path in English predicates of the type in (5a) in the same way as it seems to do in predicates denoting events of change such as *walk in the room*, considered by the reviewer to be ambiguous between a locative and a change-of-location reading (but see, e.g., Folli & Ramchand 2005: 83 and Gehrke 2008: 90 for a different opinion).

The remainder of this subsection is devoted to showing that satellite-framed predicates of creation/consumption should not be taken to involve a null Path element in their argument structure. I argue that several reasons support this conclusion, even though the contrast in (25), at first sight, might seem to suggest otherwise. First, the claim that the PP in *walk in the room* involves a phonologically null Path is disputable since Path, in such a predicate, has been argued in previous works to be expressed by the verb *walk* (Alexiadou 2015; further see Ramchand 2008: 112, fn. 1; Nikitina 2008; Beavers et al. 2010). This verb, given the right context, may be coerced by some speakers into an interpretation as involving directionality and hence goal of motion. This explains the existence of contrasts like the one depicted in (26). Unlike *walk*, *dance* denotes an activity that typically does not imply directionality. As a result, this verb is less likely to

---

[9]The judgment in (25b) is by an anonymous reviewer.

express Path, which must therefore be expressed independently in order for the verb to appear in the change-of-location frame.

(26) a. John walked in the room.   (in a change-of-location reading)
      b. #John danced in the room.   (in a change-of-location reading)
                                                          (Alexiadou 2015: 1093)

Additionally, if the satellite-framed predicates of creation/consumption dis-cussed in this chapter involved a phonologically null Path, the question would arise as to why Path *must* be null in these predicates. Even by assuming that (26a) is compatible with a change-of-location reading, Path can optionally be overtly realized independently of the verb in resultative predicates of this type, as (27) shows.

(27) John walks in(to) the room.   (in a change-of-location reading)

More strikingly, Path is mandatorily realized by a morpheme different from the verb in transitive resultatives featuring direct objects that are not semantically selected by the verb (meaning that they are not a traditional object of the verb based on what lexicalist approaches consider to be the verb's lexical argument structure, and would not be suitable objects of such a verb outside the resultative construction); see the contrast between the example in (28a) and the one in (28b), both examples displaying direct objects that are not semantically selected by their respective verb. In (28b), which involves a bona fide resultative predicate, the presence of an overt Path (*to*) is mandatory. This is not the case in (28a), in contrast to what one would expect if the predicate in (28a) was resultative.[10]

(28) a. Brush a hole in(#to) one's coat.
      b. The children run themselves *in/(in)to exhaustion.   (Iwata 2020: 281)

A further piece of evidence against considering the locative PP in satellite-framed predicates of creation/consumption as containing a null Path comes from the observation that such a PP can also be headed by the preposition *at*, as shown in (29). Unlike *in*, *at* is only compatible with a non-directional reading and is in complementary distribution with *to*. This strongly suggests that there is no null Path in the locative PPs found in the examples considered in this study.

---

[10]Arguably, a literal interpretation of the predicate in (28a) could be considered grammatical with the presence of *to*, but pragmatically aberrant, as the predicate would be interpreted as roughly meaning 'move a hole to the inside of one's coat using a brush-like object / in a brush-like manner' (Jaume Mateu, p.c.).

(29)   a.   They removed the coriaceous bracteoles wrapped outside of the corolla, bit a hole at the base of the corolla where the nectarines are located, and lapped up all the nectar in each flower.   (Web)

   b.   To really make it resemble a tea bag, Murphy punched a hole at the top, then added a length of twine and a "tag".   (COCA)

This said, that the locative PP can be omitted in (25a) but not in (25b) is not necessarily due to grammatical reasons. Other factors, e.g., conceptual/pragmatic ones, might be involved. Note that only (25b) involves a direct object which is not semantically selected by the verb. *Brush* is a verb of surface contact, and it typically appears with direct objects denoting the surface that is brushed. It can then be expected that the *v*P in (25b) requires additional contextual information in order to be interpreted under a creation reading. In the absence of the spatial PP *in the coat*, the default inferable reading would be the pragmatically aberrant (not ungrammatical, in my view) one in which *a hole* is a selected object of *brush* (that is to say, it is an existing entity that undergoes an event of *brushing*). Such a reading disappears when the locative PP is added, as the PP introduces the semantic argument of the verb (i.e. the surface which is brushed, e.g., *her coat*), favoring the interpretation of the direct object *a hole* as an effected object thanks to the additional context. Further notice, in this respect, that locative PPs do not always appear in predicates of this type. For instance, no locative PP appears in the complex creation predicates in (21), (23) and (24), nor in the complex consumption predicates in (18) and (19). I suggest that in these predicates, the intended creation/consumption reading arises based on world knowledge/pragmatic considerations regarding the scene denoted by the event which are clear enough without the necessity of additional contextual information.[11]

2.2 Method and results

The examples in (6) to (24) were presented to the speakers in a randomized order. Translations, glosses, and grammaticality judgments were collected by consulting one linguist native speaker per language.[12] For each of the examples tested,

---

[11]This is in contrast to resultative predicates like (28b), where the licensing of a direct object that is not semantically selected by the verb always requires the presence of a phrase (e.g., a result PP) acting as a secondary predicate. Such a contrast can be taken to reflect the different status of the PPs appearing in complex creation/consumption predicates and the result PPs appearing in resultative predicates with non-selected objects, the former being adjuncts while the latter are arguments of the predicate.

[12]One exception is the native speaker of Ukrainian, who is not a linguist but who is a proficient speaker of English.

it was ensured that the intended (creation/consumption) meaning of the predicate was clear to the speakers before soliciting a grammaticality judgment. Two caveats were further considered in gathering judgments from the native speakers of the Slavic languages selected. First, considering that, as I will discuss in Section 4, perfective aspect in many Slavic languages is achieved through prefixes which have been argued to play a role in the event domain and interfere with the data being analyzed, the English examples were presented in the imperfective aspect when soliciting corresponding translations from the native speakers of the Slavic languages tested. For instance, the availability of the English example in (6) was checked in Slavic languages using the imperfective construction *John was singing a song*. Additionally, the speakers were asked to provide translations involving unprefixed verbs only. As a second caveat, when possible, the availability of a transitive non-creation use of those verbs which gave rise to ungrammatical translations in the languages tested was checked for each language, in order to exclude possible cases of ungrammaticality due to unrelated lexical restrictions on the transitivity of the verbs involved.[13]

The results obtained are graphically summarized in Table 1, Table 2, and Table 3 for satellite-framed languages, verb-framed languages, and Slavic languages, respectively.[14] The data collected are provided in the Appendix.

Overall, the native speakers of the satellite-framed languages tested accepted a literal translation for the vast majority of the complex creation/consumption predicates provided from English (Table 1), consistently with Talmy's typology.[15]

---

[13]Such a non-creation use pertains to transitive predicates where the direct object is understood as a pre-existing entity which undergoes the action named by the verb, and is not created or consumed during the event. Compare, for instance, (12) with *Maria carved the wood* (Folli & Harley 2020: 439), where the direct object pre-exists the carving event and undergoes the change of state specified by the verb.

[14]In the tables, empty slots correspond to cases where a direct translation of the English verb is not available in the target language. For reasons of space, the languages examined are identified in the tables using the ISO 639-2/B standardized nomenclature (US Library of Congress).

[15]I assume that Mandarin Chinese is a standard satellite-framed language of the English type. Acedo-Matellán (2016) argues that some varieties of Chinese are weak satellite-framed because the satellite-framed constructions they display present the Path and the co-event components as univerbated in a sort of V-V compound (see also Fan 2014). However, the idea that the Path and the co-event components in Chinese resultatives form a complex head is disputed. For instance, Wang (2010) presents evidence of phrasal elements that may intervene between the two members of the V-V compound in Chinese resultatives. We can see this in (i), where the complex negation *bu tai* 'not too' disrupts the adjacency between *da* 'hit' and *si* 'die'.

(i)   Wo da bu   tai si   na   zhi zhanglang.
      I   hit NEG too die that CL   cockroach
      'I can hardly hit the cockroach to death.'   (Chinese; Wang 2010: 38)

Table 1: Creation/consumption predicates in satellite-framed languages

Example	Dut	Ger	Chi	Hun
(6) John sang a song	✓	✓	✓	✓
(7) They danced a Sligo jig	✓	✓	✓	✓
(8) Ariel ate the mango	✓	✓	✓	✓
(9) He dug a hole in the ground	✓	✓	✓	✓
(10) She wove the tablecloth	✓	✓	✓	✓
(11) Marco painted a sky	✓	✓	✓	✓
(12) Maria carved a doll	✓	✓	✓	✓
(13) She burned a hole in her coat	✓	✓	✓	✓
(14) He scratched a hole in the ground	✓	✓	✓	✓
(15) She punctured a wound in her finger	✓	✓	✓	★
(16) She cut a wound in her foot	✓	✓	✓	✓
(17) She bit a hole in the bag	✓	✓	✓	✓
(18) The adventurer walked the trail	✓	✓	✓	??
(19) The adventurer swam the channel	★	★	★	★
(20) Deanne kicked a hole in the wall	✓	✓	✓	✓
(21) She magicked a cursor	??		✓	✓
(22) She brushed a hole in her coat	✓	✓	✓	✓
(23) John smiled his thanks	★	★	★	★
(24) Elna frowned her discomfort	★			★

Table 2: Creation/consumption predicates in verb-framed languages

Example	Ita	Cat	Spa	Baq	Gre
(6) John sang a song	✓	✓	✓	✓	✓
(7) They danced a Sligo jig	✓	✓	✓	✓	✓
(8) Ariel ate the mango	✓	✓	✓	✓	✓
(9) He dug a hole in the ground	✓	✓	✓	✓	✓
(10) She wove the tablecloth	✓	✓	✓	✓	✓
(11) Marco painted a sky	✓	✓	✓	✓	✓
(12) Maria carved a doll	✓	✓	✓	✓	✓
(13) She burned a hole in her coat	★	★	★	✓	★
(14) He scratched a hole in the ground	★	★	★	✓	✓
(15) She punctured a wound in her finger	??	★	★	??	★
(16) She cut a wound in her foot	★	★	★	??	★
(17) She bit a hole in the bag	★	★	★	?	★
(18) The adventurer walked the trail	★	★	?	✓	✓
(19) The adventurer swam the channel	★	★	?		✓
(20) Deanne kicked a hole in the wall	★		★	★	★
(21) She magicked a cursor					★
(22) She brushed a hole in her coat	★	★	★	★	★
(23) John smiled his thanks	★ ?	★	?		
(24) Elna frowned her discomfort	★	★	★		★

Table 3: Creation/consumption predicates in Slavic languages (imperfective, unprefixed predicates)

Example	Rus	Ukr	Pol	Slo	Ser	Hrv
(6) John sang a song	✓	✓	✓	✓	✓	✓
(7) They danced a Sligo jig	✓	✓	✓	✓	✓	✓
(8) Ariel ate the mango	✓	✓	✓	✓	✓	✓
(9) He dug a hole in the ground	✓	✓	✓	✓	✓	✓
(10) She wove the tablecloth	✓	✓	✓	✓	✓	✓
(11) Marco painted a sky	✓	✓	✓	✓	✓	✓
(12) Maria carved a doll	★	★	✓	?	✓	✓
(13) She burned a hole in her coat	★	✓	?	??	★	✓
(14) He scratched a hole in the ground	✓	✓	★	✓	??	★
(15) She punctured a wound in her finger	★	✓	★	★	??	★
(16) She cut a wound in her foot	★	★	★	★	★	★
(17) She bit a hole in the bag	★	★	★	★	?	★
(18) The adventurer walked the trail	★	★	★	★	??	★
(19) The adventurer swam the channel	★	★	★	★	??	★
(20) Deanne kicked a hole in the wall	★	★	★	★	★	★
(21) She magicked a cursor	★	★	★	★	★	✓
(22) She brushed a hole in her coat	★	★	★	✓	★	★
(23) John smiled his thanks						
(24) Elna frowned her discomfort						

The results obtained from the native speakers of the verb-framed languages tested are considerably different when it comes to predicates that are understood as involving the expression of a co-event by the verb (Table 2). A literal translation of the English examples gets progressively more difficult to obtain in the verb-framed languages as the predicates shift from a verb-framed strategy (the verb implying the creation/consumption of the object) to a satellite-framed strategy (the verb being understood as specifying a co-event of the main event of creation/consumption), in accordance with the typology.

As Table 3 makes clear, Slavic languages behave on a par with verb-framed languages in disallowing creation/consumption predicates where the meaning of the verb cannot be taken to imply the creation/consumption of the entity denoted by the object. The literal translations in (30) of the satellite-framed example in (22) (also in (5a)) in Russian, Ukrainian, and Polish illustrate this.

(30) a. * Ona   čes-a-l-a   dyrku   v pal'to.   (Russian)
         she.NOM brush.IPFV-TH-PST-AGR hole.ACC in coat.LOC

     b. * Vona   ter-l-a   dyrku   na kurtci.   (Ukrainian)
         she.NOM brush.IPFV-PST-AGR hole.ACC in coat.LOC

     c. * Ona   czes-a-ł-a   dziurę   w płaszczu.   (Polish)
         she.NOM brush.IPFV-TH-PST-AGR hole.ACC in coat.LOC

     Intended: 'She was brushing a hole in her coat.'

In such cases, a verb-framed construction displaying a verb whose meaning implies the creation/consumption of the direct object has to be used instead, the manner co-event being optionally expressed as an adjunct.[16]

---

[16]The results obtained further warn against making generalizations about the typological behavior of a language based on individual examples. For instance, the example in (12) seems to be generally available in the verb-framed languages examined, but it does not fare well in Slavic languages such as Russian, Ukrainian, and Slovak. Instead, the example in (14) presents a high degree of variation both in verb-framed languages and in weak satellite-framed Slavic languages, as it is accepted in half of the Slavic languages and in two of the five verb-framed languages examined. Additionally, none of the native speakers of the satellite-framed languages checked seems to accept the example in (19), even though they accept the similar example in (18) and even though (19) is accepted by the native speaker of verb-framed Greek. It is also worth noticing that the examples in (23) and (24), despite being well-formed in English, do not fare well in any of the other satellite-framed languages tested according to the native speakers consulted. Arguably, some level of idiomaticity is present in these two constructions of English, which is not shared by the speakers of the other satellite-framed languages tested. Further similar irregularities are detected, which nonetheless do not affect the emergence of clear trends consistent with the predictions following from Talmy's typology.

(31) a. Ona   del-a-l-a   dyrku   v pal'to   ščëtkoj.   (Rus)
          she.NOM make.IPFV-TH-PST-AGR hole.ACC in coat.LOC brush.INS

      b. Vona   rob-y-l-a   dyrku   na kurtci ščitkoju.
          she.NOM make.IPFV-TH-PST-AGR hole.ACC in coat.LOC brush.INS
          (Ukr)

      c. Ona   rob-i-ł-a   dziurę   w płaszczu szczotką.   (Pol)
          she.NOM make.IPFV-TH-PST-AGR hole.ACC in coat.LOC brush.INS

      'She was making a hole in her coat with a brush.'

In the next section I propose a formal account of the patterns observed in terms of a PF requirement holding of the functional head *v* involved in verbal predicates. The requirement is argued to affect both verb-framed and weak satellite-framed languages, explaining the uniformity of results observed in these languages.

3 A morphophonological account of Talmy's typology

3.1 A syntactic approach to argument structure

I adopt a neo-constructionist view of argument structure along the lines of Mateu & Acedo-Matellán (2012), according to which argument structure is conceived of as consisting of the relations established between a head and its arguments (i.e. its specifier and complement) in syntax. A fundamental distinction is drawn between functional heads, which are abstract relational elements that are necessary for the building of syntactic structures, and roots, regarded as units of conceptual content that provide real world details to syntactic predicates and are devoid of grammatically relevant information (Mateu 2002; Borer 2005; Acedo-Matellán 2010; 2016, among others).

In this approach, satellite-framed constructions are understood as involving the conflation, i.e. e(xternal)-merge (Haugen 2009), of a root with a phonologically null verbal head *v*, whose complement receives a morphological realization independently of the verb. The root conflated with *v* is understood as specifying a co-event of the main event arising from the predicate (Embick 2004; Harley 2005; Mateu & Acedo-Matellán 2012; Ausensi & Bigolin 2023, among others). In the case of resultative (change-of-state/location) predicates, such as (32) (whose syntactic structure is illustrated in Figure 1), *v* takes a small clause as complement (PredP in Figure 1), where the undergoer of the transition and the final state/location are introduced (Hoekstra 1988).[17]

---

[17]In the structures, I represent roots with small capitals, following Acedo-Matellán (2016).
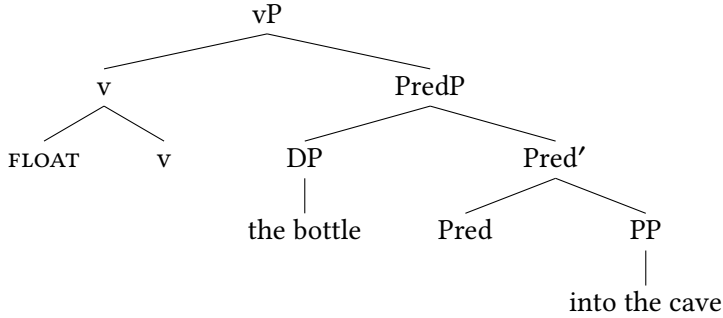
Figure 1: Syntactic structure of (32)

(32)   The bottle floated into the cave.   (Talmy 2000: 227)

Verb-framed languages are different from satellite-framed languages in that they never show the conflation pattern depicted in Figure 1 (Mateu 2012). In verb-framed languages, the predicative complement of the small clause always forms a unit with the *v* head, whereby the only resultative predicates attested are those formed via incorporation (Mateu 2002; 2017; Mateu & Rigau 2002; Folli & Harley 2020, among others).[18] The syntactic argument structure in Figure 2, relative to the Spanish verb-framed change-of-location predicate in (33), illustrates this.

(33)   La botella entró (flotando) a la cueva.
        'The bottle entered the cave (floating).'   (Talmy 2000: 227)

As for creation/consumption predicates, these are argued to involve an unergative configuration (à la Hale & Keyser 1993) consisting of a *v* head that takes as its complement either a root, which subsequently incorporates into it (the overt object emerging as a hyponym of the verb; Hale & Keyser 1997; 2002), or an independent DP. In the latter case, *v* may either appear as an overt light verb (e.g., *make*, as in the Spanish example in (5b)) or conflate with another root, giving rise to the complex creation/consumption predicates that are peculiar to the satellite-framed languages (Mateu 2012). The root incorporation pattern, corresponding to predicates of the type in (1) (see, e.g., (1a), repeated in (34)), is represented in

---

[18]Following Hale & Keyser (2002) and Mateu & Rigau (2002, 2010), I consider overt PPs expressing the final location of change-of-location events in verb-framed predicates (e.g., *a la cueva* in (33)) as hyponymous arguments that further specify the result provided by the root that incorporates into *v*. In the syntactic structures, hyponymous arguments are omitted for ease of exposition. For discussion of possible syntactic representations of hyponymous arguments, see Hale & Keyser (1997, 2002); Mateu (2008); Haugen (2009); Gallego (2012); Real-Puigdollers (2013), among others.
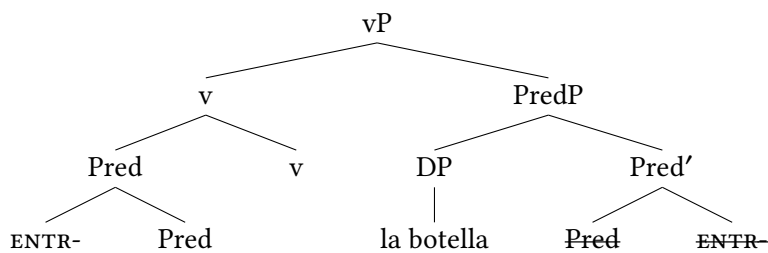
Figure 2: Syntactic structure of (33)

Figure 3. The pattern involving conflation is shown in Figure 4, which represents the syntactic structure of (5a) (repeated in (35)).[19]

(34)   He dug a hole in his garden.   (Washio 1997: 46)
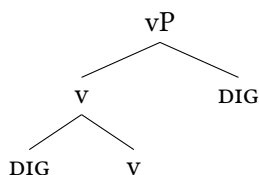

Figure 3: Syntactic structure of (34)

(35)   She brushed a hole in her coat.
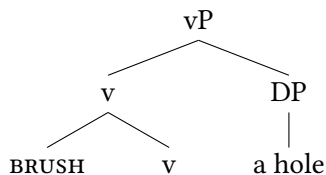   (Mateu & Rigau 2002: 213, based on Levin & Rapoport 1988)

Figure 4: Syntactic structure of (35)

---

[19]The spatial PPs in (34) and (35) are treated as *v*P-external adjuncts (see also footnote 11) and are omitted from the syntactic representations for ease of exposition. For the same reason I omit the representation of the external argument, which, following considerations in Marantz (1984); Kratzer (1996); Pylkkänen (2008), among others, I assume to be introduced by a functional head Voice merged on top of the *v*P.

At first sight, the presence vs. absence of the operation conflating a root with *v* in the syntax of a given language might seem to successfully account for the language's behavior with respect to Talmy's typology. However, there are at least two reasons, one theoretical and one empirical, why the availability of this syntactic operation in a given language cannot be taken as such as an effective way of explaining the typology. On the theoretical side, as noted in Folli & Harley (2020), parameterizing the availability of a specific syntactic operation comes at the cost of giving up on the basic minimalist assumption that variation is not located in narrow syntax. On the empirical side, the results presented in Section 2 show that correlating Talmy's typology with the presence vs. absence of the syntactic operation conflating a root with *v* leads to a wrong prediction when it comes to the possibility of licensing complex creation/consumption predicates in weak satellite-framed languages like Slavic languages (see Table 3).

In what follows, I propose an account of Talmy's typology which locates the source of the cross-linguistic variation at the PF level, understanding it in terms of differing morphophonological realization conditions of individual functional and lexical items. Not only does such an account seem to make the correct predictions with respect to the relevant patterns of cross-linguistic variation, it also provides a solution to the conundrum whereby verb-framed languages appear to consistently lack a structure-building operation (*viz.* the conflation of a root with *v*) that is instead available in satellite-framed languages.

3.2 A PF requirement on the *v* head in verb-framed languages

I endorse a view of cross-linguistic variation as primarily consisting in differing morphophonological realization conditions of functional heads (Acedo-Matellán 2016; Mateu 2017, among others). In order to account for the variation observed in relation to Talmy's typology, I posit that the *v* head in verb-framed languages is associated with a PF requirement which imposes the incorporation of *v*'s complement into *v* when *v* is phonologically null.[20]

(36)   *Verb-framed languages' PF requirement*:
       A phonologically null *v* must incorporate its complement.

---

[20]The requirement in (36) may ultimately be understood as an instance of Arregi & Pietraszko's (2021) "Generalized Head Movement" (GenHM) operation. This operation is captured by Arregi & Pietraszko (2021) by means of a feature [hm] on syntactic heads which, when present, requires them to form a single morphological word with the closest head of their complement. Although Arregi & Pietraszko (2021) formalize GenHM as a syntactic operation, they leave open the possibility that such an operation is carried out in the PF branch of the derivation (see Kwapiszewski 2022 for a PF implementation of GenHM). I am grateful to Víctor Acedo-Matellán for drawing my attention to the work of Arregi & Pietraszko (2021).

The requirement in (36) predicts that the typological patterns noted by Talmy hold regardless of whether a result component is involved (as in the case of change-of-location/state predicates) or not (as in creation/consumption predicates). This is so because the *v* head is found in both resultative predicates and creation/consumption predicates, as discussed in Section 3.1.[21] In Figure 5 and Figure 6, illustrating the syntactic structures of the verb-framed resultative predicate in (33) and the verb-framed creation predicate in (34), respectively, I represent the PF requirement on the *v* head by means of an index *[i]* which is deleted when the requirement is satisfied.
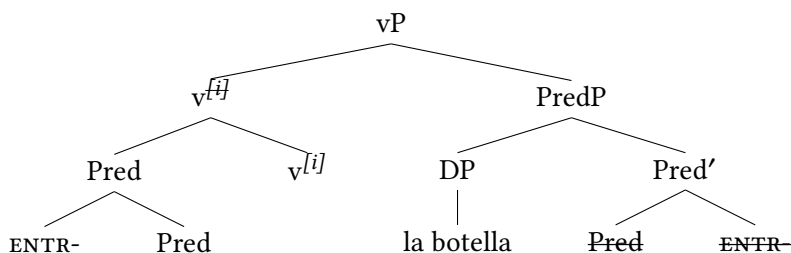
Figure 5: Syntactic structure of (33) (with a visual representation of the PF requirement (*[i]*) on *v*)
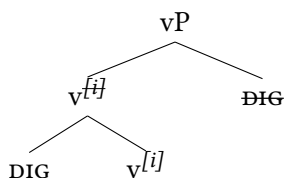
Figure 6: Syntactic structure of (34) (with a visual representation of the PF requirement (*[i]*) on *v*)

In the present account, the absence of the operation conflating a root with *v* in verb-framed languages arises as a by-product of *v*'s PF requirement. No parameterization of specific syntactic operations thus needs to be invoked. Verb-framed languages give the impression of lacking the operation conflating a root with *v*, because the syntactic configuration produced by such an operation is incompatible with the morphophonological context needed for the incorporation of *v*'s

---

[21]A reviewer wonders whether the PF requirement of verb-framed languages can be argued to apply to phonologically null functional heads in general in these languages. In the remainder of this presentation I continue to focus on the functional head involved in the argument structure of verbal predicates, leaving the exploration of this hypothesis to further research.

complement into *v* at PF in these languages. The syntactic representations in Figure 7 and Figure 8, corresponding to the Spanish ungrammatical satellite-framed resultative predicate in (37) and satellite-framed creation/consumption predicate in (38), respectively, illustrate this.[22] In the case of (37), ungrammaticality arises because neither the AP *limpia* 'clean' nor its root LIMP- can function as prefixes of verbs in Spanish, whereby the fulfillment of *v*'s PF requirement would give rise to an unpronounceable sequence of morphemes.[23] Similarly, the ungrammaticality of (38) is due to the DP complement of *v* (*su agradecimiento* 'his thanks') not being able to incorporate onto *v*, which leaves the PF requirement on *v* unsatisfied.[24]

(37)   * Él fregó   la   mesa limpia.
      he wipe.PST.AGR the table clean
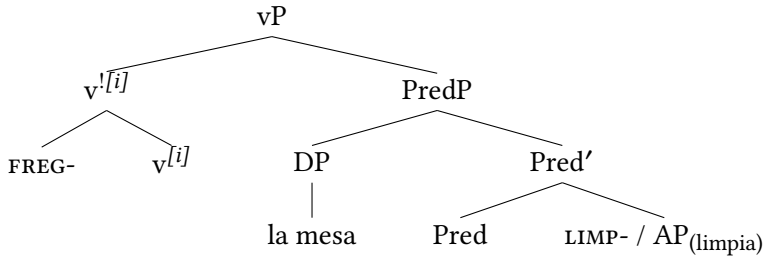   'He wiped the table clean.'   (Bigolin & Ausensi 2021: 519)

Figure 7: Syntactic structure of (37) (with a visual representation of the PF requirement (*[i]*) on *v*)

(38)   * Juan sonríe   su   agradecimiento.
      Juan smile.PRS.AGR POSS gratitude
      'Juan smiles his thanks.'   (Bigolin & Ausensi 2021: 527)

    I propose that the requirement in (36), found in verb-framed languages, is also responsible for the pattern illustrated in (4) concerning Slavic languages (and

---

[22](37) is grammatical in Spanish in the irrelevant readings involving a depictive or attributive interpretation of the AP (Jaume Mateu, p.c.).

[23]In Distributed Morphology terms, one could formalize the context of insertion of the Vocabulary Item associated with LIMP- as requiring that no roots intervene between LIMP- and *v*.

[24]See Martínez Vázquez (2014) for the claim that, to a certain extent, complex creation/consumption predicates can be found in verb-framed Spanish. Further see Bigolin & Ausensi (2021) for an analysis of the examples in Martínez Vázquez (2014) as involving a verb-framed strategy.
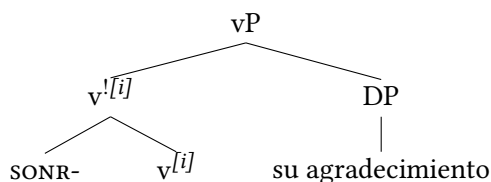
vP

v'$^{[i]}$ DP

SONR-   v$^{[i]}$ su agradecimiento

Figure 8: Syntactic structure of (38) (with a visual representation of the PF requirement (*[i]*) on *v*)

weak satellite-framed languages in general), in which the result component of resultative predicates with manner-denoting verbs must form a prosodic word with the verb (Talmy 2000; Acedo-Matellán 2010; 2016). That is to say, I propose that weak satellite-framed languages are fundamentally verb-framed languages. The reasoning goes as follows. As the account of (37) and (38) shows, in addition to PF requirements on functional heads of the type in (36), a relevant factor in determining the availability of specific constructions in a given language is whether or not the constructions in question can be spelled out consistent with the PF restrictions on the individual items that make up the lexical inventory of the language concerned. For instance, I have argued that in Spanish (and, more generally, in verb-framed languages), there are no instances of constructions involving the conflation of a root with *v* because such a syntactic configuration prevents the fulfillment of *v*'s requirement that it incorporate its complement, as the lexical inventory of Spanish does not contain morphemes capable of expressing a Talmian Path in the form of a verbal prefix. I argue that weak satellite-framed languages differ from standard verb-framed languages in the domain of resultative predicates in that their lexicon has result-denoting morphemes which can be realized as verbal prefixes (e.g., *iz-* 'out' in (4)). The prefixal nature of such morphemes may satisfy *v*'s requirement that it incorporates its complement, by concomitantly leaving open the possibility of conflating an independent root with *v*.[25] This gives rise to a satellite-framed behavior in the domain of resultative

---

[25]The pattern is discussed by Mateu (2017), who, however, continues to consider Slavic languages (and weak satellite-framed languages in general) as fundamentally satellite-framed languages. The parallelism between prefixed resultative predicates with manner-denoting verbs of Slavic languages and English satellite-framed resultative constructions is proposed in Spencer & Zaretskaya (1998) and Mateu (2008). See Snyder (2012) for the claim that Russian patterns with verb-framed languages with respect to The Compounding Parameter of Snyder (1995; 2001). Russian and Czech have been argued to be verb-framed languages also by Gehrke (2008), who relates their verb-framedness to their realizing accomplishment structures in the verb (also by means of prefixes). I come back to Gehrke's (2008) proposal in Section 4.2, where I elaborate on its potential relevance for the data dealt with in the present chapter.

predicates, as noted in Talmy (2000) and further discussed in Acedo-Matellán (2016). The structure in Figure (9), corresponding to the Russian predicate in (4) (repeated here as (39)), illustrates this.

(39) Ona   iz-pis-a-l-a   svoju ručku.
      she.NOM out-write-TH-PST-AGR POSS   pen.ACC
      'Her pen has run out of ink.' (Lit. 'She has written her pen out (of ink).')
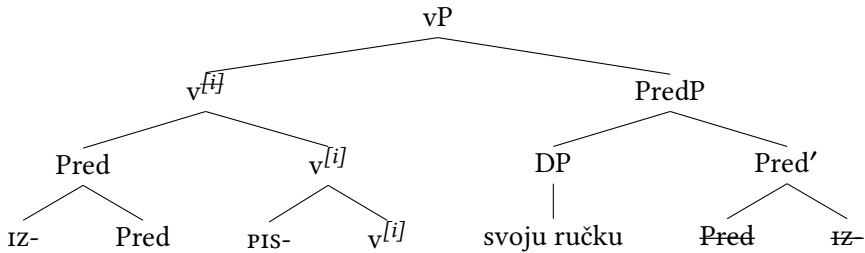   (Russian; Spencer & Zaretskaya 1998: 17)

Figure 9: Syntactic structure of (39) (with a visual representation of the PF requirement (*[i]*) on *v*)

From the hypothesis that weak satellite-framed languages are actually verb-framed languages (in the sense of (36)), it also follows that such languages should display a clear verb-framed behavior in the domain of creation/consumption predicates. No prefixal morpheme capable of referring to the object of creation/consumption predicates is present in the lexicon of these languages, whereby only creation/consumption predicates that involve the incorporation of a root into *v* can be licensed, in addition to predicates involving overt light verbs such as *do* or *make* (e.g., (31)). See this in Figure 10, where the syntactic representation of the ungrammatical Russian predicate in (30a) (repeated in (40)) is provided.

(40) *Ona   čes-a-l-a   dyrku   v pal'to.   (Russian)
       she.NOM brush.IPFV-TH-PST-AGR hole.ACC in coat.LOC
      'She was brushing a hole in her coat.'

The present account provides a solution to the minimalist conundrum whereby verb-framed languages seem to lack a structure-building operation (that of conflating a root with the *v* head) which is instead available in satellite-framed languages (see discussion in Folli & Harley 2020). In present terms, the resultative predicates with manner-denoting verbs and a prefixal result found in weak satellite-framed languages like Slavic languages are precisely to be regarded as constructions where a root is conflated with *v* in a verb-framed system.
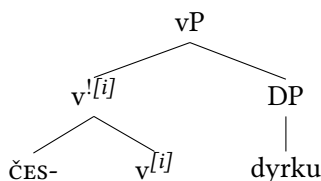
Figure 10: Syntactic structure of (40) (with a visual representation of the PF requirement (*[i]*) on *v*)

3.3 A comparison with some predecessors

Previous neo-constructionist approaches to Talmy's typology emphasize either that verb-framed languages always express the Path component in the main verb (Acedo-Matellán & Mateu 2013; Acedo-Matellán 2016; Folli & Harley 2020, among others) or that verb-framed languages lack predicates where, more generally, the verb expresses a manner co-event (Mateu 2012). The former approach runs into problems when considering that verb-framed languages and weak satellite-framed languages do not display complex creation/consumption predicates of the type displayed by satellite-framed languages, as nothing in this approach precludes the realization of such predicates – where no result component is involved – in these languages. Put differently, complex creation/consumption predicates are predicted to be universally available by this approach, contrary to facts.[26] The latter approach correctly predicts the unavailability of complex creation/consumption predicates in verb-framed languages, but it also predicts that weak satellite-framed languages should behave on a par with standard satellite-framed languages in allowing complex creation/consumption predicates. Furthermore, as discussed in Section 3.1, the generalization provided by this ap-

---

[26]Aware of this prediction, Folli & Harley (2020) argue that complex creation/consumption predicates indeed do not give rise to cross-linguistic variation related to Talmy's typology and are generally available in verb-framed languages. This fact would then constitute the empirical proof that the expression of a co-event in the main verb is a universally available linguistic process. Specifically, Folli & Harley (2020) note that creation/consumption predicates such as (10), (11), and (12) are licensed both in satellite-framed English and in verb-framed Italian, and they assume that these predicates involve the expression of a manner co-event in the verb, similar to what is observed in satellite-framed resultative predicates. However, as shown in Table 2, that these specific examples do not give rise to significant cross-linguistic variation cannot be taken to conclude that no typological variation exists in the domain of creation/consumption predicates. Namely, the examples in Folli & Harley (2020) can be taken to involve verbs whose conceptual meaning implies the creation of the direct object, which in turn is interpreted as a hyponym of the verb (in the sense of Hale & Keyser 1997; 2002). As such, they can be argued to involve the verb-framed incorporation pattern exemplified in Figure 3, whereby they are allowed in verb-framed Italian.

proach can only be taken as a descriptive one, as it cannot itself be considered explanatory without entailing a conception of syntax as a locus of parametric variation.

4 The role of perfectivizers

4.1 Internal prefixes and events of creation/consumption

In Slavic languages, the contrast between the imperfective and the perfective aspectual viewpoints is typically achieved by means of verbal prefixation and suffixation. In a standard case, basic verbal stems have an imperfective reading, which is turned perfective via the addition of a prefix. The Russian examples in (41) illustrate this.

(41) a. My   pis-a-l-i   pis'mo.
          we.NOM write.IPFV-TH-PST-AGR letter.ACC
          'We were writing a letter.'   (Russian; Smith 1991: 302)

      b. On   na-pis-a-l   pis'mo.
          he.NOM PFV-write-TH-PST   letter.ACC
          'He wrote a letter.'   (Russian; Smith 1991: 301)

Normally, the perfective prefix comes from the same inventory of morphemes which can provide the Talmian Path component in resultative predicates. Indeed, it has been argued that prefixes of this type – hereafter referred to as "internal" prefixes – denote the incorporation of a non-referential result into the verb, in a resultative structure (Ramchand & Svenonius 2002; Gehrke 2008; Acedo-Matellán 2016; Kwapiszewski 2022, among others).[27]

In the present framework, Slavic predicates perfectivized via internal prefixes (such as the Russian one in (41b)) are thus attributed the syntactic structure in Figure 11. I assume that predicates depicting events of creation/consumption made

---

[27]Internal (or "lexical") prefixes are contrasted with external (or "superlexical") ones. The distinction is motivated by a series of factors which point toward the idea that internal prefixes are merged inside the vP (hence the name), while external prefixes are merged higher in the functional spine of the clause. For discussion of the distinction between internal and external verbal prefixes in Slavic languages, see Babko-Malaya (1999); Romanova (2004); Svenonius (2004); Borik (2006); Arsenijević (2006, 2007); Gehrke (2008); Žaucer (2009); Łazorczyk (2010); Tatevosov (2011); Milosavljević (2022); Kwapiszewski (2022), among many others. The classification of Slavic prefixes has also been argued to be more nuanced than the traditional bi-partite division found in the literature. For instance, Tatevosov (2008) argued that in Russian there exists a class of prefixes (e.g., *do-* and *pere-*) that exhibit an intermediate behavior between internal and external prefixes. Since the examples from Russian collected in this study do not involve such prefixes, I do not pursue this issue further here.

vP

v   PredP

Pred   v   DP   Pred′

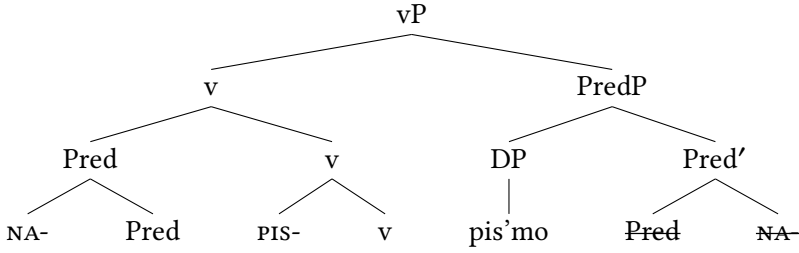NA-   Pred   PIS-   v   pis'mo   ~~Pred~~   ~~NA-~~

Figure 11: Syntactic structure of (41b)

perfective via internal prefixes consistently involve the argument structure that is found in resultative (change of state/location) predicates, the direct object being interpreted as a created or consumed entity due to pragmatic factors arising from the conceptual interpretation of the construction. Insofar as these predicates involve the incorporation of *v*'s complement into *v*, as shown in Figure 11, they are predicted to be possible in Slavic languages in the same way as resultative predicates with manner-denoting verbs are, the incorporation of the prefix fulfilling the verb-framed requirement of the language as understood in (36). In what follows, I present the results of a study exploring the validity of such a prediction.

In order to verify the prediction, I have conducted the same test run for bona fide creation/consumption predicates which was described in Section 2. This time, however, the English examples have been left in their non-progressive form, to check whether the presence of the perfective prefixes in their Slavic counterparts affects the grammaticality of their literal translation in the Slavic languages. The results obtained, summarized in Table 4, show that Slavic languages clearly behave on a par with satellite-framed languages (cf. Table 1) when a perfective prefix is present, confirming the prediction.[28] The grammatical renditions of the English example in (22) in Russian, Ukrainian, and Polish illustrate this (see (42)). The structure for the Russian example in (42a), which is understood to hold also for the rest of the data, is provided in Figure 12.

(42) a. Ona   pro-čes-a-l-a   dyrku   v pal'to.   (Russian)
          she.NOM pfv-brush-th-pst-agr hole.ACC in coat.LOC

      b. Vona   pro-ter-l-a   dyrku   na kurtci.   (Ukrainian)
          she.NOM pfv-brush-pst-agr hole.ACC in coat.LOC

---

[28]The native speakers of Serbian and Croatian seem more conservative than the native speakers of the other Slavic languages tested in disallowing a creation/consumption reading for several of the predicates involved. At the moment, I am agnostic as to why the pattern displayed by Serbian and Croatian in this test differs in this way from that of the other Slavic languages.

c. Ona   wy-czes-a-ł-a   dziurę   w płaszczu.   (Polish)
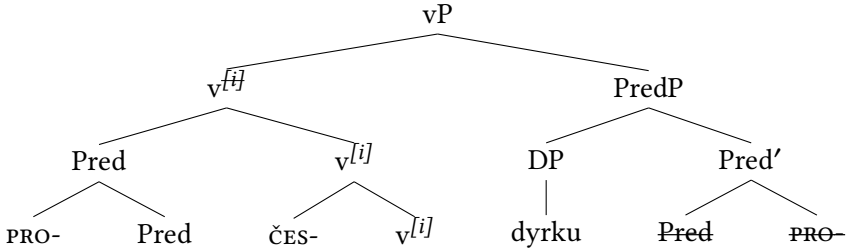she.NOM PFV-brush-TH-PST-AGR hole.ACC in coat.LOC
'She brushed a hole in her coat.'

Figure 12: Syntactic structure of (42a) (with a visual representation of the PF requirement (*[i]*) on *v*)

The contrast in acceptability between predicates with unprefixed verbs and predicates with prefixed verbs in the expression of complex events of creation/

Table 4: Perfective predicates with creation/consumption reading in Slavic languages (prefixed predicates)

Example	Rus	Ukr	Pol	Slo	Ser	Hrv
(6) John sang a song	✓	✓	✓	✓	✓	✓
(7) They danced a Sligo jig	✓	✓	✓	✓	✓	✓
(8) Ariel ate the mango	✓	✓	✓	✓	✓	✓
(9) He dug a hole in the ground	✓	✓	✓	✓	✓	✓
(10) She wove the tablecloth	✓	✓	✓	✓	✓	✓
(11) Marco painted a sky	✓	✓	✓	✓	✓	✓
(12) Maria carved a doll	✓	✓	✓	✓	✓	✓
(13) She burned a hole in her coat	✓	✓	✓	✓	??	✓
(14) He scratched a hole in the ground	✓	✓	✓	✓	??	★
(15) She punctured a wound in her finger	✓	✓	✓	✓	??	★
(16) She cut a wound in her foot	✓	?	✓	✓	★	★
(17) She bit a hole in the bag	✓	✓	✓	✓	✓	★
(18) The adventurer walked the trail	✓	✓	✓	✓	✓	✓
(19) The adventurer swam the channel	✓	✓	✓	✓	✓	✓
(20) Deanne kicked a hole in the wall	✓	✓	✓	✓	??	★
(21) She magicked a cursor	✓	✓	✓	✓	??	✓
(22) She brushed a hole in her coat	✓	✓	✓	✓	?	★
(23) John smiled his thanks						
(24) Elna frowned her discomfort						

consumption in Slavic languages (compare Table 3 with Table 4) cannot be argued to depend on the aspectual shift from the imperfective reading of the former type of predicates to the perfective reading of the latter type of predicates. This is proved by the availability, for the examples that are ungrammatical in the imperfective reading provided by unprefixed verbs, of imperfective predicates obtained via secondary imperfectivization. Secondary imperfectivization is a strategy found in Slavic languages whereby a prefixed, perfective verb is turned into an imperfective reading by means, typically (although not necessarily), of a further process of affixation (Babko-Malaya 1999; Romanova 2004; Svenonius 2004; Kwapiszewski 2022, among others). In the examples under consideration, secondary imperfectivization gives rise to grammatical predicates also in those cases where an imperfective reading involving unprefixed verbs gives rise to ungrammaticality. This is illustrated in (43) with the Ukrainian translations of (20), which is unavailable in the imperfective unprefixed version (43a) but is grammatical both in the perfective prefixed version (43b) and in the imperfective prefixed version obtained via secondary imperfectivization (43c).

(43) a. * Din byv   dyru   u stini.   (Ukrainian)
          Din kick.IPFV.PST hole.ACC in wall.LOC
          Intended: 'Din was kicking a hole in the wall.'

      b. Din pro-byv   dyru   u stini.
          Din PFV-kick.PST hole.ACC in wall.LOC
          'Din kicked a hole in the wall.'

      c. Din pro-byv-av   dyru   u stini.
          Din PRO-kick.PST-IPFV hole.ACC in wall.LOC
          'Din was kicking a hole in the wall.'

These facts suggest that the predicate's grammaticality does not rely on the perfective reading, but on the presence of the prefix, which fulfills the verb-framed requirement of the language by incorporating into *v* from its complement.

4.2 Incrementality in complex predicates

The meaning contribution of internal prefixes in the licensing of complex predicates in Slavic languages warrants further investigation. For instance, Gehrke (2008) posits that complex predicates, in which the main verb denotes an activity, require an accomplishment event structure, which in satellite-framed resultative constructions is licensed by an incremental structure provided by a secondary predicate. She further argues that internal prefixes of Slavic languages derive accomplishment structures, and that in these languages (specifically, she refers to

Czech and Russian) accomplishment structures are realized in the verb complex, either by the verb itself or by an internal prefix. Gehrke's findings may offer an alternative explanation for why complex predicates of creation/consumption are grammatical in Slavic languages only when prefixed. Unprefixed complex predicates of creation/consumption might be infelicitous in Slavic languages due to the absence of an accomplishment structure within the verbal complex. This explanation rests on the assumption that complex predicates, cross-linguistically, require the presence of an accomplishment event structure. However, the idea that an accomplishment event structure is needed to license satellite-framed constructions is not undisputed. For instance, Folli & Harley (2006) discuss cases of satellite-framed predicates of English, in which a PP denoting an unbounded path appears as the secondary predicate, as in (44). Since the incremental structure associated with PPs of this kind does not have a culmination point, the overall predicate lacks an accomplishment event structure.[29]

(44)　a.　John waltzed Matilda around and around the room for hours.

　　　b.　John walked Mary along the river all afternoon.

　　　　　　　　　　　　　　　　　　　　　　　(Folli & Harley 2006: 125)

　　　c.　John walked Mary towards her car for 3 hours.

　　　　　　　　　　　　　　　　　　　　　　　(Folli & Harley 2006: 137)

Another explanation worth considering is that unprefixed complex creation/consumption predicates are not licensed in Slavic languages due to the absence of incrementality in these constructions. I argue that this explanation is not satisfactory either. In these languages, given the right context, predicates of creation/consumption can be telic even if unprefixed (see, e.g., Gehrke 2008: 179, fn. 41; Mehlig 2012), the incremental path structure being provided by the direct object (Rappaport Hovav 2008, 2014). For instance, Mehlig (2012) argues that such a reading of the direct object in predicates denoting events of creation/consumption is possible in Russian if the extent of the entities denoted by the object has been determined in advance (e.g., from the conversational context) and these entities are referred to in the relevant imperfective predicate by means of a demonstrative (e.g., *ėtot*/*tot* 'this/that'). This is illustrated by Mehlig (2012) with

---

[29]The temporal adverbials in the examples in (44) show that the PPs in these examples are not understood as referring to a bounded Path. I am not aware of studies concerned with the availability of such examples in Slavic languages. The unavailability of these examples in Polish (Wojciech Lewandowski, p.c.) and in Italian, however, points toward the idea that the satellite-framed/verb-framed division should not be (only) intended as a constraint in the expression of accomplishment structures in some languages and not in others.

examples like (45), where the consumption predicate *on est ėti dva banana* 'he is eating those two bananas' is successfully modified by the expression *Odin on uže s"el* 'He has already eaten one of them', which presupposes that the object has an incremental structure associated with it, because the two conditions listed above are satisfied (see the text preceding the consumption predicate in (45), where the amount of *bananas* involved in the eating event is predetermined, and see the presence of the demonstrative *ėti* 'these' in the consumption predicate).

(45) Segodnja utrom ja dal[PFV] Saše dva banana. V dannyj moment on est[IPFV] ėti dva banana. Odin on uže s"el[PFV].
'This morning I gave Sasha two bananas. At the moment he is eating those two bananas. He has already eaten one of them.'

(Russian; Mehlig 2012: 216)

According to the hypothesis under discussion, the complex creation/consumption predicates that gave rise to ungrammaticality in Russian (see Table 3) should become acceptable if the contextual conditions identified in Mehlig (2012) are met, as the predicates would then be given an incremental structure by the direct object. However, the prediction is not borne out. The same results as those listed in Table 3 are obtained in Russian if the contextual conditions discussed in Mehlig (2012) are met, as illustrated in (46) with an example based on the predicate in (12) (Dària Serés, p.c.).

(46) Segodnja utrom Deanne zakazali s-delat'
today morning Deanne.DAT commission.PFV.PST.PL PFV-make.INF
dve reznye kukly. V dannyj moment *ona režet ėti
two carved dolls.ACC in this moment she.NOM carve.IPFV.PRS these
dve kukly. Skoree vsego ona uže vy-rez-a-l-a odnu
two dolls.ACC probably she.NOM already PFV-carve-TH-PST-AGR one
iz nix.
of them
'This morning Deanne was commissioned to make two carved dolls. At the moment she is carving those two dolls. She has probably already carved one of them.'

Similar considerations apply in Slovak, which also seems to license a reading of the object as having an incremental structure associated with it under the conditions in Mehlig (2012) but does not allow complex creation/consumption predicates in such contexts (Natália Kolenčíková, p.c.). In the case of Serbian, modifying expressions equivalent to the Russian *Odin on uže s"el* 'He has already

eaten one of them' in (45) are compatible with predicates denoting events of consumption regardless of the contextual conditions in Mehlig (2012) (Predrag Kovačević, p.c.), yet complex creation/consumption predicates with unprefixed verbs are not licensed (see Table 3). The ungrammaticality of the predicate in (46) is accounted for by the account proposed in the present chapter; the predicate is not grammatical in Russian because it does not have a verbal prefix which fulfills the language's verb-framed requirement, by incorporating onto the *v* head.

In sum, the approach to Talmy's typology proposed in this chapter allows us to account for the kind of cross-linguistic variation related to the typology in the domain of predicates of creation/consumption regardless of whether an incremental structure is provided to the predicate by the verb, by a prefix of the verb or by a phrasal complement of the verb. In the next section, I propose that the present account of Slavic languages as verb-framed languages should be extended to Latin, which was argued to be a weak satellite-framed language, along with Slavic languages, by Acedo-Matellán (2010; 2016).

5 Were there complex creation/consumption predicates in Latin?

As observed in Talmy (2000), and extensively explored in Acedo-Matellán (2010; 2016), Latin behaves on a par with Slavic languages in regard to Talmy's typology, in that resultative predicates where the verb denotes a co-event are allowed as long as the Path is expressed by a verbal prefix. As such, Latin is predicted to allow complex creation/consumption predicates by previous neo-constructionist accounts (see Section 3.3). Some examples of alleged complex creation predicates from Latin are provided in Acedo-Matellán (2016) to prove this point. This goes against the prediction of the present account of Talmy's typology, according to which neither Latin nor Slavic languages, *qua* weak satellite-framed languages, should be able to license bona fide creation/consumption predicates in the absence of an incorporation process of *v*'s complement onto *v*. Indeed, the data discussed in Acedo-Matellán (2016) are surprising in light of the pattern displayed by Slavic languages in this respect. In this section, I argue that there is no clear reason for attributing a complex creation/consumption reading to the Latin examples provided in Acedo-Matellán (2016). Afterward, I present the results of a corpus search which point toward the conclusion that complex creation/consumption predicates are absent in Latin, in line with the prediction of the present approach.

The examples discussed in Acedo-Matellán (2016) are provided in (47) to (51).

(47) Qui   alteri   misceat   mulsum.
who.NOM another.DAT mix.SBJV.AGR honeyed_wine.ACC
'He who makes honeyed wine for someone else.'   (Latin; *Cic.* Fin. *2, 5, 17*)

(48) Vulnus   [...] quod   acu   punctum
wound.NOM   which.NOM needle.ABL puncture.PTCP.PFV.NOM
videretur.
seem.IPFV.SBJV.AGR
'A wound that seemed to have been punctured with a needle'.
(Latin; *Cic.* Mil. *65*)

(49) [Serpens] volubilibus   squamosos nexibus   orbes
snake.NOM looping.ABL.PL scaly.ACC.PL writhing.ABL.PL coil.ACC.PL
torquet.
twist.PRS.AGR
'The snake twists his scaly coils in looping writhings.'
(Latin; *Ov.* Met. *3, 41*)

(50) Viam   silice   sternendam   [...] locauerunt.
way.ACC flint-stone.ABL strew.PTCP.GRDV.ACC   establish.PRF.AGR
'They established that the way was to be paved with flint stone.'
(Latin; *Liv. 38, 28, 3*)

(51) Aeriam truncis   [...] cumulare pyram.
high.ACC log.ABL.PL   gather.INF pyre.ACC
'To build a high pyre out of logs.'   (Latin; *Stat.*, Teb. *6, 84*)

I suggest that most of these examples are compatible with a reading as either involving hyponymous objects or displaying resultative predicates of change of state, therefore adopting a verb-framed strategy. For instance, *pyram* 'pyre' in (51) could be interpreted as a hyponym of *cumulare* '(lit.) cumulate'. Indeed, a creation reading of this verb is also found in verb-framed Italian, as (52) shows.[30]

---

[30]An anonymous reviewer asks me to elaborate on the relevance of the Italian example in (52) for the conclusion that the Latin example in (51) is not a satellite-framed construction. Both the Latin and the Italian example refer to a creation event in which a 'cumulation' is formed. As Hale & Keyser (1997) noted, the conceptual content of the verb in predicates of this kind (in the cases at hand, *cumulare*, meaning 'cumulate', or 'gather') points non-referentially to the nature of the entity effected during the event (e.g., in (51) and (52), a 'cumulation' of some sort). The object, in turn, directly refers to such an effected entity. For instance, the predicate in (52) can be paraphrased as 'make a gathering that *consists of* experience'. Similarly, the predicate in (51) can be paraphrased as 'make a gathering that *consists of* a pyre'. For this reason, direct objects of this type have been referred to in the literature as "hyponymous arguments" of the verb. As discussed in Section 3, predicates of this kind have been argued to involve the incorporation

(52) [...] il primo dovrà   aver   cumulato   esperienza nella
        the first   must.FUT.AGR have.INF gather.PTCP.PFV experience in.the
      grande distribuzione, il   secondo sul   prodotto e   sul   contatto
      big   distribution   the second on.the product and on.the contact
      con i   grandi clienti.
      with the big   clients
      'The first one must have gathered experience in large-scale distribution,
      the second one on the product and in dealing with large clients.'

(CORIS[31])

As for (48), the availability of *puncture a wound* (cf. (15)) in weak satellite-framed Slavic languages and in verb-framed languages seems to be very limited, but the Ukrainian speaker fully accepts it (Table 3) and the Basque speaker considers it marginally acceptable (Table 2), suggesting that this predicate is not entirely precluded in these language types. Finally, I suggest that examples such as (49) and (50) can be compatible with a change-of-state reading of the direct object, which would imply the adoption of a verb-framed resultative structure. For instance, a snake can twist its coils also if the coils have been previously formed, e.g., by the position of the body prior to the *twisting*. Similarly, an existing road can be ordered to be covered with flint stone, supposing, for instance, that it was unpaved before. The proposed syntactic structure of the Latin example in (50), assuming a change-of-state reading of the predicate, is provided in Figure (13).
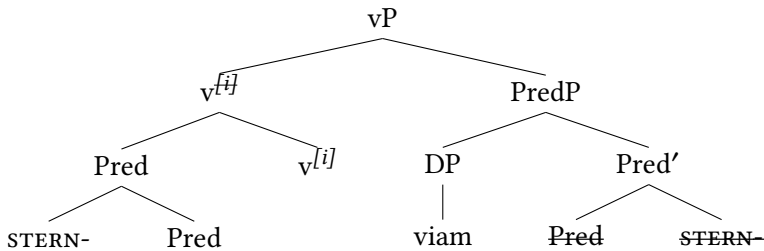

Figure 13: Syntactic structure of (50) (with a visual representation of the PF requirement (*[i]*) on *v*)

into *v* of a root e-merged as the complement of *v* (see Figure 3). Thus, they are expected to be well-formed in verb-framed languages. Assuming that Italian is a verb-framed language, that the construction in (51) can also be found in Italian, as (52) shows, provides additional evidence to the claim that such a construction is a verb-framed construction, whereby it does not constitute a counterexample to the proposal that Latin should be regarded as a verb-framed language.

[31]*Corpus di Riferimento dell'Italiano Scritto*, Università di Bologna.

In order to further substantiate the prediction that complex creation/consumption predicates could not be licensed in weak satellite-framed Latin, I carried out a corpus-based investigation checking the co-occurrence, in a creation reading, of verbs that can be associated with a manner interpretation with two direct objects that seem to be particularly productive in English complex creation predicates, namely *hole* (Lat. *foramen*) and *wound* (Lat. *vulnus*). The corpus used for Latin, comprising texts from the Early and Classical periods (up to A.D. 200), is the *Classical Latin Texts* by The Packard Humanities Institute.[32] The verbs selected, listed in (53), were taken from Acedo-Matellán (2016).

(53)   *amburo* 'burn', *caedo* 'cut, knock', *cremo* 'burn', *frico* 'rub', *rado* 'scrape', *tundo* 'beat', *uro* 'burn', *verro* 'sweep'

Importantly, the English verbs corresponding to the Latin ones in (53) can give rise to creation predicates with *hole* or *wound* as effected object, as (54) shows.

(54)   a.   A discharge of those energies burned a hole in his forehead and killed him.   (Ausensi & Bigolin 2023: 155)

   b.   [...] his words burned a wound inside her.   (Google Books)

   c.   Dad cut a hole in his chest and made me pull his heart out.   (COCA)

   d.   The Devil-Is-I pulled the knife he had used to cut a wound on his thumb and lunged forward at the leader of the twelve. (Google Books)

   e.   Weena knocked a hole in the wall.   (COCA)

   f.   But I scraped a hole in it so I could see.   (COCA)

   g.   [...] he scraped a wound on his nose that never cleared up.
                                                               (Google Books)

   h.   I erased again and again until I had rubbed a hole in the paper.
                                                                     (COCA)

   i.   [...] the mooring line has rubbed a wound in the willow bark.
                                                               (Google Books)

   j.   My 'beloved' boyfriend beat a hole in my roof and now it's awfully cold in there.   (COCA)

   k.   A sudden shift in the wind swept a hole in the blowing snow.
                                                               (Google Books)

---

[32]https://latin.packhum.org

The verbs in (53) were searched for by stem, while the objects were searched for in the nominative/accusative singular and plural forms.[33] None of the verb–complement combinations investigated provided relevant results. A verb-framed construction with *vulnus* 'wound' was found instead, as shown in (55).

(55)   Sed uulnera   facta   igne   dum sanescunt,
       But wound.NOM.PL make.PTCP.PFV.NOM.PL fire.ABL while heal.PRS.AGR
       defricare   bubula   urina   convenit.
       off_rub.INF bovine.ABL urine.ABL fit.AGR
       'But while the wounds made with fire are healing, it is appropriate to
       cleanse them with bovine urine.'   (*Col.*, De Re Rustica *6.7.4*)

I take this lack of evidence to tentatively suggest that Latin lacked complex creation/consumption predicates of the type found in satellite-framed languages, and needed to resort to verb-framed strategies in the domain of creation/consumption predicates in the same way as Slavic languages do. This is in line with the prediction, following from the present account, that complex creation/consumption predicates are unavailable in weak satellite-framed languages.

   Picking up the discussion in Section 4.2 about the possibility that unprefixed complex predicates of creation/consumption may be disallowed in Slavic languages due to their lack of an incremental structure, it is relevant to notice that direct objects could be associated with an incremental structure giving rise to telicity in the predicate in Latin. The compatibility of the consumption predicate in (56) with the time span adverbial *intra duas horas* 'within two hours' illustrates this.

(56)   [...] nitrosae   aut amarae   aquae   polenta
            nitrous.NOM.PL or bitter.NOM.PL water.NOM.PL cornmeal.ABL
       addita   mitigantur,   ut intra duas
       add.PTCP.PFV.ABL mitigate.IPFV.SBJV.PASS.AGR that within two.ACC
       horas   bibi   possint.
       hour.ACC.PL drink.INF.PASS can.IPFV.SBJV.AGR
       'Nitrous and bitter waters are softened with added cornmeal, so that they
       can be drunk within two hours.'   (*Plin.*, Nat. *24, 3, 4*)

Assuming, based on the discussion in this section, that complex creation/consumption predicates were not possible in Latin, such an absence cannot be attributed

---

[33]Being neuter, both *foramen* 'hole' and *vulnus* 'wound' appear as morphologically identical in their respective nominative and accusative forms.

to the predicate's lack of incrementality. The morphophonological account of Talmy's typology proposed in this chapter provides an alternative explanation of the phenomenon that is compatible with the observation that creation/consumption predicates could be telic in Latin without the presence of the prefix (see (56)).

6 Conclusions

I presented the preliminary results of a pilot study concerning the possibility of licensing complex creation/consumption predicates in Slavic languages. The results obtained were further compared with data gathered from native speakers of several satellite-framed languages and verb-framed languages. The study shows that Slavic languages, which are considered as fundamentally satellite-framed in the literature on Talmy's typology (Talmy 2000; Acedo-Matellán 2016), appear to behave on a par with verb-framed languages in disallowing creation/consumption predicates that involve a satellite-framed strategy.

Adopting a neo-constructionist perspective on argument structure, I have put forward a morphophonological approach to the variation related to Talmy's typology, understanding verb-framedness in terms of a morphophonological realization condition imposed at PF on the null v head involved in verbal predication. A null v is required to incorporate its complement in verb-framed languages. I have further argued that Slavic languages, and weak satellite-framed languages in general, should be regarded as fundamentally verb-framed languages, capturing the mandatory prefixation of the Path component in resultative predicates and the absence of complex predicates of creation/consumption in these languages as by-products of the verb-framed PF requirement on the v head.

With the present morphophonological account of Talmy's typology, I have additionally provided a solution to the minimalist conundrum whereby verb-framed languages seem to consistently lack the structure-building operation associated with the expression of a co-event in the verb, namely the operation of conflating a root with v. To the extent that the verb-framed PF requirement can be satisfied by means of prefixation, the compounding operation can indeed successfully take place in a verb-framed system, as shown by the availability of prefixed satellite-framed resultative predicates in weak satellite-framed languages.

Afterward, I have explored the prediction that a creation/consumption reading of predicates with manner-denoting verbs is available in Slavic languages when the predicate is perfectivized via internal prefixes, which have been argued to involve a resultative structure that receives a reading as involving an event of

creation/consumption on the basis of conceptual/pragmatic considerations. The data gathered from the native speakers of the Slavic languages tested confirmed the prediction.

Finally I have argued that Latin, as a weak satellite-framed language (Acedo-Matellán 2010; 2016), lacked complex creation/consumption predicates of the type found in bona fide satellite-framed languages in the same way as Slavic languages do. I have argued that this is the case based on the analysis of some alleged Latin complex creation/consumption predicates provided in Acedo-Matellán (2016), which have been shown to admit a reading either as involving a hyponymous object or as involving a resultative predicate of change of state. Afterward, I have presented the results of a corpus search supporting the prediction that complex creation/consumption predicates are not licensed in Latin. The results strengthen the general hypothesis that Latin and Slavic languages behave alike with respect to Talmy's typology (Acedo-Matellán 2016), meanwhile underpinning one of the main conclusion of the present account whereby weak satellite-framed languages should be considered as fundamentally verb-framed languages.

Abbreviations

ABL	ablative	NEG	negation
ACC	accusative	NOM	nominative
AGR	agreement	PASS	passive
CL	classifier	PFV	perfective
DAT	dative	PL	plural
FUT	future	POSS	possessive
GER	gerund	PRF	perfect
GRDV	gerundive	PRS	present
INF	infinitive	PST	past
INS	instrumental	PTCP	participle
IPFV	imperfective	SBJV	subjunctive
LOC	locative	SG	singular
M	masculine	TH	theme vowel

Acknowledgments

References

Acedo-Matellán, Víctor. 2010. *Argument structure and the syntax-morphology interface: A case study in Latin and other languages*. Barcelona: Universitat de Barcelona. (Doctoral dissertation).

Acedo-Matellán, Víctor. 2016. *The morphosyntax of transitions: A case study in Latin and other languages*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198733287.001.0001.

Acedo-Matellán, Víctor & Arkadiusz Kwapiszewski. 2021. Theory and facts in Talmy's typology: A reaction to Folli & Harley 2020. Seminar at Centre de Lingüística Teòrica. Universitat Autònoma de Barcelona, January 15.

Acedo-Matellán, Víctor & Jaume Mateu. 2013. Satellite-framed Latin vs. verb-framed Romance: A syntactic approach. *Probus* 25(2). 227–265. DOI: 10.1515/probus-2013-0008.

Alexiadou, Artemis. 2015. Syntax and the lexicon. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – Theory and analysis: An international handbook*, 1088–1127. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110363708-009.

Arregi, Karlos & Asia Pietraszko. 2021. The ups and downs of head displacement. *Linguistic Inquiry* 52(2). 241–289. DOI: 10.1162/ling_a_00377.

Arsenijević, Boban. 2006. *Inner aspect and telicity: The decompositional and the quantificational nature of eventualities at the syntax-semantics interface*. Utrecht: LOT Publications.

Arsenijević, Boban. 2007. Slavic verb prefixes are resultative. *Cahiers Chronos* 17. 197–213. DOI: 10.1163/9789401204446_013.

Ausensi, Josep & Alessandro Bigolin. 2023. On the argument structure realization of result verbs: A syntactic approach. *Acta Linguistica Academica* 70(1). 139–160. DOI: 10.1556/2062.2023.00567.

Babko-Malaya, Olga. 1999. *Zero morphology: A study of aspect, argument structure and case.* New Brunswick: Rutgers, The State University of New Jersey. (Doctoral dissertation).

Beavers, John, Beth Levin & Shiao Wei Tham. 2010. The typology of motion expressions revisited. *Journal of Linguistics* 46. 331–377. DOI: 10.1017/S0022226709990272.

Bigolin, Alessandro & Josep Ausensi. 2021. A new resultative construction in Spanish? A reply to Rodríguez Arrizabalaga. *Folia Linguistica* 55(2). 517–546. DOI: 10.1515/flin-2021-2078.

Borer, Hagit. 2005. *Structuring sense, vol. 2: The normal course of events.* Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199263929.001.0001.

Borik, Olga. 2006. *Aspect and reference time.* Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199291298.001.0001.

Condamines, Anne. 2013. Quand la passion autorise la transitivation d'un circonstanciel de lieu. *Journal of French Language Studies* 23(3). 335–355. DOI: 10.1017/S095926951300001X.

Davies, Mark. 2008–. The Corpus of Contemporary American English (COCA). Available online at https://www.english-corpora.org/coca/.

Embick, David. 2004. On the structure of resultative participles in English. *Linguistic Inquiry* 35(3). 355–392. DOI: 10.1162/0024389041402634.

Fan, Sheng-yang. 2014. *Argument structure in Mandarin Chinese: A lexical-syntactic perspective.* Madrid: Universidad Autónoma de Madrid. (Doctoral dissertation).

Folli, Raffaella & Heidi Harley. 2005. Flavors of *v.* In Paula Marie Kempchinsky & Roumyana Slabakova (eds.), *Aspectual inquiries,* 99–120. Dordrecht: Springer. DOI: 10.1007/1-4020-3033-9.

Folli, Raffaella & Heidi Harley. 2006. On the licensing of causatives of directed motion: Waltzing Matilda all over. *Folia Linguistica* 60(2). 121–155. DOI: 10.1111/j.1467-9582.2006.00135.x.

Folli, Raffaella & Heidi Harley. 2008. Teleology and animacy in external arguments. *Lingua* 118(2). 190–202. DOI: 10.1016/j.lingua.2007.02.004.

Folli, Raffaella & Heidi Harley. 2020. A head-movement approach to Talmy's typology. *Linguistic Inquiry* 51(3). 425–470. DOI: 10.1162/ling_a_00351.

Folli, Raffaella & Gillian Catriona Ramchand. 2005. Prepositions and results in Italian and English: An analysis from event decomposition. In Hank J. Verkuyl, Henriette de Swart & Angeliek van Hout (eds.), *Perspectives on aspect*, 81–105. Dordrecht: Springer. DOI: 10.1007/1-4020-3232-3.

Gallego, Ángel J. 2012. A note on cognate objects: Cognation as doubling. *Nordlyd* 39(1). 95–112. DOI: 10.7557/12.2289.

Gehrke, Berit. 2008. *Ps in motion: On the semantics and syntax of P elements and motion events*. Utrecht: LOT Publications.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago: The University of Chicago Press.

Hale, Kenneth & Samuel Jay Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, 53–109. Cambridge, MA: MIT Press.

Hale, Kenneth & Samuel Jay Keyser. 1997. The limits of argument structure. In Amaya Mendikoetxea & Myriam Uribe-Etxebarria (eds.), *Theoretical issues at the morphology-syntax interface*, 203–230. Bilbao: Universidad del País Vasco.

Hale, Kenneth & Samuel Jay Keyser. 2002. *Prolegomenon to a theory of argument structure*. Cambridge, MA: MIT Press.

Harley, Heidi. 2005. How do verbs get their names? Denominal verbs, manner incorporation, and the ontology of verb roots in English. In Nomi Erteschik-Shir & Tova Rapaport (eds.), *The syntax of aspect: Deriving thematic and aspectual interpretation*, 42–64. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199280445.001.0001.

Haugen, Jason D. 2009. Hyponymous objects and late insertion. *Lingua* 119(2). 242–262. DOI: 10.1016/j.lingua.2007.10.008.

Hoekstra, Teun. 1988. Small clause results. *Lingua* 74(2-3). 101–139. DOI: 10.1016/0024-3841(88)90056-3.

Iwata, Seizi. 2020. *English resultatives: A force-recipient account*. Amsterdam: John Benjamins. DOI: 10.1075/cal.26.

Kratzer, Angelika. 1996. Severing the external argument from its verb. In Laurie Ann Zaring & Johan Rooryck (eds.), *Phrase structure and the lexicon*, 109–137. Dordrecht: Kluwer.

Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, MA: MIT Press.

Kwapiszewski, Arkadiusz. 2022. *Aspect and event structure: The morphosyntax of Polish verbs from a cross-linguistic perspective*. Oxford: University of Oxford. (Doctoral dissertation).

Łazorczyk, Agnieszka A. 2010. *Decomposing Slavic aspect: The role of aspectual morphology in Polish and other Slavic languages*. Los Angeles: University of South California. (Doctoral dissertation).

Levin, Beth & Tova Rapoport. 1988. Lexical subordination. In Lynn MacLeod, Gary Larson & Diane Brentari (eds.), *Papers from the 24th annual regional meeting of the Chicago Linguistic Society*, 275–289. Chicago, IL: Chicago Linguistic Society.

Marantz, Alec. 1984. *On the nature of grammatical relations*. Cambridge, MA: MIT Press.

Martínez Vázquez, Montserrat. 2014. Reaction object constructions in English and Spanish. *ES: Revista de Filología Inglesa* 35. 193–217.

Mateu, Jaume. 2002. *Argument structure: Relational construal at the syntax-semantics interface*. Barcelona: Universitat Autònoma de Barcelona. (Doctoral dissertation).

Mateu, Jaume. 2003. Parametrizing verbal polysemy: The case of *bake* revisited. In Jonathon Cihlar, Amy Franklin, David Kaiser & Irene Kimbara (eds.), *Papers from the 39th annual meeting of the Chicago Linguistic Society*, 256–269. Chicago, IL: Chicago Linguistic Society.

Mateu, Jaume. 2008. On the l-syntax of directionality/resultativity: The case of Germanic preverbs. In Anna Asbury, Jakub Dotlačil, Berit Gehrke & Rick Nouwen (eds.), *Syntax and semantics of spatial P*, 221–250. Amsterdam: John Benjamins. DOI: 10.1075/la.120.

Mateu, Jaume. 2012. Conflation and incorporation processes in resultative constructions. In Violeta Demonte & Louise McNally (eds.), *Telicity, change, and state: A cross-categorial view of event structure*, 252–278. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199693498.001.0001.

Mateu, Jaume. 2017. Two types of locative alternation. In Albert Álvarez González & Ía Navarro (eds.), *Verb valency changes: Theoretical and typological perspectives*, 51–79. Amsterdam: John Benjamins. DOI: 10.1075/tsl.120.

Mateu, Jaume & Víctor Acedo-Matellán. 2012. The manner/result complementarity revisited: A syntactic approach. In María Cristina Cuervo & Yves Roberge (eds.), *The end of argument structure?*, 209–228. Bingley: Emerald.

Mateu, Jaume & Gemma Rigau. 2002. A minimalist account of conflation processes: Parametric variation at the lexicon-syntax interface. In Artemis Alexiadou (ed.), *Theoretical approaches to universals*, 211–236. Amsterdam: John Benjamins. DOI: 10.1075/la.49.

Mateu, Jaume & Gemma Rigau. 2010. Verb-particle constructions in Romance: A lexical-syntactic account. *Probus* 22. 241–269. DOI: 10.1515/prbs.2010.009.

Mehlig, Hans Robert. 2012. Hybrid predicates in Russian. *Journal of Slavic Linguistics* 20(2). 171–227.

Milosavljević, Stefan. 2022. The delimitative prefix *po-*, durative adverbials, and Slavic aspectual composition. In Annie Holtz, Iva Kovač & Rasmus Puggaard-Rode (eds.), *Proceedings of ConSOLE XXX*, 17–42. Leiden: Leiden University Centre for Linguistics.

Nikitina, Tatiana. 2008. Pragmatic factors and variation in the expression of spatial goals: The case of *into* vs. *in*. In Anna Asbury, Jakub Dotlačil, Berit Gehrke & Rick Nouwen (eds.), *Syntax and semantics of spatial P*, 175–195. Amsterdam: John Benjamins. DOI: 10.1075/la.120.

Pylkkänen, Liina. 2008. *Introducing arguments*. Cambridge, MA: MIT Press. DOI: 10.7551/mitpress/9780262162548.001.0001.

Ramchand, Gillian Catriona. 2008. *Verb meaning and the lexicon*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486319.

Ramchand, Gillian Catriona & Peter Svenonius. 2002. The lexical syntax and lexical semantics of the verb-particle construction. In Line Mikkelsen & Christopher Potts (eds.), *Proceedings of WCCFL 21*, 387–400. Somerville, MA: Cascadilla Press.

Rappaport Hovav, Malka. 2008. Lexicalized meaning and the internal structure of events. In Susan Rothstein (ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect*, 13–42. Amsterdam: John Benjamins. DOI: 10.1075/la.110.

Rappaport Hovav, Malka. 2014. Building scalar changes. In Artemis Alexiadou, Hagit Borer & Florian Schäfer (eds.), *The syntax of roots and the roots of syntax*, 259–281. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199665266.001.0001.

Rappaport Hovav, Malka & Beth Levin. 1998. Building verb meanings. In Miriam Butt & Wilhelm Geuder (eds.), *The projection of arguments: Lexical and compositional factors*, 97–134. Stanford, CA: CSLI Publications.

Rappaport Hovav, Malka & Beth Levin. 2010. Reflections on manner/result complementarity. In Malka Rappaport Hovav, Edit Doron & Ivy Sichel (eds.), *Lexical semantics, syntax, and event structure*, 21–38. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199544325.001.0001.

Real Academia Española. 2024. *Corpus del Español del Siglo XXI (CORPES)*. https://www.rae.es/corpes/.

Real-Puigdollers, Cristina. 2013. *Lexicalization by phase: The role of prepositions in argument structure and its cross-linguistic variation*. Barcelona: Universitat Autònoma de Barcelona. (Doctoral dissertation).

Romanova, Eugenia. 2004. Lexical vs. superlexical prefixes. *Nordlyd* 32(2). 255–278. DOI: 10.7557/12.69.

Smith, Carlota S. 1991. *The parameter of aspect.* Dordrecht: Kluwer. DOI: 10.1007/ 978-94-015-7911-7.

Snyder, William. 1995. *Language acquisition and language variation: The role of morphology.* Cambridge: MIT. (Doctoral dissertation).

Snyder, William. 2001. On the nature of syntactic variation: Evidence from complex predicates and complex word-formation. *Language* 77(2). 324–342. DOI: 10.1353/lan.2001.0108.

Snyder, William. 2012. Parameter theory and motion predicates. In Violeta Demonte & Louise McNally (eds.), *Telicity, change, and state: A cross-categorial view of event structure*, 279–299. Oxford: Oxford University Press. DOI: 10.1093/ acprof:oso/9780199693498.001.0001.

Spencer, Andrew & Marina Zaretskaya. 1998. Verb prefixation in Russian as lexical subordination. *Linguistics* 36. 1–39. DOI: 10.1515/ling.1998.36.1.1.

Svenonius, Peter. 2004. Slavic prefixes inside and outside VP. *Nordlyd* 32(2). 205– 253. DOI: 10.7557/12.68.

Talmy, Leonard. 2000. *Toward a cognitive semantics,* vol. 2: *Typology and process in concept structuring.* Cambridge, MA: MIT Press.

Tatevosov, Sergei. 2008. Intermediate prefixes in Russian. In *Proceedings of Formal Approaches to Slavic Linguistics 16*, 423–445. Ann Arbor, MI: Michigan Slavic Publications.

Tatevosov, Sergei. 2011. Severing perfectivity from the verb. *Scando-Slavica* 57. 216–244. DOI: 10.1080/00806765.2011.631782.

Tenny, Carol. 1994. *Aspectual roles and the syntax-semantics interface.* Dordrecht: Kluwer. DOI: 10.1007/978-94-011-1150-8.

Truswell, Robert. 2007. Extraction from adjuncts and the structure of events. *Lingua* 117(8). 1355–1377. DOI: 10.1016/j.lingua.2006.06.003.

Volpe, Mark. 2004. Affected object unergatives. *Snippets* 8(4). 12–13.

Wang, Chyan-an Arthur. 2010. *The microparametric syntax of resultatives in Chinese language.* New York: New York University. (Doctoral dissertation).

Washio, Ryuichi. 1997. Resultatives, compositionality, and language variation. *Journal of East Asian Linguistics* 6. 1–49. DOI: 10.1023/A:1008257704110.

Žaucer, Rok. 2009. *A VP-internal/resultative analysis of 4 'VP-external' uses of Slavic verbal prefixes.* Ottawa: University of Ottawa. (Doctoral dissertation).

Chapter 3

Delimitatives, diminutive-iteratives and the secondary imperfective in North Slavic

Petr Biskup

University of Leipzig

This paper is concerned with diminutive-iterative verbs, delimitative verbs with the prefix *po-* and the secondary imperfective suffix. It is argued that diminutive-iterative *po-*verbs are derivationally based on delimitative predicates. Further, the secondary imperfective suffix is not an undifferentiated element. It is argued that the two instances of the imperfectivizing suffix – the iterative one and the progressive one – merge in distinct structural positions and that the delimitative prefix *po-* occurs between them. In the derivation of diminutive-iteratives, delimitative *po-* selects a predicate with a scalar structure and the Davidsonian event variable and contributes an extensive measure function to the base predicate. The iterative *-yva*, with its pluractional semantics, then iterates the eventuality denoted by the *po-*predicate.

1 Introduction

This section introduces diminutive-iterative *po-*verbs, delimitative *po-*verbs and secondary imperfective verbs and briefly overviews their morphosyntactic and semantic properties relevant to the analysis pursued in following sections.

1.1 Diminutive-iterative verbs

As to their form, diminutive-iterative verbs contain the prefix *po-* and the imperfectivizing/iterative suffix -YVA, as shown in (1) for Russian, in (2) for Polish, and

in (3) for Czech (see also Švedova 1980: 600, Katny 1994: 66–70, and Petr 1986: 398). In what follows, I will use -YVA as a shorthand that also stands for other allomorphs, e.g. in Russian, it stands for the allomorphs *-yva/-iva*, *-va* and *-a*.

(1) po-lëž-iva-t'
     DEL-lie-ITER-INF
     'to lie from time to time'   (Russian)

(2) po-płak-iwa-ć
     DEL-cry-ITER-INF
     'to cry from time to time'   (Polish)

(3) po-sed-á-va-t
     DEL-sit-TH-ITER-INF
     'to sit from time to time'   (Czech)

Concerning morphological (grammatical/viewpoint) aspect properties, diminutive-iterative *po*-verbs are always imperfective. With respect to their meaning, diminutive-iteratives are usually described as denoting a short action (with low intensity) that is repeated several times; see e.g. Zaliznjak & Šmelev (1997: 103) for Russian, Czochralski (1975: 23) for Polish and Karlík et al. (1995: 194, 209) for Czech. The action does not have to be repeated regularly and the iterative suffix brings about an unspecified number of instances of the particular eventuality. Typically, the number of repetitions depends on the context.

1.2 Secondary imperfective verbs

Secondary imperfective verbs contain an -YVA allomorph, which derives imperfective predicates from perfective stems, as shown in examples (4), (5) and (6). Crucially, -YVA allomorphs used in formation of diminutive-iterative verbs are identical to the -YVA allomorphs used in secondary imperfective verbs.

(4) a. za-pis-a-t'^{PFV}
         behind-write-TH-INF
         'to write down'
     b. za-pis-yva-t'^{IPFV}
         behind-write-SI-INF
         'to write down'
         'to be writing down'   (Russian)

(5) a. pod-pis-a-ć^{PFV}
         below-write-TH-INF
         'to sign'

     b. pod-pis-ywa-ć^{IPFV}
         below-write-SI-INF
         'to sign'
         'to be signing'   (Polish)

(6) a. vy-ps-a-t^{PFV}
         out-write-TH-INF
         'to excerpt'

     b. vy-pis-ova-t^{IPFV}
         out-write-SI-INF
         'to excerpt'
         'to be excerpting'   (Czech)

Secondary imperfective verbs can have (at least) four meanings. 1. progressive, expressing that a certain eventuality is in progress; 2. iterative, which expresses the successive occurrence of several instances of a certain eventuality; 3. habitual (generic), which describes an eventuality that is characteristic of an extended time period; and 4. general-factual, which typically refers to a realized or even completed eventuality (in a fashion similar to perfective verbs); see e.g. Comrie (1976: 24–40), Dahl (1985: 75–102), Dickey (2000: 49–125), Grønn (2004: 22–30), Timberlake (2004: 417–424) and Petruxina (2011: 64–76). This article is concerned with the progressive and iterative meanings.

1.3 Delimitative verbs

Delimitative verbs are formed with the help of the delimitative prefix *po-* and with the reduplicated form *popo-* in the case of motion verbs in Czech, as demonstrated by (7), (8), and (9). The prefix is adjoined to an unprefixed, imperfective stem and derives a perfective predicate, as shown in the examples under discussion.

(7) a. čit-a-t'^{IPFV}
         read-TH-INF
         'to read'
         'to be reading'

    b. po-čit-a-t'$^{\text{PFV}}$
        DEL-read-TH-INF
        'to read for a while'   (Russian)

(8)   a. siedzi-e-ć$^{\text{IPFV}}$
        sit-TH-INF
        'to sit'
        'to be sitting'
   b. po-siedzi-e-ć$^{\text{PFV}}$
        DEL-sit-TH-INF
        'to sit for a while'   (Polish)

(9)   a. nés-t$^{\text{IPFV}}$
        carry-INF
        'to carry'
        'to be carrying'
   b. popo-nés-t$^{\text{PFV}}$
        DEL-carry-INF
        'to carry sth. a little'   (Czech)

As to the meaning, the prefix delimits the eventuality denoted by the base predicate. Typically, it is a temporal delimitation, as in (7b) and (8b), but a property scale and a path scale can be delimited, too, as shown in (9b) for the path scale.

    Delimitative *po*-verbs are standardly claimed to be perfectiva tantum, i.e. they do not form secondary imperfectives; see Isačenko (1962: 391), Kopečný (1962: 110), Zaliznjak & Šmelev (1997: 94), and Łaziński (2020: 77). Thus, e.g. the delimitative verbs in (10a), (11a), and (12a) cannot be imperfectivized and receive a progressive delimitative meaning, as shown by the second translations in examples (10b), (11b), and (12b). The *po*-verb-YVA forms in the (b) examples can only have the diminutive-iterative meaning, as demonstrated by the first translations.

(10)   a. po-čit-a-t'$^{\text{PFV}}$
         DEL-read-TH-INF
         'to read for a while'
    b. po-čit-yva-t'$^{\text{IPFV}}$
         DEL-read-ITER-INF
         'to read from time to time'
         Unavailable reading: 'to be reading for a while'   (Russian)

(11) a. po-płak-a-ć[PFV]
          DEL-cry-TH-INF
          'to cry for a while'

    b. po-płak-iwa-ć[IPFV]
          DEL-cry-ITER-INF
          'to cry from time to time'
          Unavailable reading: 'to be crying for a while'   (Polish)

(12) a. popo-nés-t[PFV]
          DEL-carry-INF
          'to carry sth. a little'

    b. popo-náš-e-t[IPFV]
          DEL-carry-ITER-INF
          'to carry sth. from time to time'
          Unavailable reading: 'to be carrying a little/for a while'   (Czech)

I leave attenuative *po*-verbs like the Russian *popridéržat'* 'to hold gently' aside in this article because the attenuative *po*- and the delimitative *po*- behave as two distinct elements morphosyntactically.[1] While the attenuative *po*- selects perfective stems that are prefixed, delimitative *po*-verbs are derived from the base, imperfective stems (see e.g. Isačenko 1962: 391, 396, Zaliznjak & Šmelev 1997: 101, Petr 1986: 398). Further, in Czech, the delimitative *po*- often adds a dative reflexive argument that is licensed by the agent and is obligatory.[2] In contrast, attenuative *po*-verbs can have the dative *si* 'self' but it is never obligatory. In addition, the two types of *po*-verbs behave differently with respect to the formation of secondary imperfectives. As discussed in the preceding paragraph, delimitative *po*-verbs are considered to be perfectiva tantum. The authors mentioned there do not discuss whether or not attenuative *po*-verbs can be imperfectivized but (at least some) *po*-attenuatives form secondary imperfectives; e.g. Russian *popridéržat'* 'to hold

---

[1]This contrasts with the semantic analysis by Součková (2004a,b) and Kagan (2016), who treat the two types of *po*- prefixes as one and the same semantic element, measuring degrees on a scale. Note that an analysis with two distinct morphosyntactic *po*-s does not preclude the possibility that the two prefixes have identical or very similar semantic properties.

[2]Consider e.g. (i) and the discussion of (23b) in Section 2. The most straightforward analysis would introduce the dative *si* 'self' in the specifier of the delimitative *po*- projection, where the argument is c-commanded by the agent placed in the specifier of VoiceP.

(i) Strejda *(si) po-lyžoval a   odjel domů.
     uncle   self DEL-skied and went home
     'My uncle skied for a while and went home.'   (Czech)

gently' derives the imperfective form *popriderživat'* 'to hold gently/to be holding gently'. Tatevosov (2009: 96) and Klimek-Jankowska & Błaszczak (2022: 7–9, 2024: ex. (85), (92)) place the Russian and Polish, respectively, attenuative *pod-* below the imperfectivizing -ʏᴠᴀ. If the attenuative *po-* behaves in the same way, then attenuative *po-*verbs should be able to undergo secondary imperfectivization. Given that the attenuative *po-* selects a perfective stem, it must belong to positionally restricted prefixes in terms of Tatevosov (2009), which means that it merges below the imperfectivizing -ʏᴠᴀ.

The remainder of the article is structured as follows. Section 2 argues that diminutive-iterative *po-*verbs are derived from delimitative predicates. Section 3 then offers a morphosyntactic and semantic analysis with the relevant derivational steps. Section 4 concludes the article.

2 Diminutive-iterative verbs are derived from delimitative predicates

Recall from the preceding section that both delimitative and diminutive-iterative verbs contain the prefix *po-* and include a delimited degree scale. Given this and the imperfectivizing and iterative effects of -ʏᴠᴀ in diminutive-iterative verbs, a natural idea is that diminutive-iteratives, as in (13b), are derived by applying the imperfectivizing *-yva* to the delimitative predicate, as in (13a).

(13) a. po-lež-a-t'
          ᴅᴇʟ-lie-ᴛʜ-ɪɴꜰ
          'to lie for a while'
      b. po-lëž-iva-t'
          ᴅᴇʟ-lie-ɪᴛᴇʀ-ɪɴꜰ
          'to lie from time to time'   (Russian)

Besides the morphological, semantic and aspectual arguments, there are also phonological facts which support such an analysis. Suffixes used in the derivation of diminutive-iteratives are identical with -ʏᴠᴀ allomorphs used in the "standard" secondary imperfectivization. Also phonological processes involved in the formation of diminutive-iterative *po-*verbs are identical with phonological processes involved in the derivation of "standard" secondary imperfectives. Compare the diminutive-iterative example (14) with the secondary imperfectivization example in (15). In both examples *-yva* shifts the accent from the theme to the root vowel and in both roots, we also observe the vowel gradation (lengthening)

from /o/ to /a/.[3] Consider also (13), which displays a vowel gradation in the root, too, and which also manifests the accent shift from the theme to the root.

(14) a. po-kol-ó-t'$^{\text{PFV}}$
DEL-prick-TH-INF
'to prick a little'

b. po-kál-yva-t'$^{\text{IPFV}}$
DEL-prick-ITER-INF
'to prick from time to time'   (Russian)

(15) a. s-pros-í-t'$^{\text{PFV}}$
with-ask-TH-INF
'to ask'

b. s-práš-iva-t'$^{\text{IPFV}}$
with-ask-SI-INF
'to ask'
'to be asking'   (Russian)

The Polish examples in (16) and (17) show that in formation of both diminutive-iteratives and ordinary secondary imperfectives, /j/ is inserted to block hiatus.

(16) a. po-pi-ć$^{\text{PFV}}$
DEL-drink-INF
'to drink a little'

b. po-pi-ja-ć$^{\text{IPFV}}$
DEL-drink-ITER-INF
'to drink from time to time'   (Polish)

(17) a. wy-bi-ć$^{\text{PFV}}$
out-beat-INF
'to kill off'

b. wy-bi-ja-ć$^{\text{IPFV}}$
out-beat-SI-INF
'to kill off'
'to be killing off'   (Polish)

A hiatus-blocking process is also present in the Czech examples in (18) and (19). This time, /v/ is inserted between the theme vowel *-a* and the imperfectivizing *-a*

---

[3]The accent is represented with the diacritic length mark.

(according to the standard analysis, /v/ is the imperfectivizing suffix itself; see e.g. Karlík et al. 1995: 194; consider also Matushansky's 2009: 397 unifying analysis of imperfectivizing suffixes in Russian, arguing that /v/ derives from an underlying back rounded yer). In addition, both examples also display a lengthening of the theme after adding the imperfectivizing suffix.

(18) a. po-kašl-a-t$^{\text{PFV}}$ si
          DEL-cough-TH-INF self
          'to cough a little'

     b. po-kašl-á-va-t$^{\text{IPFV}}$
          DEL-cough-TH-ITER-INF
          'to cough from time to time'   (Czech)

(19) a. při-děl-a-t$^{\text{PFV}}$
          at-do-TH-INF
          'to fix'

     b. při-děl-á-va-t$^{\text{IPFV}}$
          at-do-TH-SI-INF
          'to fix'
          'to be fixing'   (Czech)

However, the standard literature does not adopt the analysis in which diminutive-iterative *po*-verbs are derived from delimitative predicates (see e.g. Isačenko 1960: 279–282, 1962: 407, Švedova 1980: 600, Zaliznjak & Šmelev 1997: 94, 104). According to them, delimitatives like (13a) are perfectiva tantum, diminutive-iteratives like (13b) are imperfectiva tantum and the verbs belong to different Aktionsarten: delimitative and diminutive-iterative. The authors claim that diminutive-iteratives are derived by circumfixation of *po-* and -YVA to the imperfective simplex predicate, i.e. to *ležat'* 'to lie' in the case of (13b) (see Katny 1994: 58ff. and Petr 1986: 419 for analogous claims with respect to Polish and Czech). This means that there is a strange coincidence. Both types of verbs have *po-* and some sort of a delimitative/diminutive meaning and the perfective verbs with the delimitative *po-* do not have a secondary imperfective counterpart and the imperfective verbs with *po-* (and -YVA) do not have a perfective counterpart. Tatevosov (2009: 133–134) shows that Isačenko's (1960) arguments are not strong enough and argues that there is a clear (derivational) relation between the existence of delimitative *po*-verbs and the existence of diminutive-iterative *po*-verbs in Russian.

The generative literature places delimitative *po-* and the imperfectivizing -ʏᴠᴀ in different structural positions. The delimitative prefix is mostly higher than -ʏᴠᴀ; see e.g. Romanova (2004: 271–272), Tatevosov (2008: 437–438) and Klimek-Jankowska & Błaszczak (2022: 9, 2024: ex. (85), (92)) (but see also Součková 2004a: 408–409, who assumes circumfixation, and Jabłońska 2004: 377–381, who proposes two positions for *po-*, below and above -ʏᴠᴀ). This placement has the advantage that it can account for why delimitative verbs are not (progressively) imperfectivized. However, it brings about two scope problems. First, it makes a false prediction with respect to morphological aspect. Given the position of the prefix being higher than -ʏᴠᴀ, one expects diminutive-iterative verbs to be perfective contrary to the facts; see Section 1.1 again. Second, if the delimitative prefix scoped over -ʏᴠᴀ, one would expect an interpretation with the repetition of standard actions in a delimited/short time frame, again contrary to the facts. For instance, in example (10b), repeated here as (20), one should receive a repetition of (normal) actions of reading in a short time frame.

(20) po-čit-yva-t'
      ᴅᴇʟ-read-ɪᴛᴇʀ-ɪɴꜰ
      'to read from time to time'   (Russian)

However, (20) is interpreted as a repetition of short actions of reading. This indicates reversed scope properties, with -ʏᴠᴀ being higher than *po-*.

Generally, there are three possibilities for how to derive diminutive-iterative *po-*verbs from *po-* delimitatives, as schematized in (21).

(21) a. [*po-* ... [-*yva* ... [... √root]]]
      b. [-*yva* ... [-*po-* ... [... √root]]]
      c. [-*yva* + *po-* ... [... √root]]]

The first possibility, as just discussed, makes wrong predictions with respect to interpretational possibilities and morphological aspect properties. Thus, since diminutive-iteratives are imperfective and since morphological aspect is determined by the last (highest) aspectual morpheme (e.g. Karcevski 1927: 96, Vinogradov 1947: 500, Dostál 1954: 482, Isačenko 1962: 416–418, Zinova & Filip 2015: 601–602, Tatevosov 2020: 28, Zinova 2021: 36–38, Biskup 2024: 55–61), the second option, shown in (21b), is preferred over the first possibility in (21a). Note that the delimitative *po-* cannot attach to the predicate after the imperfectivizing -ʏᴠᴀ because there are no exceptions to the perfectivizing effect of prefixation (Smith

1997: 242); there are only apparent exceptions (e.g. Schuyt 1990: 292, Zaliznjak & Šmelev 1997: 67–68, Zaliznjak 2017: 4–6).[4]

Given that interpretational properties of diminutive-iterative verbs are correctly derived by (21b), in contrast to (21a), the second option is preferred over the derivation in (21a) more generally. As to the circumfixation derivation in (21c), this option is generally disfavored, especially if both affixes exist independently (see e.g. Marušič 2023), as is the case with diminutive-iterative *po*-verbs. What is more, there is no reason to assume the special operation of circumfixation when (21b) successfully derives the relevant facts.

Since diminutive-iterative *po*-verbs cannot receive the progressive delimitative interpretation, only the iterative interpretation, as in (10b), (11b) and (12b), I assume that the iterative -YVA differs from the progressive -YVA with respect to their structural position. The progressive -YVA is generated below the delimitative *po*- and the iterative -YVA merges above the delimitative prefix, as illustrated in (22) (see Ramchand 2004: 33, who proposes that -YVA can occur in two different projections: Asp and Cuml).

(22)   […Iter -YVA … [… Del *po*- … [… Prog -YVA … [… √root]]]]

An argument against diminutive-iteratives derived from delimitatives could be based on the fact that there are derivational chains with missing links (see e.g. Součková 2004a: 409). Specifically, in contrast to the Russian *pokašljat'* 'to cough a little' and the Polish *pokaszleć/pokasłać* 'to cough a little', Czech (and Slovak) do not have the middle step of the derivational chain *pokašlat* (there is only *pokašlat si*), as shown in (23b).[5]

(23)   a.   kašl-a-t
             cough-TH-INF
             'to cough'

---

[4]Zinova & Filip (2015: 605–607) argue that iterative *pere*- and attenuative *po*- do not have to perfectivize in Russian. They assume that imperfectives like *perezapisyvat'* 'to (be) rerecord(ing)' can be derived by attaching *pere*- to the imperfective *zapisyvat'* 'to (be) record(ing)'. Their supporting argument is however based on borrowings, which are known to be anomalous in various respects. Specifically, when the prefixes discussed attach to a borrowed biaspectual verb, the new verb is still biaspectual like *perekvalificirovat'* 'to requalify'. A comparison with other languages suggests that in Russian the verb is not adapted enough to the language system to be able to accept the perfectivizing effect of the prefix. In contrast, the Czech *překvalifikovat* 'to requalify' is already perfective and there is also its imperfective counterpart *překvalifikovávat*.

[5]The Polish situation is somewhat controversial but *pokaszleć/pokasłać* can be found in *Wielki słownik języka polskiego* (Żmigrodzki 2022).

b. * po-kašl-a-t
   DEL-cough-TH-INF

c. po-kašl-á-va-t
   DEL-cough-TH-ITER-INF
   'to cough from time to time' (Czech)

Such an argument, however, is valid only if we assume a lexicalist framework, in which verbs like diminutive-iteratives are derived from complete word forms (see e.g. the definition of derivational chains in Zinova & Filip 2015: 601-602). In morphosyntactic approaches like the one assumed here verbs are derived incrementally, morpheme by morpheme, in a bottom-up fashion, not by attaching the imperfectivizing *-va* in (23c) to the complete verb **pokašlat*.

The same also holds for other morphemes in derivational chains, e.g. prefixes. Hence, *vypisat'* 'to excerpt' in (24a) is not derived by prefixation of *vy-* to the complete verb *pisat'* 'to write' since in the morphosyntactic approach assumed here, prefixes typically merge before the tense morpheme (infinitival *-t'* in (24)) and in certain cases also before the theme marker.[6]

(24) a. vy-pis-a-t'
     out-write-TH-INF
     'to excerpt'

  b. pis-a-t'
     write-TH-INF
     'to write'

To conclude this section, the imperfectivizing marker -YVA can appear in different structural positions and consequently, it can spell out distinct semantic properties.

---

[6]An analogous reasoning also applies to disappearing elements, e.g. *si* 'self' in (i). One might argue that in cases like (i), incrementality is violated because *si* does not have to be present in (i.c), in contrast to (i.b). Again, since the verb in (i.c) is not derived from the complete verb with *si* in (i.b), there cannot be disappearing *si* in (i.c).

(i) a. křič-e-t
     shout-TH-INF
     'to (be) shout(ing)'

  b. po-křič-e-t   *(si)
     DEL-shout-TH-INF   self
     'to shout a little for oneself'

  c. po-křik-ova-t   (si)
     DEL-shout-ITER-INF self
     'to shout from time to time' (Czech)

3 Analysis

In what follows, I elaborate the proposal from the preceding section, first from the semantic perspective, then from the syntactic point of view.

3.1 Combining *po*- and the verb stem

According to Isačenko (1962: 392), there are several restrictions on the formation of delimitative *po*-verbs. The restrictions are found not only in Russian, as shown in (25), (26) and (27). At least partially, they can be accounted for if it is assumed that the delimitative prefix needs a scale (for scale (degree) approaches to *po*- see Filip 2000, 2003; Jabłońska 2004, Součková 2004a,b; Kagan 2016 and Zinova 2021).

(25)   * po-zyskać
       DEL-win
       Intended: 'to win for a while'   (Polish)

(26)   * po-bodnout
       DEL-stab
       Intended: 'to stab for a while'   (Czech)

(27)   * po-stoit'
       DEL-cost
       Intended: 'to cost for a while'   (Russian)

Given that achievements denote a momentaneous change of state, there is no (protracted) scale that could be delimited by the prefix and cases like (25) are ungrammatical.[7] In the same vein, since semelfactive verbs denote punctual eventualities, the delimitative *po*- cannot measure them and examples like (26) are ungrammatical as well.

An interesting case is stative predicates like the one in (27), which also cannot be prefixed with the delimitative *po*-. The ungrammatical status cannot be ascribed to the fact that a temporal extent scale is missing because the base predicate of such states can be modified by durative adverbials, as is the case e.g. with Russian *znat' kogo tri dnja* 'to know somebody for three days'. The reason also cannot be the homogeneity requirement of delimitative *po*- (Filip 2000: 61, 2003:

---

[7]The so-called degree achievements are known for their special behavior; they provide an appropriate change-of-state scale.

91; Mehlig 2006: 247) because states are homogeneous. Atelicity does not play a role either because states are atelic and other atelic predicates – e.g. activities like the Polish *płakać* 'to cry' – can be prefixed with delimitative *po-*. So, one might argue that the problem lies in the concept of change because stative predicates do not entail a change, in contrast to dynamic predicates (activities, accomplishments and achievements). However, this reasoning is not correct, either, since there are also states that can be prefixed with the delimitative *po-*; consider the examples (28), (29) and (30).

(28)   Russian
    a. po-stojat'   b. po-sidet'   c. po-spat'
       DEL-stand   DEL-sit   DEL-sleep

(29)   Polish
    a. po-stać   b. po-siedzieć   c. po-spać
       DEL-stand   DEL-sit   DEL-sleep

(30)   Czech
    a. po-stát   b. po-sedět   c. po-spat
       DEL-stand   DEL-sit   DEL-sleep
       'to stand for a while'   'to sit for a while'   'to sleep for a while'

The unprefixed verbs in (28)–(30) belong to the class of "interval statives" (Dowty 1979: 173–180) and differ from "static states" (Bach 1986: 6) like 'to know', 'to own' and 'to cost', which are less dynamic and do not accept the delimitative *po-* in North Slavic. Maienborn (2003, 2005) analyzes dynamic states like 'to stand' as "Davidsonian states" because they refer to eventualities in the sense of Davidson, so they introduce a Davidsonian event argument. In contrast, static states belong to "Kimian states" in her analysis. They do not have a Davidsonian event variable but introduce a specific Kimian-state referential argument. In fact, the German modifier *ein bisschen* 'a little bit' distinguishing Davidsonian states from Kimian states in her eventuality diagnostic (Maienborn 2005: 297–299) behaves like delimitative *po-* with respect to grammaticality judgements. Here I follow Maienborn's proposal and assume that delimitative *po-* selects a predicate with a Davidsonian event argument (that in addition has some scalar structure, as discussed above).

   According to Filip (2000: 61–66, 2003: 89–90), delimitative *po-* applies to a homogeneous predicate and contributes an extensive measure function, which is contextually specified and meets or falls short of some contextually determined

expectation value. Součková (2004a: 410, 2004b: 73) modifies Filip's proposal and argues that the measure function only applies to events, as shown in (31).[8]

(31) $\llbracket po\text{-}\rrbracket = \lambda P \lambda e[P(e) \wedge m(e) = c_{\text{relatively.small}}]$

P is a variable over predicates, *m* stands for the extensive measure function applied to an event and *c* means that its value is contextually determined. The function *m* measures events that contain some scalar structure, concretely, it measures the degree of change on the appropriate scale and it can apply to various dimensions (types of scales). It depends on lexical properties of the particular verb, which dimension – if at all; recall the discussion of achievements and semelfactives in (25) and (26) – is accessible. For instance, in the case of motion verbs, there is a scale of progress along the path; in degree achievements, there is an increase in the degree on a property scale; but in most cases, the prefix is applied to a time scale. Note, however, that it is not possible to reduce all delimitative cases to the time scale and that particular scales do not have to coincide with respect to the degree of change. For instance, in the case of Czech *poponést* 'to carry sth. a little' in (9b), it is the progress on the appropriate path that is measured and the relatively small value of this progress (the short path) can be in contrast to the duration of the carrying eventuality, which under appropriate circumstances can even be very long.

As just mentioned, (31) states that the value of the measure function is relatively small in the context. I use the equal relation in (31) instead of the less-than-or-equal relation (in contrast to e.g. Kagan 2015: 47, 2016: 310; Klimek-Jankowska & Błaszczak 2024: ex. (68)) since it directly brings about the quantization property of *po*-delimitatives. necessary for iteration, as discussed in the following section. More concretely, if a delimited event of crying takes five minutes in the specific context, i.e. the value of the measure function is contextually determined to be equal to five minutes, then there will be no proper part of the event that also falls in the denotation of this crying *po*-predicate. In contrast, if the less-than-or-equal relation were used in (31), there could be a proper part – e.g. crying for just four minutes – falling in the denotation of the crying *po*-predicate.

Thus, building on Součková's proposal, the meaning of the Polish *popłaka(ć)* 'to cry for a while' – after applying the delimitative prefix *po-* to the predicate *płaka(ć)* 'to cry' – will look like (32).

---

[8]The proposal that the measure function applies to events can be already found in Piñón (1994: 362–363). Součková (2004a,b) also argues that the prefix can also apply to non-homogeneous predicates since she unifies delimitative *po-* and the attenuative *po-*. Given that I keep the two prefixes apart (see Section 1.3), I assume that the delimitative *po-* applies to homogeneous predicates. The attenuative *po-* then probably only applies to non-homogeneous predicates, as in the case of the Russian *poprideržat'* 'to hold gently'.

(32) $\llbracket po\text{-}płaka\text{-}\rrbracket = \lambda e[\text{CRY}(e) \wedge m(e) = c_{\text{relatively.small}}]$

Given the meaning in (31), with *P* as a predicate over events and the measure function applied to an event, it is obvious that delimitative *po-* cannot combine with states like *stoit'* 'to cost' in (27), which only have the Kimian-state referential argument.

3.2 Combining the delimitative *po*-predicate and the iterative -YVA

According to the proposal in Section 2, diminutive-iteratives are derived from delimitative predicates in the way that the eventuality delimited/measured by *po-* is iterated by means of -YVA. In other words, delimitative *po-* is responsible for individuation and the iterative -YVA then for pluralization. Thus, given that for iteration and counting, discrete elements are necessary, the question arises as to which concept is responsible for the individuation here. Is quantization sufficient or is telicity necessary as well? According to Filip (2003: 91), delimitative *po-* ("attenuative" in her terms) makes predicates not only quantized, analogously to measure functions like *a* (*relatively*) *small quantity* and *a few*, but also telic given her definition of telicity based on atomicity (see Filip 2003: 60–61). However, if the standard adverbial test is a reliable diagnostic for (a)telicity, then delimitative *po*-verbs must be atelic because they are compatible with durative adverbials like 'an hour (long)' and incompatible with time-span adverbials like 'in an hour' in the relevant reading. In the light of grammatical diminutive-iteratives like *polëži-vat'* 'to lie from time to time' in (1), that in turn means that quantization brought about by an extensive measure function like the delimitative *po-* is sufficient for applying the iterative *-yva-* and that telicity is not a necessary condition in this case.

As mentioned in Section 1.1, the iterative marker brings about an unspecified number of instances of the particular eventuality and the cardinality of repetitions is contextually determined. Since it is difficult to determine the smallest number of repetitions here, I assume the weakest position and use the meaning of plurals as a base. This means that the cardinality of the iterated eventuality is greater than one, as shown in (33), where $|e|$ stands for the number of atomic events.

(33) $\llbracket \text{ITER} \rrbracket = \lambda P \lambda e[P(e) \wedge e = \sigma e'[P(e') \wedge e' \subset e \wedge |e| > 1] \wedge \forall e'[\text{ATOM}(e') \rightarrow \neg \exists e''[P(e'') \wedge e'' \subset e \wedge \text{ATOM}(e'') \wedge \tau(e') \supset \subset \tau(e'')]]]$

For deriving pluralities, usually, Link's (1983) σ-operator and the *-operator are used. Therefore, for the iterative -YVA, I use Kratzer's proposal (Kratzer 2008:

296, see also Boneh & Doron 2008: ex. (31) and Ferreira 2016: 358), according to which the sum of all events e' that are proper parts of the event e and have the property P is identical to e. Since it brings about a weak notion of plurality, with singularities as special cases, I also use the conjunct $|e| > 1$ in (33), as discussed above.

The iterated events are not temporally adjacent, e.g. the meaning of the Polish *popłakiwać* 'to cry from time to time' in (2) is characterized by Żmigrodzki (2022) as *płakać z przerwami* 'to cry with pauses' (see also Katny 1994: 67). For this reason, I add to the meaning in (33) the restriction on temporal traces. Specifically, for every atomic proper part e' of the event e, it holds that there is no atomic subevent e'' with the property P that is a proper part of e and whose temporal trace abuts with the temporal trace of e'. Under the assumption that the abut relation (precluding any contact) is stronger than the overlap relation, it holds that if the temporal traces of e' and e'' do not abut, then they also do not overlap. The size of the time interval between e' and e'' is not defined here since it depends on the lexical meaning of the appropriate predicate and on the context. Moreover, intervals between the particular subevents can be of different sizes.

I do not use the classical non-overlap condition by Lasersohn (1995: 256) (see also e.g. Wood 2007: 126), with the function f standing for temporal, spatio-temporal or participant-based distributivity, since in the case of diminutive-iterative *po*-verbs, participants and spaces can overlap. As an illustration, consider example (34), in which the referent of the expression *Kasia* is identical for the crying subevents.[9] The crying subevents of example (34) also can (but do not have to) happen in an identical space.

(34) Kasia cały   dzień po-płak-iwa-ł-a.
      Kasia whole day   DEL-cry-ITER-PTCP-F
      'Kasia cried repeatedly the whole day.'   (Polish)

To receive the separated-in-time reading, Lasersohn (1995: 254) adds a between-ness condition to the non-overlap condition, which introduces a time that inter-venes between temporal traces of the singular events. I do not follow his proposal because I assume that the time variable is introduced by the aspectual head later in the derivation. Instead of the betweenness and non-overlap conditions, I use the abut condition, as stated in (33).

When the meaning of the iterative operator is applied to the meaning of the de-limitative predicate *popłaka(ć)* in (32), we obtain a diminutive-iterative predicate over events, as demonstrated in (35).

---

[9]With plural subjects, distributive readings are possible but not necessary.

(35) $[\![\textit{popłakiwa-}]\!]$
$= \lambda e[*\text{CRY}(e) \wedge m(e) = c_{\text{relatively.small}} \wedge e = \sigma e'[*\text{CRY}(e') \wedge m(e') = c_{\text{relatively.small}} \wedge e' \subset e \wedge |e| > 1] \wedge \forall e'[\text{ATOM}(e') \rightarrow \neg\exists e''[*\text{CRY}(e'') \wedge m(e'') = c_{\text{relatively.small}} \wedge e'' \subset e \wedge \text{ATOM}(e'') \wedge \tau(e') \supset\subset \tau(e'')]]]$

Thus, having established the meaning of iterative -YVA, let us now look at the lower instantiation of the marker, progressive -YVA.

3.3 The progressive -YVA and imperfectivity

As already mentioned, when -YVA attaches to a delimitative predicate, the form cannot have the progressive meaning. It can only receive an iterative interpretation; consider e.g. the Russian *počityvat'* 'to read from time to time' in (10b) again. This results from the splitting of -YVA into two different syntactic positions and from the positioning of the delimitative *po-* between them, as discussed in Section 2. So, what is the meaning of the progressive -YVA?

Progressivity is often defined in terms of partitivity (Filip 1999: 171–175, 213 and references therein), as is (secondary) imperfectivity (e.g. Łazorczyk 2010: 134–139). Although progressivity and (Slavic) imperfectivity are close notions (Zucchi 1999: 200), they are not identical; see e.g. Comrie (1976: 33) and Dahl (1985: 92). Recall also from Section 3.2 that the aspectual projection, which is going to encode (im)perfectivity in the current proposal, occurs in a higher syntactic position, hence it is not identical to the progressive projection spelled out by the progressive *-yva-*. Thus, since the progressive brings about an internal part of the eventuality, I assume the (for simplicity extensional) meaning in (36) for the progressive -YVA.

(36) $[\![\text{PROG}]\!] = \lambda P \lambda e' \exists e[P(e) \wedge e' < e]$

It is based on Krifka (1992: 47) progressive operator but uses the proper-part-of relation instead of just part-of relation. The reason is that we need to exclude the possibility that the event culminates with the progressive -YVA.[10]

Given the proposal that there are two different syntactic positions (Prog and Iter; see (22) again) with distinct meanings that are spelled out as -YVA, the question arises whether the progressive -YVA can co-occur with the iterative -YVA. It is possible to test it with Czech since it allows combinations of more -YVA markers.

---

[10]Filip (2005) uses the part-of relation for the meaning of the imperfective morphological aspect since it can also derive the meaning of general-factuals, which can refer to a culminated eventuality. Also Łazorczyk (2010) uses the part-of relation in the meaning of her secondary imperfective operator.

Consider example (37), in which the imperfectivizing *-vá* is attached to the un-prefixed perfective stem *dá-* and then another *-(y)va* marker is adjoined, forming the imperfective *dávával*.[11]

(37) (*Včera)   dá-vá-va-l   peníze chudým.
      yesterday give-PROG/ITER-HAB-PTCP money poor.ADJ.DAT.PL
      (Intended:) '(Yesterday) he had the habit of giving money to the poor.'

Although the lower *-vá* can bring about the progressive or the iterative interpretation, the higher *-va* can only bring about the habitual meaning in (37). Therefore, the sentence cannot be interpreted non-habitually. The ungrammatical status of the adverbial 'yesterday' indicates that an episodic reading is impossible. If an iterative adverbial is used, the sentence must also receive a habitual interpretation, as shown by the translation in (38).

(38) Dá-vá-va-l   peníze chudým dvakrát.
      give-PROG/ITER-HAB-PTCP money to.poor   twice
      'He had the habit of giving money to the poor twice.'

For instance, a person could have the habit of giving money to the poor daily at 9 a.m. and 7 p.m. In contrast, it is not possible that there were only two occasions of giving money to the poor in total. Thus, the outer -YVA in cases like (38) always instantiates a habitual operator that is structurally higher than the iterative and progressive suffixes.[12]

The incompatibility of the iterative and progressive -YVA has a semantic reason. As discussed in Section 3.2, the iterative operator applies to quantized predicates. However, the progressive operator derives a homogeneous (i.e. cumulative and divisive) predicate, which cannot serve as an input for the iterative -YVA. According to Łazorczyk (2010: 137–138), the secondary imperfective operator takes a telic predicate and returns a homogeneous subpart of it (see also Filip 1999: 167,

---

[11]The imperfective *dáva-(t)* can have either the progressive or the iterative interpretations, hence I gloss *-vá* with PROG/ITER.

[12]For more on the habitual -YVA, see Filip & Carlson (1997), Esvan (2007: 321–340), Berger (2009), Nádeníček (2011: 132–135), Nübler (2017), and also Biskup (2023), who shows that the habitual -YVA differs from the secondary imperfective -YVA – the progressive and iterative -YVA in the current approach – in phonological, aspectual and interpretational properties. Interpretationally, the habitual -YVA brings about a (generic) quantificational semantics in contrast to the pluractional iterative -YVA. Since the habitual -YVA is beyond the scope of this article, I will not discuss it any further here.

who argues that progressive sentences are cumulative).[13] Unsurprisingly, this behavior of the progressive operator corresponds with morphological aspect properties; we saw in Section 1.2 that the imperfectivizing -YVA selects a perfective stem and derives an imperfective predicate.[14]

Consequently, since the progressive -YVA selects a perfective, telic stem, the *po*-verbs *popisyvat'* 'to write from time to time', *popłakiwać* 'to cry from time to time' and *posedávat* 'to sit from time to time' cannot be derived in the way schematized in (39), i.e. as "delimited progressives". The reason for this is that in derivations like (39), in contrast to what was just said, the progressive -YVA would have to apply to stems that are imperfective, homogeneous and atelic: 'to write', 'to cry' and 'to sit'.

(39) a. * [po-[pis-yva]]-t'
          DEL-write-PROG-INF   (Russian)

      b. * [po-[płak-iwa]]-ć
          DEL-cry-PROG-INF   (Polish)

      c. * [po-[sed-á-va]]-t
          DEL-sit-TH-PROG-INF   (Czech)

Another alternative would be to assume that the progressive reading of *po*-verbs with -YVA is excluded because the delimitative *po-* applies first and makes the predicate quantized and individuated, with the consequence that proper parts of the denoted event are not accessible to the progressive -YVA (which applies as second). This proposal, however, would be very restricted because it could only work in the case of an incremental theme (see Filip 2005: 273). Let us test it with other prefixes.

(40) Pavel teď vy-pis-uj-e   / o-pis-uj-e   celou
      Pavel now out-write-PROG-3.SG   about-write-PROG-3.SG whole
      přednášku.
      talk
      'Pavel is excerpting/copying the whole talk right now.'   (Czech)

---

[13]See also Tatevosov (2015: 489), who builds on Paslawska & von Stechow (2003: 346) and treats the imperfectivizing -YVA as an eventizer.

[14]Note that despite the fact that the iterative -YVA and the progressive -YVA cannot co-occur, sentences with a diminutive-iterative verb can receive a simultaneous interpretation when the reference time (expressed e.g. by the temporal clause) is included in the event time, as in (i), based on (34).

(i) Kiedy wszedł   do pokoju, Kasia po-płak-iwa-ł-a.
     when came.3SG.M into room   Kasia DEL-cry-ITER-PTCP-F
     'When he came into the room, Kasia was crying a little.'

Although there is a secondary imperfective predicate with prefixes making the base verb quantized, combined with a quantified incremental theme in (40), the progressive reading is available. Thus, this alternative explanation of the impossibility of the progressive delimitative interpretation also does not work.

At this point, the question arises as to how the difference between progressivity and imperfectivity is modelled in the current approach. The meaning of the progressive operator in (36) shows that progressivity concerns the internal structure of eventualities. As to the morphological aspect, I make the standard assumption that it concerns temporal properties of eventualities, i.e. relates the event time to the reference time via the inclusion relation (see e.g. Paslawska & von Stechow 2003: 322). Concretely, in the case of the imperfective aspect, the reference time is included in the event time.

3.4 Morphosyntactic derivation

Building on Biskup (2019: 36−42, 2024), I assume that the value of the morphological aspect of diminutive-iterative *po*-verbs is determined in the aspectual head via the operation Agree. Combining the structure in (22) with the proposal that all aspectual markers are separated from the aspectual interpretation of the aspectual head (Biskup 2023), we receive (41), with parts relevant to our discussion.[15]

(41) [Asp$_{[Asp:?]}$ ... [Iter -YVA$_{[IPFV]}$ ... [Del *po*-$_{[PFV]}$ ... [Prog -YVA$_{[IPFV]}$ ... [√root]]]]]

It shows that aspectual markers like the iterative -YVA, the delimitative *po*- and the progressive -YVA bear an aspectual feature – either with the value [perfective] or [imperfective] – which can value the unvalued aspectual feature of the aspectual head.

Concerning the aspectual projection, it is standardly placed above the projection introducing the agent (see Babko-Malaya 2003, Błaszczak & Klimek-Jankowska 2012, Gribanova 2015). This placement is also supported by the fact that although Russian *-nie* nominals are aspectless (e.g. Schoorlemmer 1995, Gehrke 2008), they can have an agent argument, as demonstrated by the agent-oriented modifier in (42a) and the agentive *by*-phase in (42b).[16]

---

[15]With the aspectual interpretation of the Asp head, the relation between the reference time and the event time is meant, as discussed in the last paragraph of Section 3.3. This interpretation is different from the semantics of particular aspectual markers, like prefixes, the semelfactive and habitual suffixes and the iterative and progressive -YVA, as discussed in Section 3.2 and Section 3.3 (see also Paslawska & von Stechow 2003, Grønn 2004, Tatevosov 2011, 2015).

[16]Regarding the aspectless status of Russian *-nie* nominals, a reviewer asks whether (42a) can have a durative modifier. Given that *for*-adverbials and *in*-adverbials diagnose (a)telicity rather than (im)perfectivity, this test is not effective.

(42)   a. umyšlennoe prestuplenie
            deliberate   crime
            'a wilful crime'

        b. soveršenie   prestuplenija licom...
            perpetration crime.GEN.SG person.INS.SG
            'a perpetration of the crime by a person'

                                            (Russian; Biskup 2023: ex. (69))

It has been argued that the secondary imperfective suffix merges inside the verbal domain below the agent (see e.g. Romanova 2004: 272 and Tatevosov 2015: 488 for Russian, Kwapiszewski 2022 for Polish and Biskup 2023 for Czech). However, given the splitting of -YVA into the iterative and progressive -YVA in the current proposal, we need to know more about the positioning of the agent. There are agent nominalizations ending in *-tel'* in Russian, *-ciel* in Polish and *-tel* in Czech that can contain the imperfectivizing -YVA. Semantically, the suffixes *-tel'*, *-ciel* and *-tel* (and others, like the agentive version of the Czech *-č* in (44)) relate to the projection containing the agent (external) argument since they introduce an entity – predominantly, a person – that carries out the action denoted by the predicate to which they are attached.[17] Thus, based on Baker & Vinokurova's (2009: 531) analysis of nominalizing affixes like the English *-er*, I consider the morphemes *-tel'*, *-ciel* and *-tel* to be nominal versions of the agentive Voice head. What is crucial for us is that the imperfectivizing -YVA is always closer to the root than the agentive nominalizing suffix, as demonstrated in (43a).

(43)   a. do-pis-ova-tel
            to-write-ITER-NMLZ
            'correspondent'

        b. ob-jev-i-tel   Plut-a
            about-show-TH-NMLZ Pluto-GEN
            'the discoverer of Pluto'   (Czech)

Since (43a) refers to a person repeatedly performing the event of making a report, I take the *-ova* suffix to represent the iterative head. Note that the iterative meaning cannot be a property of *-tel* since this suffix also attaches e.g. to perfective predicates denoting a single event of discovering without changing the cardinality of the action, as shown in (43b). Thus, the order of the imperfectivizing -YVA and the nominalizing suffix can be taken to mean that the projection of Voice

---

[17]In the traditional terminology, nouns in *-tel* are called *nomina agentis*.

– hosting the agent – is structurally higher than the iterative projection spelled out by -YVA.[18]

I am not aware of diminutive-iterative *po*-predicates with *-tel'*, *-ciel* or *-tel* but there is at least the deverbal nominalization *pojídač* 'eater', in which the agentive *-č* occurs outside the diminutive-iterative predicate *pojída(t)* 'to eat from time to time', as shown in (44). The presence of the iterative meaning (expressed by *-a*) is confirmed by the fact that the singular complement *krevety* is ungrammatical in contrast to the plural *krevet*.

(44) po-jíd-a-č   krevet   / *krevety
      DEL-eat-ITER-NMLZ shrimp.PL   shrimp.SG
      'shrimp eater'   (Czech)

Building on these nominalization facts and the placement of the aspectual head above the projection introducing the agent, the relevant piece of morphosyntactic structure of diminutive-iterative *po*-verbs looks like (45), which is based on (41). Note that I added the standard verbalizing head *v* and that the structure does not contain the progressive projection now since the progressive reading of delimitative *po*-verbs with -YVA is excluded.

(45)   [Asp$_{[Asp:IPFV]}$ [Voice [Iter -YVA$_{[IPFV]}$ [Del *po*-$_{[PFV]}$ [*v* [√root]]]]]]

Given that the aspectual value is determined by the aspectual marker that is attached last (i.e. that is closest to Asp; see discussion in Section 2), when the aspectual head probes, then it finds the aspectual feature of the iterative -YVA first. Consequently, the Agree operation uses this [imperfective] feature and diminutive-iterative *po*-verbs always occur as imperfective, as shown in (45).

Structures of diminutive-iterative *po*-verbs are not difficult to linearize since all affixes are structurally higher than the root, as illustrated for the elements under discussion in (45). If the Affix-Specific Linearization by Harley (2013) is assumed – which encodes the prefixal versus affixal property directly in the specific marker – then no head movement is necessary. Only argument phrases need to be evacuated from the extended verbal projection. Consider e.g. the relevant part of linearization of the Polish verb *popłakiwać* 'to cry from time to time' in Figure 1.

The verbalizing head *v* is phonologically empty in this derivation but in other cases a theme vowel can be inserted. Since theme vowels are suffixes, the *v* head

---

[18]The vocabulary item -YVA can be inserted into the head of the iterative and progressive projections since it is specified as [imperfective].

```
                              ...
                         ╱          ╲
                      ...              T
                       │              │
                    VoiceP            -ć
                  ╱        ╲
               IterP        Voice
             ╱       ╲
          DelP        Iter
        ╱      ╲        │
      Del       vP     -iwa
       │      ╱    ╲
      po-   √P       v
             │
           płak
```
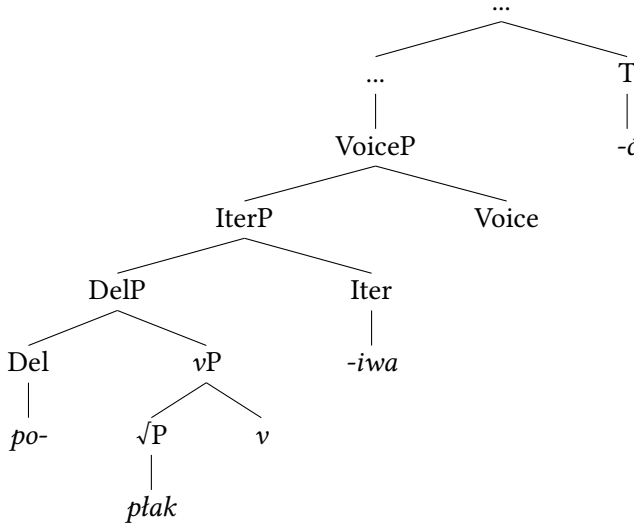
Figure 1: Linearization of Polish *popłakiwać* 'to cry from time to time'

is placed to the right also here. Delimitative *po-* and iterative *-iwa* are placed to the left and to the right, respectively, in accordance with their prefixal and suffixal status. I do not decompose imperfectivizing suffixes in this article but it is possible to split them, e.g. the Russian *-yva* into *-yv-a*, analogously the Polish *-ywa* into *-yw-a* and the Czech *-ova* into *-ov-a*. The second element (*-a*) could be analyzed as a theme vowel that, e.g. in the structure of *popłakiwać*, spells out the Voice head. Therefore, I put the head to the right in (1). Finally, *-ć* can be taken to represent the infinitival T head, linearized to the right in accordance with the suffixal status of the marker.

One could assume that Figure 1 in fact represents the syntactic structure already before linearization (if one prefers a more powerful syntactic module). This would however go against the standard SVO analysis of Slavic languages, which supposes that heads take their complements to the right in Slavic (but see also Haider & Szucsich 2022: 34–35).

Delimitative *po-* is not the only prefix displaying the special, iterative type of behavior. Comitative *pod-*, e.g. in the Russian *podpevat'* and in the Polish *podśpiewywać*, both with the meaning 'to sing a little with sth.', seems to behave in the same way.

Flier (1985: 41) shows that perdurative ("delimitative" in his terms) *pro-* manifests very similar behavior; consider example (46).

(46)   a. Často vesennie večera   ona pro-siž-iva-l-a   na vysokom
           often spring   evenings she through-sit-ITER-PST-F.SG on high
           kryl'ce.
           porch
           'She would often spend entire evenings in the spring sitting on the
           high porch.'

       b. * Ona dolgo prosiživala$^{\text{IPFV}}$ na vysokom kryl'ce.
          she long   through.sit.ITER.PST.F.SG on high   porch
          Intended: 'She would spend a long time sitting on the high porch.'

       c. * Ona sejčas prosiživaet$^{\text{IPFV}}$ na vysokom kryl'ce.
          she now   through.sit.ITER.PRS.SG on high   porch
          Intended: 'She is now spending time sitting on the high porch.'

                                         (Russian; Flier 1985: 41)

As shown by (46a), the event of sitting is iterated. The incompatibility of the imperfective verb with the adverbial *dolgo* 'for a long time' in (46b) demonstrates that the durative meaning is excluded and the incompatibility of the verb in present with the adverbial *sejčas* 'now' in (46c) shows the impossibility of the progressive meaning. A larger context showing the impossibility of the progressive reading is provided in (47).

(47)   * Ona pro-siž-iva-l-a   vesennie večera   na vysokom kryl'ce
          she   through-sit-ITER-PST-F.SG spring   evenings on high   porch
          v tot moment, kogda bitva načalas'.
          in the moment when battle begin.PST
          Intended: 'She was spending entire evenings in the spring sitting on the
          high porch when the battle began.'

Thus, this perdurative *pro-* seems to be the next candidate for the positioning between the progressive and iterative -YVA.

    The current proposal has consequences for the overall architecture of verbal predicates. Before the splitting of the secondary imperfective marker into the progressive and the iterative -YVA, there were two structural possibilities for verbal prefixes: below and above the secondary imperfective morpheme. Now, after the splitting, there are three options, as shown in (48).

(48)   [Asp [Voice [SP$_{\text{high}}$ [Iter -YVA [SP$_{\text{high}}$ (Del *po-*) [Prog -YVA [SP$_{\text{low}}$ [*v*
      [√root [LP]]]]]]]]]]

Specifically, (i) below the progressive -YVA for (lexical, intermediate and lower superlexical) prefixes in predicates that can receive the progressive and iterative interpretations, like the Russian *pro-davat'*[IPFV] 'to sell'; (ii) between the progressive and the iterative -YVA for higher superlexical prefixes in predicates which cannot be progressivized but can have an iterative interpretation, like delimitative *po-* in diminutive-iteratives and perdurative *pro-* in (46); (iii) above iterative -YVA for higher superlexical prefixes in predicates which cannot be imperfectivized – like ingressive *roz-* in the Polish *rozboleć*[PFV] 'to start to ache' – and for higher superlexical prefixes which perfectivize a secondary imperfective predicate, like distributive *po-* in the Russian *po-vytalkivat'*[PFV] 'to push out one after another'.[19]

4 Conclusions

I have argued that diminutive-iterative *po*-verbs are derived from delimitative *po*-predicates. The secondary imperfective marker is split into two distinct elements, both syntactically and semantically: iterative -YVA and progressive -YVA. The iterative marker has a pluractional meaning and merges higher than delimitative *po-*, whereas progressive -YVA is a partitive operator that occurs below the prefix. In diminutive-iterative verbs, the progressive operator is not present and delimitative *po-*, with its measure function meaning, applies to the simplex predicate. Then, the event denoted by the quantized predicate is iterated by the pluractional imperfectivizing -YVA. I have discussed certain restrictions on the formation of delimitative *po*-verbs and argued that only eventualities with a scalar structure and an event variable can be delimited. Static states, which contain the Kimian state referential argument, are not compatible with delimitative *po-*. As to morphosyntactic structure, I have argued that the iterative projection, spelled out by -YVA, occurs inside the verbal domain below the Voice projection, which introduces the external argument. Morphological aspect properties of diminutive-iteratives are determined in the aspectual projection via Agree with the closest aspectual feature, i.e. the imperfective feature of the iterative head. We have also seen that the overall (aspectual) architecture of Slavic predicates is more fine-grained and that there are more prefixal positions than usually assumed.

---

[19]Mehlig (2007, 2012) discusses examples of delimitative *po*-verbs in Russian like *pootkryvat' okno* 'to open a window for a while', which at first sight, suggest that the delimitative *po-* can also merge higher than the iterative *-yva*. Such predicates denote an attempt to attain the change of state through several different actions. These cases probably are not problematic since they could be analyzed in terms of the partitive progressive operator (which I argued to be lower than the delimitative *po-*), as in Tatevosov & Ivanov (2009).

Petr Biskup

Abbreviations

DEL	delimitative	PFV	perfective
F	feminine	PL	plural
GEN	genitive	PROG	progressive
HAB	habitual	PRS	present
INF	infinitive	PST	past
INS	instrumental	PTCP	participle
IPFV	imperfective	SG	singular
ITER	iterative	SI	secondary imperfective
M	masculine	TH	theme (vowel)
NMLZ	nominalizer		

Acknowledgments

References

Babko-Malaya, Olga. 2003. Perfectivity and prefixation in Russian. *Journal of Slavic Linguistics* 11(1). 5–36. https://www.jstor.org/stable/24599703.

Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 9(1). 5–16. DOI: 10.1007/BF00627432.

Baker, Mark C. & Nadya Vinokurova. 2009. On agent nominalizations and why they are not like event nominalizations. *Language* 85(3). 517–556. DOI: 10.1353/lan.0.0144.

Berger, Tilman. 2009. Anmerkungen zur Produktivität der tschechischen Iterativa. In Lenka Scholze & Björn Wiemer (Hg.) (eds.), *Von Zuständen, Dynamik und Veränderung bei Pygmäen und Giganten: Festschrift zum 60. Geburtstag von Walter Breu*, 25–43. Bochum: Universitätsverlag Dr. Norbert Brockmeyer.

Biskup, Petr. 2019. *Prepositions, case and verbal prefixes: The case of Slavic.* Amsterdam: John Benjamins.

Biskup, Petr. 2023. Aspect separated from aspectual markers in Russian and Czech. In Petr Biskup, Marcel Börner, Olav Mueller-Reichau & Iuliia Shcherbina (eds.), *Advances in formal Slavic linguistics 2021*, 61–98. Berlin: Language Science Press. DOI: 10.5281/zenodo.10123631.

Biskup, Petr. 2024. An Agree analysis of the morphological aspect in Slavic. In Silke Fischer, Doreen Georgi, Fabian Heck, Johannes Hein, Anke Himmelreich, Andrew Murphy & Philipp Weisser (eds.), *Strictly cycling*, 55–74. Leipzig: Institut für Linguistik, Universität Leipzig. https://www.philol.uni-leipzig.de/fileadmin/Fakult%C3%A4t_Philo/Linguistik/Forschung/LAB/LAB_97/LAB97_04_biskup.pdf.

Błaszczak, Joanna & Dorota Klimek-Jankowska. 2012. Futures in Polish and Slovenian: 'A hole in a sock' theory. In Alexander Podobryaev (ed.), *Formal Approaches to Slavic Linguistics 20: The second MIT meeting 2011*, 17–32. Ann Arbor: Michigan Slavic Publications. https://lsa.umich.edu/content/dam/slavic-assets/slavic-documents/20FASL.pdf.

Boneh, Nora & Edit Doron. 2008. Habituality and the habitual aspect. In Susan Rothstein (ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect*, 321–347. Amsterdam: John Benjamins. DOI: 10.1075/la.110.14bon.

Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

Czochralski, Jan A. 1975. *Verbalaspekt und Tempussystem im Deutschen und Polnischen: Eine konfrontative Darstellung*. Warszawa: Państwowe Wydawnictwo Nauk.

Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Blackwell.

Dickey, Stephen M. 2000. *Parameters of Slavic aspect: A cognitive approach*. Stanford: CSLI Publications.

Dostál, Antonín. 1954. *Studie o vidovém systému v staroslověnštině*. Praha: Státní pedagogické nakladatelství.

Dowty, David R. 1979. *Word meaning and Montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ*. Dordrecht: Reidel. DOI: 10.1007/978-94-009-9473-7.

Esvan, François. 2007. *Vidová morfologie českého slovesa*. Praha: Nakladatelství Lidové noviny.

Ferreira, Marcelo. 2016. The semantic ingredients of imperfectivity in progressives, habituals, and counterfactuals. *Natural Language Semantics* 24. 353–397. DOI: 10.1007/s11050-016-9127-2.

Filip, Hana. 1999. *Aspect, eventuality types and noun phrase semantics*. New York: Garland Publishing. DOI: 10.4324/9780203827413.

Filip, Hana. 2000. The quantization puzzle. In Carol Tenny & James Pustejovsky (eds.), *Events as grammatical objects: The converging perspectives of lexical semantics and syntax*, 39–95. Stanford: CSLI Publications.

Filip, Hana. 2003. Prefixes and the delimitation of events. *Journal of Slavic Linguistics* 11(1). 55–101. https://www.jstor.org/stable/24599705.

Filip, Hana. 2005. Measures and indefinites. In Gregory N. Carlson & Francis Jeffry Pelletier (eds.), *Reference and quantification: The Partee effect*, 229–288. CSLI Publications. DOI: 10.1007/1-4020-3232-3_7.

Filip, Hana & Gregory N. Carlson. 1997. Sui generis genericity. *University of Pennsylvania Working Papers* 4. 91–110. https://repository.upenn.edu/handle/20.500.14332/45451.

Flier, Michael S. 1985. The scope of prefixal delimitation in Russian. In Michael S. Flier & Alan Timberlake (eds.), *The scope of Slavic aspect*, 41–58. Columbus: Slavica Publishers.

Gehrke, Berit. 2008. *Ps in motion: On the semantics and syntax of P elements and motion events*. Utrecht: LOT. https://www.lotpublications.nl/ps-in-motion-ps-in-motion-on-the-semantics-and-syntax-of-p-elements-and-motion-events.

Gribanova, Vera. 2015. Exponence and morphosyntactically triggered phonological processes in the Russian verbal complex. *Journal of Linguistics* 51(3). 519–561. DOI: 10.1017/S0022226714000553.

Grønn, Atle. 2004. *The semantics and pragmatics of the Russian factual imperfective*. Oslo: University of Oslo. (Doctoral dissertation).

Haider, Hubert & Luka Szucsich. 2022. Slavic languages: "SVO" languages without SVO qualities? *Theoretical Linguistics* 48(1–2). 1–39. DOI: 10.1515/tl-2022-2035. https://doi.org/10.1515/tl-2022-2035.

Isačenko, Aleksandr V. 1960. *Grammatičeskij stroj russkogo jazyka. Morfologija. Častj vtoraja [Grammatical structure of the Russian verb. Morphology. Part two]*. Bratislava: Vydavatelstvo Slovenskej Akademie vied.

Isačenko, Aleksandr V. 1962. *Die russische Sprache der Gegenwart. Teil 1: Formenlehre*. Halle (Saale): Max Niemeyer.

Jabłońska, Patrycja. 2004. When the prefixes meet the suffixes. *Nordlyd* 32(2). 363–401. DOI: 10.7557/12.73.

Kagan, Olga. 2015. *Scalarity in the verbal domain: The case of verbal prefixation in Russian*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781316136195.

Kagan, Olga. 2016. Measurement across domains: A unified account of the adjectival and the verbal attenuative *po-*. *Journal of Slavic Linguistics* 24(2). 301–336. DOI: 10.1353/jsl.2016.0013.

Karcevski, Serge. 1927. *Système du verbe russe: Essai de linguistique synchronique.* Praha: Plamja.

Karlík, Petr, Marek Nekuta & Zdenka Rusínová (eds.). 1995. *Příruční mluvnice češtiny.* Praha: Nakladatelství Lidové noviny.

Katny, Andrzej. 1994. *Zu ausgewählten Aktionsarten im Polnischen und deren Entsprechungen im Deutschen.* Rzesów: Wydawnictwo WSP.

Klimek-Jankowska, Dorota & Joanna Błaszczak. 2022. The status of secondary imperfectivization in Polish: Evidence from VP idioms. *Journal of Slavic Linguistics* 30(FASL 29 issue). 1–19. https://ojs.ung.si/index.php/JSL/article/view/99.

Klimek-Jankowska, Dorota & Joanna Błaszczak. 2024. The interaction of idioms and aspect in Polish. *Natural Language & Linguistic Theory* 42(1). 153–207. DOI: 10.1007/s11049-023-09577-2.

Kopečný, František. 1962. *Slovesný vid v češtině.* Praha: Nakladatelství Československé akademie věd.

Kratzer, Angelika. 2008. On the plurality of verbs. In Johannes Dölling, Tatjana Heyde-Zybatow & Martin Schäfer (eds.), *Event structures in linguistic form and interpretation*, 269–300. Chicago: De Gruyter. DOI: 10.1515/9783110925449.269.

Krifka, Manfred. 1992. Thematic relations as links between nominal reference and temporal constitution. In Ivan A. Sag & Anna Szabolcsi (eds.), *Lexical matters*, 29–53. Stanford: CSLI Publications.

Kwapiszewski, Arkadiusz. 2022. *Aspect and event structure: The morphosyntax of Polish verbs from a cross-linguistic perspective.* Oxford: University of Oxford. (Doctoral dissertation).

Lasersohn, Peter. 1995. *Plurality, conjunction and events.* Dordrecht: Kluwer. DOI: 10.1007/978-94-015-8581-1.

Łaziński, Marek. 2020. *Wykłady o aspekcie polskiego czasownika.* Warszawa: Wydawnictwo Uniwersytetu Warszawskiego. DOI: 10.31338/uw.9788323542346.

Łazorczyk, Agnieszka A. 2010. *Decomposing Slavic aspect: The role of aspectual morphology in Polish and other Slavic languages.* Los Angeles: University of Southern California. (Doctoral dissertation).

Link, Godehard. 1983. The logical analysis of plural and mass terms: A lattice theoretical approach. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, use, and interpretation of language*, 302–323. Berlin: de Gruyter. DOI: 10.1515/9783110852820.302.

Maienborn, Claudia. 2003. *Die logische Form von Kopula-Sätzen.* Berlin: Akademie-Verlag. DOI: 10.1524/9783050082271.

Maienborn, Claudia. 2005. On the limits of the Davidsonian approach: The case of copula sentences. *Theoretical Linguistics* 31(3). 275–316. DOI: 10.1515/thli.2005.31.3.275.

Marušič, Franc. 2023. Circumfixation. In Peter Ackema, Sabrina Bendjaballah, Eulàlia Bonet & Antonio Fabregas (eds.), *The Wiley Blackwell companion to morphology*. Hoboken: John Wiley & Sons. DOI: 10.1002/9781119693604.morphcom013.

Matushansky, Ora. 2009. On the featural composition of the Russian back yer. In Gerhild Zybatow, Uwe Junghanns, Denisa Lenertová & Petr Biskup (eds.), *Studies in formal Slavic phonology, morphology, syntax, semantics and information structure: Proceedings of FDSL 7, Leipzig 2007*, 397–410. Frankfurt am Main: Peter Lang.

Mehlig, Hans Robert. 2006. Glagol'nyj vid i vtoričnaja gomogenizacija oboznačaemoj situacii posredstvom kvantifikacii: K upotrebleniju delimitativnogo sposoba dejstvija v russkom jazyke. In Volkmar Lehmann (ed.), *Semantika i struktura slavjanskogo vida IV* (Slavolinguistica 7), 235–276. München: Sagner.

Mehlig, Hans Robert. 2007. Aspect and bounded quantity complements in Russian. In Susan Rothstein (ed.), *Theoretical and Crosslinguistic Appoaches to the Semantics of Aspect*, 257–290. Amsterdam: John Benjamins. DOI: 10.1075/la.110.11meh.

Mehlig, Hans Robert. 2012. Hybrid predications in Russian. *Journal of Slavic Linguistics* 20(2). 171–227. DOI: 10.1353/jsl.2012.0010.

Nádeníček, Petr. 2011. *Das tschechische Aspektsystem im Vergleich mit dem Verbalaspekt des Russischen und Polnischen*. Hamburg: Verlag Dr. Kovač.

Nübler, Norbert. 2017. Iterativnost. In Petr Karlík, Marek Nekula & Jana Pleskalová (eds.), *Czechency – Nový encyklopedický slovník češtiny*. https://www.czechency.org/slovnik/ITERATIVNOST.

Paslawska, Alla & Arnim von Stechow. 2003. Perfect readings in Russian. In Artemis Alexiadou, Monika Rathert & Arnim von Stechow (eds.), *Perfect explorations*, 307–362. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110902358.307.

Petr, Jan. 1986. *Mluvnice češtiny 1: Fonetika – fonologie – morfonologie a morfemika – tvoření slov*. Praha: Academia.

Petruxina, Elena V. 2011. *Aspektual'nye kategorii glagola v russkom jazyke v sopostavlenii s češskim, slovackim, pol'skim i bolgarskim jazykami*. Second edition. Moskva: URSS.

Piñón, Christopher. 1994. Aspectual composition and the 'pofective' in Polish. In Sergey Avrutin, Steven Franks & Liljana Progovac (eds.), *Formal approaches to Slavic linguistics: The MIT meeting 1993 (FASL 2)*, 341–373. Ann Arbor: Michigan Slavic Publications.

Ramchand, Gillian Catriona. 2004. Time and the event: The semantics of Russian prefixes. *Nordlyd* 32(2). 23–361. DOI: 10.7557/12.72.

Romanova, Eugenia. 2004. Superlexical versus lexical prefixes. *Nordlyd* 32(2). 255–278. DOI: 10.7557/12.69.

Schoorlemmer, Maaike. 1995. *Participial passive and aspect in Russian.* Utrecht: Utrecht University. (Doctoral dissertation).

Schuyt, Roel. 1990. *The morphology of Slavic verbal aspect: A descriptive and historical study.* Amsterdam: Rodopi. DOI: 10.1163/9789004654044.

Smith, Carlota S. 1997. *The parameter of aspect.* 2nd edition. Dordrecht: Springer. DOI: 10.1007/978-94-011-5606-6.

Součková, Kateřina. 2004a. *Measure prefixes in Czech: Cumulative* na- *and delimitative* po-. Tromsø: University of Tromsø. (MA thesis).

Součková, Kateřina. 2004b. There is only one *po-*. *Nordlyd* 32(2). 43–419. DOI: 10.7557/12.74.

Švedova, Natalija J. 1980. *Russkaja grammatika.* Moskva: Nauka.

Tatevosov, Sergei. 2008. Intermediate prefixes in Russian. In Andrei Antonenko, John F. Bailyn & Christina Y. Bethin (eds.), *Formal Approaches to Slavic Linguistics 16: The Stony Brook Meeting 2007*, 423–445. Ann Arbor: Michigan Slavic Publications.

Tatevosov, Sergei. 2009. Množestvennaja prefiksacija i anatomija russkogo glagola. In Ksenia Kisileva, Vladimir Plungjan, Ekaterina Raxilina & Sergej [=Sergei] Tatevosov (eds.), *Korpusnye issledovanija po russkoj grammatike*, 92–156. Moskva: Probel-2000.

Tatevosov, Sergei. 2011. Severing perfectivity from the verb. *Scando-Slavica* 57. 216–244. DOI: 10.1080/00806765.2011.631782.

Tatevosov, Sergei. 2015. Severing imperfectivity from the verb. In Gerhild Zybatow, Petr Biskup, Marcel Guhl, Claudia Hurtig, Olav Mueller-Reichau & Maria Yastrebova (eds.), *Slavic grammar from a formal perspective: The 10th Anniversary FDSL Conference*, 465–494. Frankfurt am Main: Peter Lang. DOI: 10.3726/978-3-653-05335-7.

Tatevosov, Sergei. 2020. *On the aspectual architecture of Russian.* Ms. Lomonosov Moscow State University. http://otipl.philol.msu.ru/staff/people/tatevosov/aspectless_verb_3.1.pdf.

Tatevosov, Sergei & Mikhail Ivanov. 2009. Event structure of non-culminating accomplishments. In Lotte Hogeweg, Helen de Hoop & Andrej Malchukov (eds.), *Cross-linguistic semantics of Tense, Aspect, and Modality*, 83–130. Amsterdam: John Benjamins. DOI: 10.1075/la.148.05tat.

Timberlake, Alan. 2004. *A reference grammar of Russian.* Cambridge: Cambridge University Press. DOI: 10.1007/s11185-006-0705-7.

Vinogradov, Viktor V. 1947. *Russkij jazyk: Grammatičeskoe učenie o slove.* Moskva: Izdat. "Vysšaja Škola".

Wood, Esther J. 2007. *The semantic typology of pluractionality.* Berkeley: University of California. (Doctoral dissertation).

Zaliznjak, Anna A. 2017. Russian prefixed verbs of motion revisited. *Zeitschrift für Slawistik* 62(1). 1–28. DOI: 10.1515/slaw-2017-0001.

Zaliznjak, Anna A. & Aleksej D. Šmelev. 1997. *Lekcii po russkoj aspektologii.* München: Verlag Otto Sagner. DOI: 10.3726/b12573.

Zinova, Yulia. 2021. *Russian verbal prefixation: A frame semantic analysis.* Berlin: Language Science Press. DOI: 10.5281/zenodo.4446717.

Zinova, Yulia & Hana Filip. 2015. The role of derivational history in aspect determination. In Gerhild Zybatow, Petr Biskup, Marcel Guhl, Claudia Hurtig, Olav Mueller-Reichau & Maria Yastrebova (eds.), *Slavic grammar from a formal perspective: The 10th anniversary FDSL conference*, 595–609. Frankfurt am Main: Peter Lang. DOI: 10.3726/978-3-653-05335-7.

Żmigrodzki, Piotr. 2022. Wielki słownik języka polskiego PAN jako słownik tematyczny. *Studia Językoznawcze: Synchroniczne i diachroniczne aspekty badań polszczyzny* 21. 153–166. DOI: 10.18276/sj.2022.21-11.

Zucchi, Sandro. 1999. Incomplete events, intensionality and imperfective aspect. *Natural Language Semantics* 7. 179–215. DOI: 10.1023/A:1008310800632.

Chapter 4

Bare nouns in Slavic and beyond

Olga Borik[a], Bert Le Bruyn[b], Jianan Liu[b] & Daria Seres[c]

[a]Universidad Nacional de Educacion a Distancia [b]Utrecht University
[c]University of Graz

The article presents a study of the distribution of singular bare nouns in three Slavic languages, Russian, Polish and Macedonian, based on parallel translation corpora. The distribution of bare singulars in Russian and Polish shows that they freely appear in definite and indefinite contexts, which makes it possible to classify these languages as truly articleless. Macedonian bare singulars frequently appear in indefinite contexts, alongside with nouns accompanied by an indefinite marker *one*, whose status require further scrutiny. The data reported in this study call for a theoretical account of bare nouns which allows for fine-grained variation in their distribution across domains and languages, taking into consideration a broader/ narrower use of competing expressions.

1 Introduction

Referring is one of the main functions of natural language, and speakers of different languages use a variety of linguistic means and mechanisms to express different types of reference. In the empirical study that we present here, we focus on the syntax–semantics interface of bare nouns (BNs) and examine their distributional properties in Russian, Polish and Macedonian, languages that belong to the East Slavic, West Slavic and South Slavic subgroups, respectively. In particular, we address the issue of a comparative distribution of bare singular nouns (BSs) in the definite and indefinite domain across the three languages.

In terms of definiteness/indefiniteness marking, Russian and Polish are typically classified as articleless languages (Dryer & Haspelmath 2013), that is, having no dedicated morphosyntactic marker to express definiteness or indefiniteness.

We thus expect nominals to appear in their bare form in all argument positions in both languages. This straightforward expectation is in line not only with the traditional descriptive grammars, such as Švedova (1980), but also with some formal semantic literature, such as Chierchia (1998), Geist (2010), among others. Other formal approaches, most notably Dayal (2004, 2018), argue that number plays a crucial role in the distribution of BNs in articleless languages, making different predictions for bare plurals (henceforth BPs) and BSs. In particular, Dayal (2004) argues that BSs do not get an indefinite interpretation in languages without articles, while BPs can get narrow scope indefinite readings. Therefore, BSs are predicted to be largely restricted to definite contexts. Our focus on BSs allows us to check the predictions made by Dayal's theory as opposed to more traditional approaches.

Macedonian is usually described as a language with a definite article (Friedman 1993, Tomić 2006, among many others). The definite article in Macedonian is postpositive and morphologically bound. It is typically added to the first element of a nominal phrase[1] (e.g., *kuče-to* 'the dog', *ubavo-to kuče* 'the beautiful dog'), and is inflected for number and gender (e.g., *maž-ot* M.SG 'the man', *žena-ta* F.SG 'the woman', *dete-to* N.SG 'the child', *maži-te/ženi-te* M/F.PL 'the men/the women', *deca-ta* N.PL 'the children'). BSs are also admissible in argument positions in Macedonian, while there is no agreement on their interpretation in the literature (Weiss 2004, Topolinjska 2009, among others). The most widely accepted assumption is that BSs in Macedonian appear in indefinite contexts, although it has also been noticed that a determiner *eden* 'one' is often used to mark indefiniteness in this language (Tomić 2006).

Our study aims at answering the following research questions:

1. What is the distribution of BSs in Russian and Polish as languages without articles? Do they appear in both definite and indefinite domains or do we observe significant differences in the distribution of BSs across domains?

2. What is the status of BSs in Macedonian in indefinite contexts as compared to Russian and Polish?

3. What is the status of BSs in Macedonian in definite contexts?

To address these questions, we ran a parallel corpus study to analyze nominal phrases that appear in both definite and indefinite contexts in the three languages, with a critical look at the distribution of BSs in each of the domains.

---

[1]We use the term *nominal phrase* to abstract away from the DP/NP debate, prominent mostly in the syntactic literature on Slavic. See, for instance, Bošković (2008).

For the definite domain, the expectations are rather straightforward: both traditional descriptive and formal literature seem to converge on the idea that BSs freely appear in definite contexts in Russian and Polish, whereas in Macedonian we expect the definite article to dominate. However, the status of the definite marker as an article in Macedonian is not uncontroversial: Rudin (2021: 313), for instance, suggests that it might be a type of demonstrative rather than an article. Semantic literature repeatedly stresses similarities between demonstrative NPs and definite descriptions (e.g., Roberts 2002, Elbourne 2008), as well as the need to differentiate between the two (Lyons 1999). We include demonstrative nominals in our empirical study and look at the relative distribution of NPs specified by demonstratives vs. definite nominal phrases in Macedonian or BSs in Russian and Polish in the definite domain.

For the indefinite domain, existing analyses diverge when it comes to predictions. Traditional descriptions do not report any irregularities or asymmetries in the distribution of BSs across definite vs. indefinite domains, so they seem to predict that BSs can freely appear in indefinite contexts. However, claims have been made that in Polish and Macedonian, the indefinite marker ONE[2] is acquiring (or has acquired) the status of an indefinite article (Hwaszcz & Kędzierska 2018, Molinari 2022 with reference to Polish; Tomić 2006 with reference to Macedonian). The prediction that these proposals make is that the marker ONE will frequently appear in the indefinite domain in these languages, competing with or prevailing over BSs. The same prediction is made by Dayal (2004, 2018), who takes Hindi as a representative example of an articleless language and argues that it typically resorts to a construction with ONE in those contexts where English uses the indefinite article. Applying Dayal's analysis to Russian and Polish,[3] we expect BSs in these languages to be severely restricted in the indefinite domain, as opposed to the construction with ONE, which should dominate. In other analyses, the marker ONE in Russian is assumed to mark specificity rather than function as an indefinite article (Ionin 2013), which predicts its appearance only in specific indefinite contexts, converging with the predictions of Geist (2010), who argues that BSs in Russian can only get a non-specific reading.[4]

To get a broader cross-linguistic perspective, we compare parallel-corpus data for Russian, Polish and Macedonian to corpus results for Mandarin and German,

---

[2]The English ONE is used as a cover term for language specific *odin* (Russian), *jeden* (Polish) and *eden* (Macedonian) and their respective forms.

[3]Dayal does discuss Russian, and we assume that the proposal extends to other languages without articles like Polish, as it is based on general, language-independent semantic principles and mechanisms.

[4]Geist's (2010) predictions should be relativized to the information structure since she argues that indefinite BSs cannot serve as aboutness topics.

two non-Slavic languages. Mandarin functions as a control language for Russian and Polish, as it is usually assumed to be an articleless language (Li 2021), whereas German functions as a control language for Macedonian, as both have a definite article.

In order to investigate the distribution of BSs in definite and indefinite contexts in the three Slavic languages we ran a parallel-corpus study, described in detail in §2. We present the results of our study in §3, followed by a general discussion in §4. §5 concludes the paper.

2 Data and methodology

We use parallel corpora to study the distribution of grammatical items in different languages in parallel, an approach that has recently gained traction in the formal literature for the study of a variety of empirical domains, for example, tense and aspect (see – among others – Fuchs & González 2022; Gehrke 2022; Mo 2022; Mulder et al. 2022), negation (de Swart 2020) and reference (Bremmers et al. 2021). Parallel corpus research builds on the assumption that the meanings of the original and the translations are as closely related to each other as the grammars of the respective languages allow them to be. Another important assumption is that translations are representative of their target languages (*the target language representativeness hypothesis*). For a more detailed discussion of the methodology and its caveats, see Le Bruyn et al. (2022), Le Bruyn & de Swart (2022).

This study uses a translation corpus built on the first chapter of J. K. Rowling's *Harry Potter and the Philosopher's Stone*, a novel written in English and translated into many typologically diverse languages. English grammatically marks the distinction between definiteness and indefiniteness, which allows us to easily detect all definite and indefinite referential expressions in the source text. We selected all (in)definite referential expressions (*a N, the N, N-s, the N-s*) with their aligned translations in Russian, Polish, Macedonian, Mandarin and German (n=284) and manually annotated the corresponding NP forms in all the target languages.[5] At this point, it is important to emphasize that our methodology involves the annotation of forms (but not meanings) in the same contexts across the languages under study.

---

[5]Because some referential expressions are not translated and because of issues of automatic alignment, some data are literally lost in translation. Our dataset for this study includes referential expressions that have translations in all five languages under scrutiny. These numbers are expected not to be identical to the ones in Liu et al. (2023), a study that we conducted for a wider set of languages using the same methodology.

Since this paper focuses on the singular domain, we limit our quantitative analysis to the singular paradigm only.[6] Apart from theoretical reasons discussed in §1, plurals were excluded due to their relatively low frequency in our dataset and the interaction of plural definites with proper names (e.g., *The Potters, The Dursleys*). Thus, our final dataset includes the translations of *a N* (n=82) and *the N*$_{\text{sing}}$ (n=124) constructions into Russian, Polish, Macedonian, as well as Mandarin and German, which are used as control languages in this study.

An example of an English source *the N* expression (1a) and its translations from the parallel corpus are shown below.

(1)   a.   Mr Dursley might have been drifting into an uneasy sleep, but *the cat* on the wall outside was showing no sign of sleepiness.

    b.   Dolgoždannyj i nespokojnyj son uže prinjal v svoi ob"jatija mistera Darsli, a sidevšaja na ego zabore *koška* spat' soveršenno ne sobiralas'.
<div align="right">Russian [N]</div>

    c.   Pan Dursley zapadł w niezbyt zresztą spokojny sen, ale *kot* na murku nie okazywał najmniejszych oznak senności.   Polish [N]

    d.   Gospodinot Darsli možebi potona vo nemiren son, no *mačkata* na dzidot nadvor ne pokažuvaše ni troška sonlivost.
<div align="right">Macedonian [N+the]</div>

    e.   Mr Dursley mochte in einen unruhigen Schlaf hinübergeglitten sein, doch *die Katze* draußen auf der Mauer zeigte keine Spur von Müdigkeit.   German [the N]

    f.   Désīlǐ xiānshēng mímíhúhú, běnlái kěnéng húluàn shuì-shàng yí jiào, kě huāyuán qiángtóu shàng *nà zhī māo* què méiyǒu sīháo shuìyì.
<div align="right">Mandarin [demonstrative+classifier+N]</div>

In the definite domain, we examine the forms that Russian, Polish and Macedonian use for the translation of the English *the N*. In particular, we check whether and to what extent BSs that we expect to find in Russian and Polish, and singular definites that we expect to find in Macedonian, interact with demonstratives in the definite domain. We then contrast the results obtained for the three Slavic languages with Mandarin (as a control for Russian and Polish) and German (as a control for Macedonian).

In the indefinite domain, we look to determine which forms are used for the translations of the English *a N* in all three languages. We evaluate to which extent BSs are used in singular indefinite contexts in Russian vs. Polish vs. Macedonian and check for the interactions with the forms using the marker ᴏɴᴇ. Once

---

[6]Although there will be a short discussion of plurals in §4.1.2.

again, we compare the results obtained for Slavic languages with the results for Mandarin in the indefinite domain.

3 Results

3.1 Singular definite contexts

As far as definite contexts are concerned, there are no major surprises found in our data. The overall results are presented in Figure 1, which reflects absolute frequencies and includes all translations in the target languages. The category *Rest* contains all those translations that do not present any immediate interest for us (e.g., pronouns, possessives, etc.).

As we can observe, BSs are, indeed, the default option for rendering English *the N* both in Russian and in Polish, as shown in Figure 1. The differences in the occurrence of bare nominals in definite contexts are not significant for these two languages ($p = 0.37$, Fisher's Exact Test (FET)). Regarding Macedonian, the most prominent form in singular definite contexts is the one with a definite article, a result which is also fully in accordance with our initial expectations. In all three languages, there are practically no demonstratives used in singular definite contexts.

Comparing the results of Russian and Polish with their control language, Mandarin, we see that BNs in Mandarin are the most frequent form in the definite context as well. However, we also observe an important difference in the relative distribution of BNs vs. NPs specified by demonstratives in the definite domain: in Mandarin, the tendency to resort to demonstratives is higher. The differences are significant for the comparison of Mandarin and Polish ($p < 0.001$, FET), and for Mandarin and Russian ($p = 0.016$, FET).

As for Macedonian and its control language German, the two languages are quite uniform in the distribution of nominal forms in the definite domain. BSs and demonstrative NPs are either absent or clearly outnumbered in singular definite contexts in both Macedonian and German.[7] In §4.1.1, we will come back to the issue of definiteness marking in Macedonian and discuss some of the examples with BSs.

---

[7]More specifically, there is only one BS found in German and three in Macedonian.
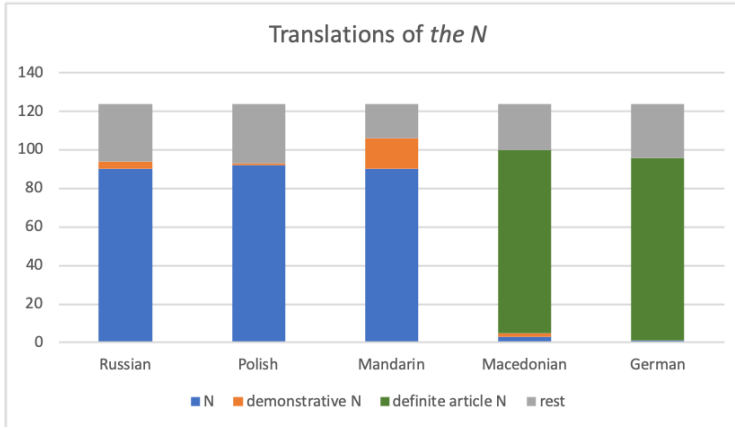
Figure 1: Russian, Polish and Mandarin BNs vs. demonstrative-N; Macedonian and German the-N vs. demonstrative-N

3.2 Singular indefinite contexts

The parallel-corpus data showed that a bare noun is the default option for rendering singular indefinite nominals in both Russian and Polish (see Figure 2). These two languages do not use the ONE+N construction in indefinite contexts in a statistically relevant way. The differences in distribution of bare nominals and nominals preceded by ONE are not significant for Russian and Polish ($p = 0.5$ FET).

In Macedonian, however, while a BS is still the most frequent form in the indefinite domain, the English *a N* construction is more often translated with the numeral ONE than in Russian or Polish. The differences are significant for the comparison of both Macedonian and Russian ($p < 0.001$, FET), and Macedonian and Polish ($p < 0.001$, FET).

As for the control language, Mandarin, where the numeral ONE precedes the nominal in a large number of cases in indefinite contexts, it shows a sharp contrast with Russian and Polish, which hardly ever use this structure. Moreover, Mandarin also shows contrast with Macedonian, where the use of ONE is not as frequent. The differences are significant for Mandarin and Russian ($p < 0.001$, FET), and Mandarin and Polish ($p < 0.001$, FET), as well as for Macedonian and Mandarin ($p < 0.001$, FET).
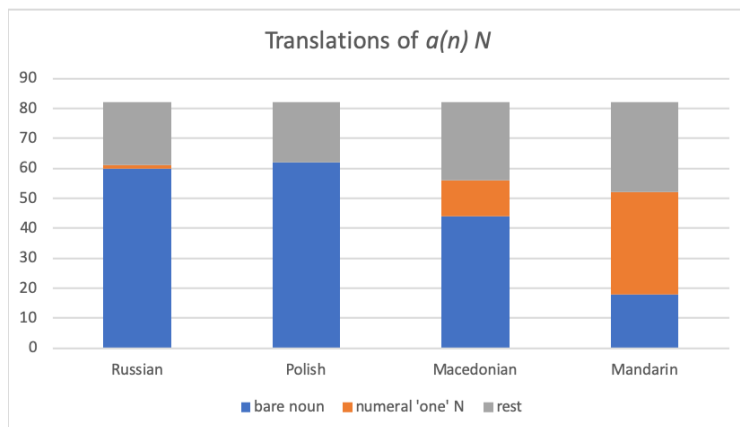
Figure 2: Russian, Polish, Macedonian and Mandarin bare nominals vs. ᴏɴᴇ+N

3.3 Recap

Summing up the results of our parallel corpus study, it can be said that Russian and Polish freely use bare nouns in both singular indefinite and singular definite contexts, in accordance with the Slavic descriptive literature. They are, however, in sharp contrast with Mandarin, where the numeral ᴏɴᴇ seems to be the default option in the indefinite domain and the demonstrative is competing with bare NPs in the definite domain.

As for Macedonian, in the indefinite domain it seems to occupy an intermediate position between Russian and Polish, on the one hand, and Mandarin on the other: The ᴏɴᴇ+N construction appears in the translations of *a N* quite frequently, but not as often as in Mandarin. In the definite domain, Macedonian uses NPs with a definite marker in *the N$_{sing}$* contexts as often as German.

4 Discussion

We structure our main discussion points in the same way we presented the results of the study, that is, according to the distribution of various forms in a specific context. We begin by evaluating the parallel corpus results obtained for the definite domain. In the discussion of the indefinite domain, we reflect not only on the distribution of BSs, but also on the role of ᴏɴᴇ+N construction in the grammar of all three target languages.

4.1 Definite contexts in Russian, Polish and Macedonian

Both traditional descriptions and formal semantic analyses seem to be in full agreement on attributing a possible definite reading to BSs in languages without articles. Our data cast no doubt on this claim for Russian or Polish: BSs prevail in definite contexts in both languages and the distributional behaviour of BSs is therefore in full accordance with their standard semantic descriptions and/or analyses. In Macedonian, the prevailing form is the definite singular, i.e., our data also confirm the status of Macedonian as a language with a definite article. Even though the interpretation of our main results seems to be rather straightforward, there are two points of interest that we would like to discuss.

The first observation concerns the distribution of BSs vs. NPs specified by demonstratives in the definite domain. In the previous section, we pointed out that demonstratives do not seem to occupy a prominent place in either of the three Slavic languages in *the* N$_{sing}$ contexts. Russian and Polish as languages without articles can be contrasted to Mandarin in this respect, one of the two control languages used in this study, where the higher rate of demonstratives in the definite domain suggests that the demonstrative plays a much more significant role in definiteness marking in Mandarin. In fact, Liu et al. (2023) hypothesise that Mandarin is developing a definite article (and an indefinite one), but in this paper, we limit ourselves to empirical statements with respect to Mandarin. In Macedonian, there were only three contexts where a demonstrative was used, all corresponding to anaphoric uses of *the N* in the source text.

The second point that we would like to discuss is the status of the definite article in Macedonian, which will be examined in the next subsections.

4.1.1 The Macedonian definite article in the singular domain

The status of the Macedonian definite article has been subject to some debate in the semantic literature, as pointed out in §1. No consensus emerges from the literature concerning the semantic contribution of this marker. One of the features that it exhibits (and that distinguishes it from a typical definite article) is that it morphologically marks a proximal–neutral–distal distinction, just like demonstratives in many languages do (Lyons 1999). It should be noted that the question about the status of a definite marker in any language is essentially semantic and cannot be definitively resolved without looking into the meaning of this expression, but the distribution of any definite marker/article also plays a significant role in a potential analysis and this is what our study can inform about.

We looked into the properties of the definite article in Macedonian by running a comparative study of the distribution of the definite article in Macedonian and

German. In particular, we measured their co-occurrence in the same contexts by calculating Normalized Pointwise Mutual Information (NMPI, Bouma 2009), which is a bidirectional measure for parallel data (Le Bruyn et al. 2022). The result shows that the NPMI of the two articles reaches 0.48 (with a maximum of 1). That means that the likelihood of the articles in the two languages occurring in the same contexts is higher than chance but not at ceiling. In other words, the bi-directional mapping pattern suggests that the distribution of definite articles in German and Macedonian across the definite contexts is not completely identical.

Table 1: Bi-directional mapping patterns between the German and Macedonian (singular) definite article

| | | Macedonian | | |
		definite article N	rest	
German	definite article N	88	24	112
	rest	20	74	94
		108	98	206

Looking into the contexts where Macedonian and German did not coincide in the use of the definite article, we find some interesting examples of BSs. For instance, in (2), Macedonian uses a BS, while German opts for a definite article, just like the English source:

(2) a. At half past eight, Mr Dursley picked up his briefcase, pecked Mrs Dursley on *the cheek* and tried to kiss Dudley goodbye but missed...
                                                            English (source): [the N]

     b. Vo osum i pol gospodinot Darsli ja zede svojata aktenčanta, ja kolvna gospoǵáta Darsli vo *obraz* i se obide da go bakne Dadli za razdelba, no ne uspea...   Macedonian: [N]

     c. Um halb neun griff Mr Dursley nach der Aktentasche, gab seiner Frau einen Schmatz auf *die Wange* und versuchte es auch bei Dudley mit einem Abschiedskuss...   German: [the N]

Another example of the same type of article mismatch is presented in (3):

(3) a. A man appeared on the corner the cat had been watching, appeared so suddenly and silently you'd have thought he'd just popped out of *the ground.*   English: [the N]

b. Na agolot što go nabljuduvaše mačkata se pojavi čovek, tolku nenadejno i tivko, kako da izniknal od *zemja*.   Macedonian [N]

c. An der Ecke, die sie beobachtet hatte, erschien ein Mann, so jäh und lautlos, als wäre er geradewegs aus *dem Boden* gewachsen.

German: [the N]

Although we cannot reach any firm conclusions on the basis of only few examples, we can hypothesise that they both present cases of weak definites (possessive weak definites, Barker 2005, in the case of (2)), so that the contexts where the uniqueness of a definite description is questioned are potentially very good candidates for the absence of the definite article in Macedonian. Needless to say, additional empirical investigation is needed to check this hypothesis.

4.1.2 Some remarks on the plural definite domain

Even though we did not run statistical analyses for the plural domain and there were not too many data points in our dataset, we would like to draw attention to some observations concerning the use of the definite article with plurals in Macedonian that appear important. For instance, Macedonian seems to use definite articles in plural generic contexts, while English resorts to bare plurals and German presents variation.

(4) a. *Cats* couldn't read maps or signs.   English (source): [Ns]

b. *Mačkite* ne možat da čitaat ni mapi ni oznaki.   Macedonian: [Ns+the]

c. *Katzen* konnten weder Karten noch Schilder lesen.   German: [Ns]

The use of the definite article with generic plurals as illustrated in (4) may suggest that Macedonian – at least in some aspects – is rather comparable to Romance languages in its use of definite plurals than to Germanic languages.

Existential contexts in the plural definite domain require further scrutiny. We detected several examples in our dataset where both Macedonian and German use a definite article whereas English uses a bare plural in the same context. One of those examples is (5).

(5) a. And finally, *bird-watchers* everywhere have reported that the nation's owls have been behaving very unusually today.   English: [Ns]

b. I konečno *nabljuduvačite* na ptici od site strani javija deka buvo - vite vo našata zemja deneska se ondesuvale mnogu neobično.

Macedonian: [Ns + the]

c. Und hier noch eine Meldung. Wie *die Vogelkundler* im ganzen Land berichten, haben sich unsere Eulen heute sehr ungewöhnlich verhalten.   German: [the Ns]

In this particular case, the presence of the article in Macedonian could be due to a specific syntactic construction used in the example (prepositional phrase *watchers of birds* instead of the nominal compound *bird-watchers* in the source text). This, however, would not explain the presence of the definite article in German. It might also be the case of a so-called FUNCTIONAL reading of BPs in English, discussed at length for English by Condoravdi (1994). The availability of this reading for BPs is language-specific, so we conclude that our data demonstrate some cross-linguistic variation worth a more systematic investigation. It is not surprising to see this variation in the distributional patterns, as cross-linguistic differences in the use of the definite article are very well documented and widely discussed in the literature. The corpus data of the current study is not sufficient to arrive at any firm conclusions, but it may be reasonably suggested that German and Macedonian, just like many other languages with grammatical marking of definiteness, do not fully coincide in definiteness marking patterns: the overlap in the use of the definite article is only partial, not absolute.

4.2 Indefinite contexts in Russian, Polish and Macedonian

Indefinite contexts constitute the most interesting case in our study, as they convincingly illustrate several theoretically relevant points. First, there is variation both within and outside the Slavic family in the distribution of BSs in the indefinite domain, which has direct repercussions for existing theoretical analyses of BSs. Second, intricate interactions of ONE+N with BSs in the indefinite domain can elucidate the grammatical status of ONE in a given language. Third, our data pose some very specific constraints and requirements for an accurate and empirically adequate theoretical analysis of BSs in languages without articles. We discuss each of these points in the three subsections that follow.

4.2.1 BSs in Russian, Polish and Macedonian

One of the main results of our study concerns the distributional pattern of BSs in Russian and Polish. In particular, the data from the parallel corpus show that in Russian and Polish BSs freely appear in indefinite singular contexts as counterparts of *a N* in the source text. One rather typical example of an indefinite in an existential context is given below:

(6)   a.   English (source): [a N]
           There was *a tabby cat* standing on the corner of Privet Drive, but
           there wasn't a map in sight.
      b.   Russian: [N]
           Na uglu Praivet Draiv dejstvitel'no sidela *polosataja koška*, no nikakoj
           karty vidno ne bylo.
      c.   Polish: [N]
           Na rogu Privet Drive rzeczywiście stał *bury kot*, ale nie studiował
           żadnej mapy.

Our Russian and Polish data directly support traditional descriptive approaches
to BSs in Slavic languages without articles and those formal approaches which
do not rule out an indefinite interpretation for BSs, e.g., Chierchia (1998), Krifka
(2003). The results of our study are also compatible with the proposal that
bare NPs in Russian are essentially indefinite and a definite reading is achieved
through pragmatic strengthening (Seres & Borik 2021).

On the other hand, our empirical findings are in conflict with Dayal's (2004)
proposal, whose prediction – as we mentioned in §1 – is that BSs should never
give rise to indefinite readings in regular argument position in languages without
articles. Dayal examines the behavior of BSs in Hindi, Russian and Mandarin, and
argues that an overt indefiniteness marker has to appear in those contexts where
an indefinite reading has to be expressed. This prediction holds for Hindi, where
ONE functions as such a marker,[8] but it is very clear that Russian and Polish
behave differently. In fact, in our data ONE is only used twice in Russian in the
indefinite domain, whereas the Polish data do not contain a single occurrence of
this item. Thus, our data allow us to conclude that both Russian and Polish are
truly articleless languages where BSs dominate in both definite and indefinite
contexts. No competing forms emerge in our study in either of the two contexts
in either of the two languages.

In contrast to Russian and Polish, Macedonian uses both BSs and ONE+N constructions. Our data show that Macedonian differs from truly articleless languages, and the construction ONE+N competes with BSs in the indefinite domain
in Macedonian. This difference can be illustrated with the translation of example (6a) above into Macedonian: where Russian and Polish use a BS, Macedonian
uses ONE+N.

(7)   Na agolot na Šimširovata uliča stoeše *edna neobična šarena mačka*, no
      nikade nemaše mapa.

---

[8]This result has been confirmed by a parallel corpus study reported in Liu et al. (2023).

If we look outside the Slavic family, our control language, Mandarin, shows a strong tendency for the ONE+N construction to appear in singular indefinite contexts (see Figure 2). Macedonian clearly occupies an intermediate position between Mandarin (relatively low percentage of BSs) and Russian/Polish (predominantly BSs) with respect to the use of BSs in the singular indefinite domain.

Note that this kind of variation in the use of BSs comes out unexpected on most analyses. In general, articleless languages are perceived as a homogeneous group that either do or do not use BSs in a certain domain, but the kind of variation that we see in our data is rather challenging for theoretical approaches. We will come back to this point at the end of this section, but first we will take a better look at the closest competitor of a BS in the indefinite domain, the indefinite marker ONE.

4.2.2 The status of ONE in the indefinite domain

It is well known that the numeral ONE is a predecessor of the indefinite article in many languages (Heine 1997, van Gelderen 2011, among many others). Looking once again at the distribution of nominal forms in the indefinite domain in Figure 2, we observe a clear interaction between BSs and ONE+N constructions: the frequency of ONE+N in our data goes from being at floor in Russian and Polish to a significant percentage in Macedonian and to predominance in Mandarin. This raises a question about the grammatical status of the marker ONE in different languages.

The differences in the use of the ONE+N construction across languages may be accounted for by different stages of its grammaticalisation as an article. Typically, the stages of grammaticalisation of the indefinite article are defined in the following order: 1. the numeral, 2. the presentative marker, 3. the specificity marker, 4. the non-specific marker, 5. the generalised article (Givón 1981, Heine 1997, among others).[9] Even though defining the exact stage of grammaticalisation of ONE in the languages under study is out of the scope of this paper, our data offer several discussion points relevant for the issue.

Our empirical findings for Russian and Polish, where BSs overwhelmingly dominate in the indefinite domain, seem to be in conflict with the proposal of Hwaszcz & Kędzierska (2018), who claim that in Russian ONE is grammaticalised as a presentative marker, that is, it marks a newly introduced referent, which is intended to be used in the subsequent discourse and is usually specific and topical. The authors also claim that in Polish ONE is grammaticalised even further, being

---

[9]These stages are coarsely defined and may have substages.

used as a specific and sometimes as a non-specific marker. Neither of the two claims is confirmed by our data, as some representative examples can illustrate:

(8) a. English (source): [a N]
         The Dursleys had *a small son* called Dudley and in their opinion there was no finer boy anywhere.
     b. Russian: [N]
         U mistera i missis Darsli byl *malen'kij syn* po imeni Dadli, i, po ix mneniju, ėto byl samyj čudesnyj rebenok na svete.
     c. Polish: [N]
         Syn Dursleyów miał na imię Dudley, a rodzice uważali go za najwspanialszego chłopca na świecie.

(9) a. English (source): [a N]
         He was sure there were lots of people called Potter who had *a son* called Harry.
     b. Russian: [N]
         Mister Darsli legko ubedil sebja v tom, čto v Anglii živet množestvo semej, nosjaščix familiju Potter i imejuščix *syna* po imeni Garri.
     c. Polish: [N]
         Mnóstwo ludzi może się nazywać Potter i mieć *syna* Harry'ego.

Example (8) is a typical context where a new specific referent is introduced by a modified indefinite in the source text, which is then rendered by a BS both in Russian and in Polish, just like the non-specific indefinite *a son* in (9). At least in Russian, ONE+N cannot be used instead of N in (8) and (9), unless ONE is interpreted as a numeral.[10] Our data show no sign of any significant difference between Russian and Polish with respect to the grammatical status of ONE: this marker does not show up regularly or systematically in either a presentative, specific or any other type of context.

 Macedonian ONE, on the other hand, is more frequent. We have not conducted any specific study of the contexts where ONE appears in Macedonian, as our dataset is too small to yield sensible results, but we can provide some indicative examples here that can help us map out a path for future research. For instance, Macedonian uses a BS in translations of both example (8) and (9) above, but there are other specific and non-specific contexts where ONE+N construction appears:

---

[10]We thank an anonymous reviewer for stressing this point.

(10) a. English (source): [a N]   (non-specific)
My dear Professor, surely *a sensible person* like yourself can call him
by his name?

b. Macedonian [one N]
Draga moja profesorke, ne misliš li deka *edna tolku razumna ličnost'*
kako što si ti slobodno može da go narekuva po ime?

(11) a. English (source): [a N]   (specific)
Professor McGonagall pulled out *a lace handkerchief* and dabbed at
her eyes beneath her spectacles.

b. Macedonian [one N]
Profesorkata Mekgonagl izvadi *edno tanteleno maramče* i gi protri
očite pod očilata.

The mixed data across specific and non-specific contexts indicate that the ONE+N
construction is not really established in these types of contexts. The data obtained in our study are, in principle, in line with Hwaszcz & Kędzierska (2018),
who claim that ONE in Macedonian is used with both specific and non-specific
indefinite NPs. Our Macedonian data show that both specific and non-specific indefinite NPs may also appear as bare, as illustrated in the above examples, which
may indicate a certain degree of optionality in the use of ONE for marking specific
and non-specific nominals.[11] This flexibility (possibly translated as optionality)
provides a contrast with English and German, languages where an indefinite article is obligatorily used in all the examples discussed in this subsection. Thus,
Macedonian does differ from languages with established indefinite articles, and
we therefore conclude this discussion by saying that the status of ONE cannot be
unequivocally defined as an indefinite article in Macedonian, contra, e.g., Tomić
(2006).[12] Rather, ONE is an indefinite marker that might evolve into an article,
but further research is needed to substantiate this claim.

4.2.3 Theoretical implications

As the discussion in the previous sections indicates, the main challenge that our
data pose for theoretical approaches striving for empirical adequacy is the problem of language variation. The variation in the definite domain, especially in the

---

[11]One of the limitations of corpus studies is that it is impossible to determine the optionality of
an element. In order to research the (non-)obligatoriness of ONE in certain linguistic environments, linguistic experiments with native speakers need to be carried out.

[12]In this respect, the Macedonian data resemble the situation in Bulgarian, as reported in Geist
(2013).

distribution of the definite article across languages, is relatively well known and discussed in the semantic literature (e.g., Dryer 2005). Our analysis of the definite article in Macedonian vs. German adds one more study case to this discussion.

In the indefinite domain, however, variation in the distribution of BSs in languages without articles (or without an indefinite article) is less expected. For instance, the approach to BNs in general and BSs in particular developed in Dayal (2004), Dayal (2018), and Dayal & Sağ (2020) is based on the claim that BSs do not allow for indefinite readings in articleless languages. The formal machinery of this approach does not leave much room for variation: the denotation of a noun in regular argument positions is derived by type-shifting operators and, crucially, Dayal's analysis cuts off the possibility of an existential type-shift for BSs. The logic behind this move, we believe, applies universally. Our data for Russian and Polish, though, strongly suggest that there should be an easy way to allow for a BS to appear in the singular indefinite domain, which may be achieved via standard type-shifting operations, like an existential type shift. However, allowing for this type shift to be subject to parametric variation will considerably weaken Dayal's formal theory, at least in the absence of any independent principle underlying such variation.

The Macedonian data, where we see a competition between BSs and the ONE+N construction, suggest that there should be a way to allow for BSs in those contexts where the other construction does not appear on a regular basis. In other words, there should be an account of an interaction between nominal forms that coexist in the indefinite domain. Dayal's approach cannot easily accommodate such interaction either, because ONE+N is predicted to be the only option in the indefinite domain in the absence of an indefinite article. Thus, we conclude that the semantic theory of bare nominals advocated in Dayal (2004), Dayal (2018), and Dayal & Sağ (2020) has considerable difficulties accounting for an overall empirical picture that emerges from our data.[13]

Mandarin, our control language, clearly prefers the ONE+N construction to BNs in the indefinite domain. As Liu et al. (2023) argue, this fact does not really follow from Dayal's analysis either, since in Mandarin, which lacks grammatical number, BNs are expected to easily get an indefinite reading, just like BPs in other languages do. If Mandarin BNs behave like BPs rather than BSs, they are predicted to get a narrow scope indefinite reading and hence, they should be visibly prominent in indefinite (singular and plural) contexts. In our data, however, the ONE+N construction wins over BNs in the singular indefinite domain. In fact, it looks like what Dayal (2004, 2018) predicts for Mandarin occurs in Russian and

---

[13]See also Liu et al. (2023).

Polish, with a proviso for number marking, and what her analysis predicts for Russian and Polish seems to hold for Mandarin.

An analysis that our data calls for should allow for a formal way to derive an existential interpretation of a BS via type-shifting, but only if there is no competing form with an overt marker that would block this shift. Chierchia's (1998) or Krifka's (2003) classical analyses, for instance, state that while in some languages type shifts are indicated by overt determiners, in languages that lack them, type shifts apply covertly whenever the linguistic context requires it. Covert type-shifting is restricted by the Blocking Principle, which roughly states that if a language has an overt means to express a type shift, then it must be used. This analysis seems to be much better equipped to handle our data. For instance, we have seen no evidence that ONE+N in Russian (*odin N*) and Polish (*jeden N*) function as an article-like expression. Thus, the covert application of the existential type-shift is not blocked, which allows for BSs to be freely used in indefinite contexts. For Macedonian, a language with an emerging indefinite marker ONE, the existential type shift would be blocked for a BN only in those contexts where *eden* appears. Our cross-linguistic data provide a serious argument in favour of a classical blocking semantic analysis of bare nominals, in which fine-grained variation in the distribution of bare nominals follows from the broader/narrower use of article-like expressions.

5 Conclusions

In this paper, we have reported the results of a parallel translation corpus study on the distribution of BSs in three Slavic languages, Russian, Polish and Macedonian. We built our corpus on the text of the first chapter of *Harry Potter and the Philosopher's Stone* and complemented the results obtained for Slavic languages with the results for Mandarin as a control language for Russian and Polish, and German as a control language for Macedonian.

In view of the empirical data presented here, it can be concluded that Russian and Polish are truly articleless languages and freely allow their BSs to take on definite and indefinite readings across domains. In Macedonian, BSs are restricted to the indefinite domain where they compete with the indefinite marker ONE, whereas in the definite domain, Macedonian uses the definite article, just as expected. Therefore, we conclude that Macedonian is a language with a definite article and with an emerging indefinite marker whose exact grammatical status requires further empirical investigation.

Slavic languages present challenging theoretically relevant contrasts with their control languages. In case of Macedonian, we have stressed the need to

further scrutinize the conditions and the contexts where the definite article is used because we have shown that the overlap between the definite articles in Macedonian and German is partial. We also see the need to extend the investigation to the plural domain to get a full picture of the distribution of the definite article in Macedonian. As for Russian and Polish, they present a striking contrast with Mandarin in the indefinite singular domain, where the two Slavic languages show a clear preference for BSs and Mandarin opts for the ONE+N construction as a counterpart of the English *a N*. Macedonian occupies an intermediate position: ONE+N is used rather frequently in Macedonian, but not as often as in Mandarin singular indefinite contexts.

We have argued that these contrasts call for a theoretical approach where the observed variation in the distribution of BSs and competing forms can be naturally accounted for. We suggest that the Blocking Principle as formulated in Chierchia (1998) can serve as a foundation for such an approach.

Abbreviations

F	feminine	PL	plural
M	masculine	SG	singular
N	neuter		

Acknowledgments

References

Barker, Christian. 2005. Possessive weak definites. English (US). In Kim, Ji-yung, Lander, Yury, Partee & H. Barbara (eds.), *Possessives and beyond: Semantics and syntax*. 89–113. Amherst, MA: GLSA Publications.

Bošković, Željko. 2008. What will you have, DP or NP? In Emily Elfner & Martin Walkow (eds.), *Proceedings of NELS 37*, 101–114. University of Illinois Urbana-Champaign.

Bouma, Gerlof J. 2009. Normalized (pointwise) mutual information in collocation extraction. In Christian Chiarcos, Richard Eckart de Castilho & Manfred Stede (eds.), *Proceedings of the Biennial International Conference of the German Society for CLLT*, 31–40. Tübingen: Gunter Narr Verlag.

Bremmers, David, Jianan Liu, Martijn van der Klis & Bert Le Bruyn. 2021. Translation mining: Definiteness across languages (A reply to Jenks 2018). *Linguistic Inquiry* 53(4). 735–752. DOI: 10.1162/ling_a_00423.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6(4). 339–405. DOI: 10.1023/a:1008324218506.

Condoravdi, Cleo. 1994. *Descriptions in context*. New Haven: Yale University. (Doctoral dissertation).

Dayal, Veneeta. 2004. Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy* 27(4). 393–450. DOI: 10.1023/b:ling.0000024420.80324.67.

Dayal, Veneeta. 2018. (In)definiteness without articles: Diagnosis, analysis, implications. In Ghanshyam Sharma & Rajesh Bhatt (eds.), *Trends in Hindi linguistics*, 1–26. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110610796-001.

Dayal, Veneeta & Yağmur Sağ. 2020. Determiners and bare nouns. *Annual Review of Linguistics* 6(1). 173–194. DOI: 10.1146/annurev-linguistics-011718-011958.

de Swart, Henriëtte. 2020. Double negation readings. In Viviane Déprez & M. Teresa Espinal (eds.), *The Oxford handbook of negation*, 479–496. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780198830528.013.26.

Dryer, Matthew S. 2005. Definite articles. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 154–157. Oxford: Oxford University Press.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *WALS online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/.

Elbourne, Paul D. 2008. Demonstratives as individual concepts. *Linguistics and Philosophy* 31(4). 409–466. DOI: 10.1007/s10988-008-9043-0.

Friedman, Victor A. 1993. Macedonian. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic Languages*, 249–305. London: Routledge.

Fuchs, Martín & Paz González. 2022. Perfect-perfective variation across Spanish dialects: A parallel-corpus study. *Languages* 7(3). 166. DOI: 10.3390/languages7030166.

Gehrke, Berit. 2022. Differences between Russian and Czech in the use of Aspect in narrative discourse and factual contexts. *Languages* 7(2). 155. DOI: 10.3390/languages7020155.

Geist, Ljudmila. 2010. Bare singular NPs in argument positions: Restrictions on indefiniteness. *International Review of Pragmatics* 2(2). 191–227. DOI: 10.1163/187731010x528340.

Geist, Ljudmila. 2013. Bulgarian *edin*: The rise of an indefinite article. In Uwe Junghanns, Dorothee Fehrmann, Denisa Lenertová & Hagen Pitsch (eds.), *Formal Description of Slavic Languages: The Ninth Conference. Proceedings of FDSL 9, Göttingen 2011*, 125–148. Frankfurt am Main: Peter Lang.

Givón, Talmy. 1981. On the development of the numeral 'one' as an indefinite marker. *Folia Linguistica Historica* 15(2-1). 35–54. DOI: doi:10.1515/flih.1981.2.1.35.

Heine, Bernd. 1997. *Cognitive foundations of grammar*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780195102512.001.0001.

Hwaszcz, Krzysztof & Hanna Kędzierska. 2018. The rise of an indefinite article in Polish: An appraisal of its grammaticalisation stage (Part 1). *Studies in Polish Linguistics* 13. 93–121. DOI: 10.4467/23005920SPL.18.005.8744.

Ionin, Tania. 2013. Pragmatic variation among specificity markers. In Cornelia Ebert & Stefan Hinterwimmer (eds.), *Different kinds of specificity across languages*, 75–103. Dordrecht: Springer. DOI: 10.1007/978-94-007-5310-5_4.

Krifka, Manfred. 2003. Bare NPs: Kind-referring, indefinites, both, or neither? In Robert B. Young & Yuping Zhou (eds.), *Proceedings of SALT 13*, 180–203. DOI: https://doi.org/10.3765/salt.v13i0.2880.

Le Bruyn, Bert & Henriëtte de Swart. 2022. *Cross-linguistic semantics: Methodological advances*. Course materials, ESSLLI. shorturl.at/doqV6.

Le Bruyn, Bert, Martín Fuchs, Martijn van der Klis, Jianan Liu, Chou Mo, Jos Tellings & Henriëtte de Swart. 2022. Parallel corpus research and target language representativeness: The contrastive, typological, and translation mining traditions. *Languages* (3). 176. DOI: 10.3390/languages7030176.

Li, Xuping. 2021. The Semantics of Chinese Noun Phrases. In *Oxford Research Encyclopedias: Linguistics*. Oxford: Oxford University Press. https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-886.

Liu, Jianan, Shravani Patil, Daria Seres, Olga Borik & Bert Le Bruyn. 2023. 'Articleless' languages are not created equal. In Maria Onoeva, Anna Staňková & Radek Šimík (eds.), *Proceedings of Sinn und Bedeutung 27*, 381–398. DOI: 10.18148/sub/2023.v27.1076.

Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511605789.

Mo, Chou. 2022. *The compositionality of Mandarin Aspect: A parallel corpus study*. Utrecht: University Utrecht. (Doctoral dissertation). https://www.lotpublications.nl/Documents/628_fulltext.pdf.

Molinari, Luca. 2022. The syntax of Polish *jeden* 'one' as an indefinite determiner. *Annali di Ca' Foscari. Serie occidentale* 56. 63–84. DOI: http://doi.org/10.30687/AnnOc/2499-1562/2022/10/004.

Mulder, Gijs, Gert-Jan Schoenmakers, Olaf Hoenselaar & Helen de Hoop. 2022. Tense and aspect in a Spanish literary work and its translations. *Languages* 7(3). 217. DOI: https://doi.org/10.3390/languages7030217.

Roberts, Craige. 2002. Demonstratives as definites. In Kees van Deemter & Roger Kibble (eds.), *Information sharing: Reference and presupposition in language generation and interpretation*, 89–136. Stanford: CSLI Press.

Rudin, Catherine. 2021. Demonstratives and definiteness: Multiple determination in Balkan Slavic. In Andreas Blümel, Jovana Gajić, Ljudmila Geist, Uwe Junghanns & Hagen Pitsch (eds.), *Advances in formal Slavic linguistics 2018* (Open Slavic Linguistics 4), 305–338. Berlin: Language Science Press. DOI: 10.5281/zenodo.5483114.

Seres, Daria & Olga Borik. 2021. Definiteness in the absence of uniqueness: the case of Russian. In Andreas Blümel, Jovana Gajić, Ljudmila Geist, Uwe Junghanns & Hagen Pitsch (eds.), *Advances in formal Slavic linguistics 2018* (Open Slavic Linguistics 4), 339–363. Berlin: Language Science Press. DOI: 10.5281/zenodo.5155544.

Švedova, Natalija J. 1980. *Russkaja grammatika*. Moskva: Nauka.

Tomić, Olga Mišeska. 2006. *Balkan Sprachbund morpho-syntactic features*. Dordrecht: Springer. DOI: 10.1007/1-4020-4488-7.

Topolinjska, Zuzanna. 2009. Definiteness (synchrony). In Sebastian Kempgen, Peter Kosta, Tilman Berger & Karl Gutschmidt (eds.), *Die slavischen Sprachen / the Slavic languages*, 176–188. Berlin: De Gruyter Mouton. DOI: doi:10.1515/9783110214475.1.3.176.

van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the language faculty*. New York: Oxford University Press. DOI: 10.1093/acprof:oso/9780199756056.001.0001.

Weiss, Daniel. 2004. The rise of an indefinite article: The case of Macedonian *eden*. In Walter Bisang, Nikolaus P. Himmelmann & Björn Wiemer (eds.), *What makes grammaticalization? A look from its fringes and its components*, 139–168. Berlin: De Gruyter Mouton. DOI: doi:10.1515/9783110197440.2.139.

Chapter 5

Multiple wh-fronting in a typological setting: What is behind multiple wh-fronting?

Željko Bošković

University of Connecticut

The paper establishes broad typological correlations between multiple wh-fronting (MWF) and other phenomena in an attempt to understand what is behind MWF. In particular, the paper establishes a correlation between MWF and the morphological shape of wh-words, which is argued to be responsible for MWF. MWF languages are also shown to be characterized by a particular status regarding articles: they either lack articles or have affixal articles (the difference is shown to matter for superiority effects). Certain cases of non-wh indefinite interpretations of wh-phrases and the exceptional behavior of D-linked wh-phrases regarding MWF – they are not subject to it – are also discussed and captured (including Hungarian, where D-linked wh-phrases are not exceptional in this respect, which is tied to another exceptional property of Hungarian).

1 What is special about multiple wh-fronting?

The goal of this paper is to shed light on what is behind one particular language type regarding multiple questions. Most languages front one question word/wh-phrase or leave them all in situ in multiple questions. The former type is illustrated by English (1) and the latter by Chinese (2).

(1)   What did John give to who?

(2)   John gei-le   shei shenme?   (Chinese)
      John give-PFV who what
      'What did John give to who?'

Željko Bošković

There is another pattern, which is not frequent crosslinguistically: the so-called multiple wh-fronting languages (MWF), which front all wh-phrases in questions. The pattern is illustrated by Serbo-Croatian (SC) examples in (3) (note that SC is an SVO language).[1]

(3)   a.   Ko šta kupuje?   (SC)
            who what buys
            'Who is buying what?'
       b.   *Ko kupuje šta?
            who buys what

There have been quite a few works on MWF in the generative tradition since the seminal paper by Rudin (1988) (MWF has been discussed less outside of that tradition, but see e.g. Mycock 2007). These works generally focus on examining the structure and the derivation of MWF constructions. However, they do not attempt to understand what is really behind MWF, why some languages employ this strategy.

     This paper aims to address that question, but from a broad typological perspective, in particular, by establishing correlations between MWF and other phenomena. Its scope will be limited – I will not go into the derivation and the structure of

---

[1]There are some highly specific contexts where MWF languages need not front wh-phrases (just like there are contexts where English can employ wh-in-situ). I will generally not be concerned with those exceptional contexts here (apart from D-linking), just with the broad, main pattern. I merely note that, as discussed in Bošković (2002), several of those exceptional contexts involve PF issues, e.g. the case where the fronted wh-phrases would yield a sequence of homophonous elements, like Romanian (i). Bošković (2002) shows that such cases are exceptional only superficially – they still involve MWF in the syntax, with pronunciation of a lower copy of a moved wh-phrase (second *ce* in (i.a)), which is motivated by PF considerations. Thus, the second wh-phrase in (i.a) licenses parasitic gaps (see (ii)), which is a test for movement in overt syntax (compare (iii.a) and (iii.b)).

    (i)   a.   Ce precede ce?   (Romanian)
              what precedes what

         b.   *Ce ce precede?
             what what precedes

   (ii)   Ce precede ce fără să influenţeze?   (Romanian)
       what precedes what without SUBJ.PART influences
       'What precedes what without influencing?'

  (iii)   a.   *What precedes what without influencing?
         b.   What did Mary promote without influencing?

MWF constructions; the goal of the paper is simply to establish, and understand, prerequisites for the MWF pattern, in an effort to understand what is behind this strategy of forming multiple questions. The discussion will be based on the following 18 (typologically diverse) MWF languages: SC, Romanian, Polish, Russian, Bulgarian, Macedonian, Czech, Slovenian, Ukrainian, Yiddish, Lithuanian, Hungarian, Basque, Mohawk, Georgian, Ossetic, Svan, and Latin.[2] Latin will turn out to be particularly useful, since it can be compared with modern Romance languages.

What will be important for our purposes is the notion of indeterminates (the term goes back to Kuroda 1965, who actually took it from traditional Japanese grammars, which use the term "indeterminate words"). In many languages, the same forms that are used for wh-words have a variety of usages, like existentials, universal quantifiers, negative concord/negative polarity items, free choice, depending on the context where they occur (for much relevant discussion, see Haspelmath 1997). They are referred to as indeterminates since their exact quantificational force is not inherently determined – it is determined by the licensing context in which they are found.

Cheng (1991), a predecessor of this work, observes that Bulgarian, Polish, and Hungarian have indeterminate systems. It turns out that all MWF languages from above have a productive indeterminate system, which suggests that the indeterminate system is a prerequisite for MWF. But there is more to it. There are different types of indeterminate systems. I define here a particular type, which I will refer to as the sub-wh system. It is a fully productive system where addition of an inseparable affix to a wh-phrase results in a series of meanings shown in SC (4).

(4)   a. ko 'who'   (SC)

         b. *i*ko 'anyone'

         c. *ni*ko 'no one'

         d. *ne*ko 'someone'

         e. *sva*ko 'everyone'

         f. *bilo* ko 'whoever'

---

[2]The list includes languages I was able to identify as having MWF (and determine for them the additional information that is needed in the discussion below) based on literature surveys (most of them are well-known as MWF languages; for some less-known cases, see Ledgeway 2012 for Latin, Baker 1996 for Mohawk, Gillon & Armoškaitė 2015 for Lithuanian, Erschler 2012 for Ossetic, Erschler 2015 for Georgian and Svan).

There is a morphological subset-superset relationship between the wh/question usage and other usages, as stated in (5) regarding 'who'.

(5) sub-wh system: *who*+X for other pronouns (inseparable, fully productive, order doesn't matter)

What is not a sub-wh system is the situation found in Chinese, where the same form can have different functions, as illustrated by (6), or Japanese, where a particle occurs on each function – in some cases inseparable (namely, existential), in some cases separable – note that *-ka*, which is always separated on the wh-usage in Standard Japanese, need not be separated in Okinawan, as illustrated by (7).

(6) a. Ni xiang mai *shenme* (ne)?   (Chinese)
         you want buy what   Q
         'What do you want to buy?'

     b. Wo bu   xiang mai *shenme.*
         I   not want buy anything
         'I don't want to buy anything.'

     c. Wo xiang mai yi-dian *shenme.*
         I   want buy one-CL something
         'I want to buy something.'

(7) a. Taruu-ja *nuu* koota-*ga.*   (Okinawan)
         Taro-TOP what bought-Q
         'What did Taro buy?'

     b. Taruu-ja *nuu-ga* koota-ra.
         Taro-TOP what-Q bought-RA
         'What did Taro buy?'   (Kinjo & Oseki 2016)

It should be noted that it has been argued that the Q marker starts with the wh-phrase even in Standard Japanese (just as in Okinawan), see e.g. Hagstrom (1998). This is then a rather different system from SC, where the wh-form is a subset of everything.[3]

English also does not have a sub-wh system since the relevant system is not fully productive in English (compare *some*where, *every*where, *no*where,

---

[3]Japanese is, however, more similar to SC in the relevant respect than Chinese is, which may not be surprising in light of the discussion below given that Japanese in fact used to be a MWF language (i.e. Old Japanese appears to have been a MWF language; see Aldridge 2009, Dadan 2019).

*any*where with **some*who/*every*who/*no*who, **no*what/*no*when/*no*how), i.e., it is lexicalized (Cheng 1991 suggests that the good cases are lexically incorporated forms, essentially compounds).

Returning to MWF languages, it turns out that all MWF languages have exactly the sub-wh type of indeterminates, which leads me to posit (8) (note that this is a one-way correlation).

(8)   If a language has multiple wh-fronting, it has a sub-wh indeterminate system.

This was illustrated above with SC in (4). Additional confirmations of (8) are provided by the MWF languages in Tables 1–4 (the data in Tables 1–7 are from, or based on, Haspelmath 1997; only partial paradigms are given below, and not all series are illustrated – all these languages have additional series; for more complete paradigms, see Haspelmath 1997).[4]

Table 1: Russian indeterminate series

	interrogative	existential	neg-concord	free choice
person	kto	kto-to	ni-kto	kto ugodno
thing	čto	čto-to	ni-čto	čto ugodno
place	gde	gde-to	ni-gde	gde ugodno
time	kogda	kogda-to	ni-kogda	kogda ugodno
manner	kak	kak-to	ni-kak	kak ugodno

Particularly interesting for our purposes is Romance. Latin was clearly a MWF language (see Ledgeway 2012 and Dadan 2019 for extensive discussion) and had a fully productive sub-wh system. The fully productive sub-wh system got lost in

---

[4]I do not consider German as having a productive sub-wh system since in German only one series, the *irgend*-series (but not the *etwas*- or *n*-series, which are the respective second and fourth examples in (i)), is related to wh-words, as shown by (i) (data from Haspelmath 1997; note, however, that (8) is a *one-way* correlation).

(i)   a.   *person*: wer, jemand, irgend-wer / irgend-jemand, niemand

b.   *thing*: was, etwas, irgend-was / irgend-etwas, nichts

c.   *place*: wo, –, irgend-wo, nirgends

d.   *time*: wann, –, irgend-wann, nie

e.   *manner*: wie, –, irgend-wie, (auf keine Weise)

f.   *determiner*: welche, (ein), irgend-ein / irgend-welche, kein

Table 2: Bulgarian indeterminate series

	interrogative	existential	neg-concord	free choice
person	koj	nja-koj	ni-koj	koj to i da e
thing	što	ne-što	ni-što	što to i da e
place	kâde	nja-kâde	ni-kâde	kâde to i da e
time	koga	nja-koga	ni-koga	koga to i da e
manner	kak	nja-kak	ni-kak	kak to i da e

Table 3: Hungarian indeterminate series

	interrogative	existential	neg-concord	free choice
person	ki	vala-ki	sen-ki	akár-ki
thing	mi	vala-mi	sem-mi	akár-mi
place	hol	vala-hol	se-hol	akár-hol
time	mikor	vala-mikor	sem-mikor	akár-mikor
manner	hogy(an)	vala-hogy(an)	se-hogy(an)	akár-hogy(an)

Table 4: Basque indeterminate series

	interro-gative	*bait*-series (non-emphatic)	*i*-series (NPI)	*edo*-series (free choice)	*nahi*-series (free choice)
person	nor	nor-bait	i-nor	edo-nor	nor-nahi
thing	zer	zer-bait	e-zer	edo-zer	zer-nahi
place	non	non-bait	i-non	edo-non	non-nahi
time	noiz	noiz-bait	i-noiz	edo-noiz	noiz-nahi
manner	nola	nola-bait	i-nola	edo-nola	nola-nahi
determiner	zein	–	–	edo-zein	zein-nahi

all modern Romance languages except one: Romanian, which is the only modern Romance language that still has MWF, a strong confirmation of (8). A partial illustration of the Romance situation is given in Tables 5–7.

Table 5: Latin indeterminate series

	interrogative	existential	polarity	free choice
person	quis	ali-quis	quis-quam	qui-vis
thing	quid	ali-quid	quid-quam	quid-vis
place	ubi	ali-cubi	usquam	ubi-vis
time	quando	ali-quando	umquam	–

Table 6: Italian

	interrogative	existential	neg-concord
person	chi	qualcuno	nessuno
thing	che	qualche cosa, qualcosa	niente, nulla
place	dove	in qualche luogo	in nessun luogo
time	quando	qualque volta	(mai)

Table 7: Romanian indeterminate series

	interrogative	existential	free choice	*oare*-series
person	cine	cine-va	ori-cine	oare-cine
thing	ce	ce-va	ori-ce	oare-ce
place	unde	unde-va	ori-unde	oare-unde
time	cînd	cînd-va	ori-cînd	oare-cînd

I conclude therefore that a sub-wh system is a prerequisite for MWF. I will now briefly discuss why that is the case.

The crucial point is that *ko* in (4b) is actually not 'who', i.e. it does not correspond to English *who*. The form is a true indeterminate, which means that it does not have an inherent quantificational force (see below for evidence to this effect). It requires licensing, which also determines its quantificational force (i.e. its exact meaning).

The particles that indeterminates merge with normally do that – they determine the exact quantificational force, and the meaning of the indeterminate in cases like those given in SC (9) as a partial illustration of the relevant SC paradigm.[5]

(9) a. i+*ko* 'anyone'   (SC)
     b. n+i+*ko* 'no one'

Importantly, in a sub-wh system, the *only* usage on which the indeterminate is not merged with a particle is the wh-usage, which means that we are dealing here with an unlicensed indeterminate. I suggest that this is what requires fronting. The indeterminate is licensed as a wh-phrase by moving to an interrogative projection (which determines its meaning). The movement thus does not occur because of a property of the interrogative head (which is the case in English, where only one wh-phrase fronts because of that), but because of indeterminate licensing – this is why they *all* need to undergo fronting, resulting in MWF.

In short, in the sub-wh system, affixes merged with an indeterminate determine its quantificational force and license the indeterminate. When there is no such affix, the indeterminate is licensed as a wh-phrase by movement to an interrogative projection.

MWF languages do however have certain cases where the wh-phrase itself (so the form that is used in wh-questions) receives a different, non-wh interpretation, like the wh-existential in (10) (see e.g. Izvorski 1998, Bošković 2002, Šimík 2011).

(10) a.   Ima ko šta   da ti proda.   (SC)
            has who what that you sells
            'There is someone who can sell you something.'
      b. * Ima ko   da ti   proda šta.
            has who that you sells what

Importantly, the relevant elements must front here. The fronting does not occur to the interrogative projection, since the relevant clause is simply not interrogative. I suggest that since *ko* and *šta* are not merged with an indefinite particle in these cases, they are licensed as indefinites by moving to a special indefinite licensing position. What is relevant here is languages like Kaqchikel, where the

---

[5]In these particular cases, the morphology is rather transparent. *I-* also means 'even'. On the connection between 'even' and NPIs, see e.g. Rooth (1985), Haspelmath (1997), Giannakidou (2007), Crnič (2011); *n-* may indicate a connection with negation. At any rate, these details are not important for our purposes.

exact same form functions as interrogative or indefinite, and must be fronted on both functions, with the landing site of the interrogative being higher than the indefinite licensing projection, as discussed in detail in Erlewine (2016). What Kaqchikel shows is that there is a pattern where the indefinite meaning of an indeterminate is licensed by movement to a special projection that licenses this meaning (see Erlewine 2016). The suggestion is that this is precisely what happens in (10) (the movement is not to the interrogative CP projection since the relevant clauses are clearly not interrogative; note that this (i.e. (10)) can also be taken to confirm that the relevant elements are not inherently wh-phrases but bare indeterminates).[6]

It is worth noting that a number of Australian languages have the same form for wh-phrases and indefinites but while the morphology is the same the syntax is not: as wh-phrases they must front, as indefinites they stay in situ (these languages cannot be checked for MWF since they do not allow multiple questions in the first place, see Cheng 1991 for relevant discussion of these languages).

(11) Martuthunira

    a. ngana nganhu   wartirra nyina-nguru karra-ngka
       who   that.NOM woman   sit-PRS   scrub-LOC
       muyinu-npi-rra?
       hidden-INCH-CTEMP
       'Who is that woman hiding in the scrub?'

    b. ngayu   nyina-lha martama-l.yarra palykura-la
       1SG.NOM sit-PRS   press.on-CTEMP groundsheet-LOC
       nganangu-la.
       someone.GEN-LOC
       'I sat down on someone's groundsheet, holding it down.' (Dench 1987)

(12) Panyjima

    a. ngatha ngananhalu   nhantha-nnguli-nha.
       1SG.NOM something.INS bit-PASS-PST
       'I was bitten by something.'

---

[6]The movement strategy just discussed and the affixation strategy for licensing indefinites can be combined, though this option is slightly disfavored, possibly due to a parallelism for indefinite licensing being favored.

   (i)   ? Ima ko da ti proda nešto.   (SC)
         has who that you sells   something
         'There is someone who can sell you something.'

    b. ngananha-ma-rna nyinta   ngunhalku?
        what-CAUS-PST   2SG.NOM that.ACC
        'What have you done to him?'   (Dench 1981)

There is a parallel situation with MWF languages. In particular, there are similar wh-indefinites in Slavic MWF languages, as illustrated by Russian (13) (see e.g. Zanon 2022, Hengeveld et al. 2023).

(13)   Možet, kto   prixodil.   (Russian)
       maybe who   came
       'Maybe someone came.'   (Hengeveld et al. 2023)

This usage is very restricted in Slavic; in SC even more so than in Russian – (13) is in fact unacceptable in SC; regarding Russian, see especially Zanon (2022), who argues that the relevant elements are licensed by a semantically motivated and constrained null operator, which essentially plays the role of the licensing affixes discussed above hence this kind of analysis of the usage in question can be adjusted to the system developed here. Alternatively, it is possible that an indeterminate that does not have a licensing particle attached and does not move to an indeterminate-licensing projection or has a linking index (see the discussion right below) is interpreted by a default rule for unlicensed indeterminates, which would apply in the relevant contexts in the languages that allow this usage (they also differ regarding such contexts), as a simple indefinite. In this respect, it is worth noting that such indefinites cannot occur in wh-questions (see e.g. Zanon 2022 and Hengeveld et al. 2023), which can be taken to confirm the default nature of the licensing in question – it is available only if another way is not available.[7]

    Interestingly, Zanon (2022) and Hengeveld et al. (2023) observe that these wh-indefinites cannot be focused. What is important here is that real MWF/wh-fronting in Slavic has been analyzed as focus-movement (e.g. Bošković 2002, see also Stepanov 1998 for Russian as well as the discussion below), i.e. it is essentially focusing. It then makes sense that if the relevant element is focused it would be interpreted as a wh, not a non-wh (i.e. indefinite), hence the non-wh-indefinite usage does not allow focalization. Hengeveld et al. (2023) actually observe that the non-focusing requirement is not general – it does not hold in Dutch. Given the current discussion, Dutch-like exceptions should not be possible in MWF languages.

---

[7]Note that these indefinites are different from those in wh-existentials like (10) – e.g. Ksenia Zanon (p.c.) notes that the former cannot be coordinated, see Zanon (2022), while the latter can be.

A different (and independent) exception to MWF concerns D-linked wh-phrases, which need not undergo fronting, as illustrated below by SC (14). (Note, however, that this is not the case in all MWF languages; they must front in Hungarian, which is discussed in Section 2.)[8]

(14) Ko kupuje koju knjigu?   (SC)
      who buys   which book
      'Who is buying which book?'

Two issues are relevant here. First, *koju* is not an indeterminate but a wh-specific form (this may not be a general situation though). Second, as briefly noted above, Bošković (2002) argues that MWF is actually movement to a focus projection, this means that the relevant licensing takes place in the Spec of a focus-licensing head; this by itself is not surprising – focus/interrogativity connection has often been noted.[9] Furthermore, Bošković (2002) observes that D-linking is very different from focus. With D-linked wh-phrases the range of felicitous answers is restricted by a set of objects that is familiar to the speaker and the hearer as a result of it being referred to/salient in the context. In other words, the range

---

[8]It may be worth noting here that D-linked wh-phrases more generally can be special, and subject to ill-understood language variation. Thus, there are languages that disallow multiple questions, e.g. Hong Kong Sign Language (HKSL), Italian, and Mandinka. ((i)–(iv) are taken from Gan 2022).

   (i)   * WHO BUY WHAT?   (HKSL)

   (ii) * Chi ha scritto che cosa?   (Italian)
           who has written what

Gan (2022) shows that D-linking improves multiple questions in HKSL and Mandinka, but not in Italian (it is not out of question that there is some connection here with the SC vs. Hungarian difference regarding D-linked MWF questions).

   (iii) STUDENT WHO BUY COMPUTER BUY-WHICH?   (HKSL)
         'Which student bought which book?'

   (iv) * Quale studente comprerà quale libro?   (Italian)
           which student will-buy   which book

[9]I assume that as a result of this connection, indeterminates can still be licensed as interrogative in such a projection. (Possibly, being in such a projection would enable them to undergo unselective binding with interrogative C in spite of the issue noted in Section 2.1 (i.e. without a null operator, the intuition being that it is not needed in this case since the relevant element is located in an operator, in fact the right operator, position—SpecFocP), which would license their interrogative interpretation.)

of reference of D-linked wh-phrases is discourse-given. Due to their discourse givenness, such wh-phrases are not focused, hence they are not subject to focus movement. (One wh-phrase always must front for clausal typing as discussed in Cheng 1991 so when only a D-linked wh-phrase is present it fronts but Bošković 2002 shows that the landing site is different; for special behavior of D-linked wh-phrases regarding MWF see also Diesing 2003 on Yiddish, which disallows MWF with D-linked wh-phrases.)

Regarding the interpretation of D-linked wh-phrases, Enç (2003) proposes that specific arguments have a linking index ℓ which identifies the set of individuals of which the argument is a member (i.e. it gives the set which that argument must belong to). Non-specific arguments have no such index. Shields (2008) extends this to wh-phrases: D-linked wh-phrases are specific and therefore have a set-denoting (linking) index, which non-D-linking wh-phrases do not have. The linking index points to the set of entities in the discourse that a specific expression is required to be a member of.

Indeterminate pronouns are normally non-specific, D-linked ones (i.e. D-linked wh-phrases) are not. The interpretation of the latter is essentially determined by their semantics, no further licensing is needed (essentially, an indeterminate with a linking index is interpreted as D-linked – the linking index points to the set of entities in the discourse that the relevant element is required to be a member of). It is also possible that the linking index allows D-linked wh-phrases to undergo unselective binding by interrogative C and that they are licensed in that way (see Pesetsky 1987 on unselective binding of D-linked wh-phrases; see also Section 2).[10]

In conclusion, this section has established a correlation between MWF and another phenomenon. In particular, MWF languages have been shown to have a sub-wh indeterminate systems, which forces MWF (except with D-linked wh-phrases).

---

[10]There is an alternative account. A number of authors (e.g. Belletti 2004, Lacerda 2020) have argued for several languages that they have a low topic projection. It is possible that D-linked wh-phrases are licensed in a low topic-like projection (see Grohmann 2006 for D-linking as topichood). On this analysis, the D-linked wh-phrase in (14) would not actually be in situ (SC and Hungarian could then differ here regarding topic movement; see, however, below). It is worth noting here that (i) is also acceptable. Bošković (2002), however, shows that the D-linked wh-phrase in such cases is lower than the second wh-phrase in examples like (3a), i.e. it is not the case that the D-linked wh-phrase simply optionally undergoes movement that the second wh-phrase must undergo in (3) (examples like (i), i.e. optional fronting, is actually not allowed in all MWF languages, see Bošković 2002, Pesetsky 1987, Wachowicz 1974).

   (i) Ko koju knjigu kupuje?   (SC)
        who which book buys
        'Who is buying which book?'

2 Multiple wh-fronting and articles

2.1 Another generalization

I will now show that there is another property that MWF languages have in common, which is in principle independent of the one presented in Section 1 (in the sense that if one of the generalizations in questions turns out not to be correct the other one would not necessarily be affected).[11] In particular, they all either lack definite articles or have affixal definite articles (15). The relevant language cut is given in (16).[12]

---

[11]But see the generalization regarding indeterminates themselves in Oda (2022) that would actually relate (8) and (15). Oda also provides an alternative deduction of (8) based on my earlier version of this generalization given in Bošković (2020) where the prerequisite for MWF was a broader indeterminate system than the sub wh-system.

[12]For most of the languages listed in (16b), their affixal status is well-known. For arguments that Hungarian definite article is affixal (more precisely, a prefix), see MacWhinney (1976), Oda (2022), and Lewis (2024). MacWhinney observes that it undergoes a morphophonemic alternation that is typical of affixes, while Oda and Lewis observe typological generalizations where Hungarian patterns with languages with affixal articles (languages with affixal articles actually pattern with languages without articles regarding those generalizations). Regarding the affixal status of the definite article in Yiddish, which might be the least discussed case here, see Oda (2022). To mention some relevant arguments, Talić (2017) and Oda (2022) observe that languages with affixal definite articles allow article omission in contexts where such omission is not possible in free-standing article languages like English. Oda notes that this is especially the case in PPs, where due to article omission a bare noun can even receive a definite interpretation in (some) affixal article languages, which is never possible in languages with non-affixal definite articles, where a definite article is required for definite interpretation (see Bošković 2016; Oda argues that in the relevant cases the preposition essentially functions as the definite article). Thus, Zwicky (1984: 119) observes regarding (i.b): "The phrase *in gloz* in 'in the glass' is a typical example. The noun *gloz* in this expression is understood definitely, and can even be anaphoric."

   (i)   a.   lebn tir   (Yiddish)
               near door
               'near the door'
          b.   in gloz = in the glass   (Zwicky 1984)

Bošković (2016) also notes that, for the purposes of Bošković's NP/DP generalizations (see below for some relevant discussion), definite articles have a form distinct from demonstratives. Definite articles in Yiddish have the same form as demonstratives, with stress distinguishing them. Margolis (2011: 122) in fact states that: "this/these" is identical to the definite article with added stress. Essentially following Oda (2022), I thus consider Yiddish to be an affixal article language, the definite article being an affixal, hence unstressed, version of the demonstrative (there may be a change under way regarding the status of the relevant element where dialectal differences may also be relevant; not all dialects of Yiddish in fact have MWF, see Diesing 2003).

(15)   MWF languages either lack articles or have affixal definite articles.

(16)   a.   No articles: SC, Polish, Russian, Czech, Slovenian, Ukrainian, Mohawk, Latin, Georgian, Lithuanian, Ossetic, Swan

      b.   Affixal articles: Romanian, Bulgarian, Macedonian, Basque, Hungarian, Yiddish

Turning to the deduction of (15), in a series of works (e.g. Bošković 2012), based on a number of syntactic and semantic typological generalizations, where languages with and without definite articles consistently differ regarding a number of syntactic and semantic phenomena, I argued that languages without definite articles do not project DP (i.e., there are no null definite articles in such languages).

Talić (2017) argues for a refinement of the NP/DP language distinction; she shows that in many respects languages with affixal definite articles behave like a separate type (see also Oda 2022, Lewis 2024), in that they sometimes behave like languages with articles and sometimes like those without articles.[13]

In Bošković (2020) I suggested an implementation of this observation for the affixal article languages that have MWF: there is D in such languages, but there is no DP. The affixal article is base-generated adjoined to N (more precisely, its host). It should be noted that there is nothing strange about this theoretically: Adjunction through movement can involve either phrasal or head adjunction, the same should hold for adjunction through base-generation (for much relevant discussion regarding definite articles, see also Oda 2022; regarding indefinite articles, see Wang 2019).

book

book   the
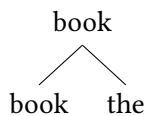
Figure 1: Noun-article base-generation

Recall now that in a sub wh-system, only on the wh-usage the indeterminate does not occur with a licensing particle. I suggest then that, in principle, such indeterminates can still be licensed at a distance *in situ*, with a null operator in SpecDP that is unselectively bound by interrogative C. This is not possible in MWF languages due to the lack of a DP projection that would be capable of such

---

[13]Below, for ease of exposition I will simply use the term (affixal) article, though what matters here (and what matters for Bošković's NP/DP generalizations) is definite articles only.

licensing. The only way to license the indeterminate on the wh-usage is then to front it to an interrogative position.[14]

A confluence of independent factors, namely the sub wh-system and a particular status regarding articles, is what is behind MWF: MWF languages have a sub-wh indeterminate system, and either lack articles or have affixal articles, which are the typological findings of this paper.

Regarding the relevance of the latter property, in languages without articles and languages with affixal articles the possibility of wh-licensing *in situ* by interrogative C through unselective binding is blocked because such licensing is done through a null operator in SpecDP (except with D-linked wh-phrases), which is absent in languages without articles and languages with affixal articles (in the former, because DP itself is lacking, and in the latter because the affixal article is base-generated adjoined to N, which means that in such languages there is D, but there is still no DP, hence no null operator in SpecDP).

2.2 Superiority variation regarding basic Superiority effects

I turn now to a case of variation within MWF languages which will also shed light on the exceptional status of Hungarian regarding D-linked wh-phrases, noted in Section 1. Already Rudin (1988) observed that MWF languages differ regarding whether they show ordering, i.e. Superiority, effects with MWF. Regarding basic cases like those shown in (17)–(18), SC does not show them, while Bulgarian does show them.

(17) a. Koj kakvo e kupil?   (Bulgarian)
          who what is bought

  b. * Kakvo koj   e kupil?
          what   who is bought
      (Intended:) 'Who bought what?'

(18) a. Ko šta   kupuje?   (SC)
          who what buys

  b. Šta   ko   kupuje?
      what who buys
      'Who is buying what?'

A survey of the literature shows the following language cut regarding Superiority effects in basic cases of this sort.

---

[14]As suggested above, D-linked wh-phrases may be able to undergo unselective binding even in the absence of DP for independent reasons, namely, due to the presence of the linking index.

(19) a. No Superiority effects: SC, Polish, Czech, Russian, Slovenian, Ukrainian, Mohawk, Lithuanian, Georgian, Ossetic, Svan, Hungarian

     b. Superiority effects: Romanian, Bulgarian, Macedonian, Basque, Yiddish

It turns out that the cut is not arbitrary – there is a correlation with (the type of) articles. Putting Hungarian aside (taking Hungarian into consideration we would have a one-way correlation in (21), which was actually noted in Bošković 2008), we have (20).

(20) MWF languages without articles do not show basic Superiority effects, those with affixal articles do.

(21) MWF languages without articles do not show basic Superiority effects.

Below, I will briefly outline a deduction of (20) that will also accommodate the Hungarian exception (given the affixal status of the Hungarian definite article, see fn. 12), tying it to another Hungarian exception, namely the exceptional behavior of Hungarian regarding D-linking.

Bošković (2002) argues that Superiority effects arise with MWF to SpecCP (English-style wh-movement), not with MWF to a lower position, which means that SC MWF targets a lower position than Bulgarian MWF (see Bošković 2002 for evidence to this effect). Now, if Superiority is taken to be a sign of true, English-style wh-movement, this can be generalized in such a way that languages with articles (non-affixal or affixal) must have true English-style wh-movement to SpecCP when fronting wh-phrases. Bošković (2008) in fact suggests that the D-feature is crucially involved in movement to SpecCP. Affixal article languages still have the D-feature, which means that they have wh-movement to SpecCP, which is Superiority inducing. This then captures (20). But what about Hungarian?

Superiority as a test for wh-movement is confirmed by single-pair (SP)/pair-list (PL) answers. Bošković (2001, 2002) shows that overt wh-movement languages require a PL answer for examples like (22). (22) cannot be felicitously asked in the following situation: John is in a store and sees somebody buying an article of clothing, but does not see who it is and does not see exactly what the person is buying. He goes to the sales clerk and asks (22).

(22) Who bought what?

Whereas German patterns with English, wh-in-situ languages Japanese, Hindi, and Chinese allow SP answers in such questions (see Bošković 2001). Importantly, French allows SP answers, but only with in-situ questions like (23a), not (23b).

(23) a. Il a   donné quoi à qui?   (French)
          he has given what to who

      b. Qu' a-t-il donné à qui?
          what has-he given to who
    'What did he give to who?'

Based on this, Bošković (2001, 2002) argues that the availability of SP answers
depends on the possibility of not moving any wh-phrase to SpecCP overtly (see
Bošković 2001 for an account of this generalization).

Turning to MWF languages, SC allows SP answers, while Bulgarian does not,
which confirms that SC MWF lands in a lower position than Bulgarian MWF (see
Bošković 2007 and references therein for additional languages confirming this).

As noted above, Bošković (2002) argues that MWF involves focus. Now,
Bošković (1999) argues that movement-attracting heads can differ regarding
the specification of the movement-attracting feature. They can be specified to
attract one element with the relevant feature, call it F, or all elements with the F
feature. English interrogative C is an attract 1-F head – it attracts one (in particu-
lar, the highest) element with the wh-feature. In SC, wh-phrases undergo focus
movement; the relevant head has the specification Attract All-focus. Bulgarian
is a combination of English and SC: It has single-fronting wh-movement as in
English (Attract 1-wh) and MWF for focus (Attract All-focus, see Bošković 1999
and fn. 16). Importantly, from this perspective, Superiority is not a diagnostic
of wh-movement, but single fronting.[15] In this respect, Bošković (2002) shows
that there are selective Superiority effects in Bulgarian. Only the first wh-phrase,
which is the only wh-phrase that undergoes wh-movement, is subject to Superi-
ority effects, other wh-phrases are not. Thus, the indirect object wh-phrase must
precede the direct object wh-phrase in (24) (because it is higher than the object

---

[15]Given the economy-of-derivation condition that every requirement be satisfied through the
shortest movement possible, Attract 1-F heads will always attract the highest phrase with the
relevant feature: thus, in (i), the relevant formal inadequacy of the interrogative C is checked
through a shorter movement in (i.a) than in (i.b) (cf. the pre-wh-movement structure in (i.c)).

(i)   a. Who$_i$ did Mary tell t$_i$ to buy the book?
      b. * What$_i$ did Mary tell who to buy t$_i$?
      c. Mary tell who to buy what

With Attract All-F heads, like the SC focus-licensing head, all relevant elements must move:
Regardless of the order of movement, the same number of nodes are crossed with such move-
ment, hence the order of movement of wh-phrases is free (see Bošković 1999 for a more detailed
discussion).

wh-phrase before wh-fronting) but not in (25), where a subject wh-phrase, which is higher than both indirect and direct object wh-phrase before wh-fronting, is present.[16]

(24) a. Kogo kakvo e pital Ivan?   (Bulgarian)
          whom what   is asked Ivan

      b. ?* Kakvo kogo   e pital   Ivan?
             what   whom is asked Ivan
      'Who did Ivan ask what?'

(25) a. Koj kogo   kakvo e pital?   (Bulgarian)
          who whom what   is asked

      b. Koj kakvo kogo   e pital?
          who what   whom is asked
      'Who asked who what?'

All this raises a question: Is there a MWF language where D-linked wh-phrases also must front? That would be a true MWF counterpart of English, with an Attract All-wh specification (note that Attract All-wh affects D-linked wh-phrases, in contrast to Attract All-focus). As noted above, and as discussed in Bošković (2007) and É. Kiss (2002), both D-linked and non-D-linked wh-phrases must move in Hungarian. This is illustrated by (26)–(27).

(26) a. * Ki   irt   mit?   (Hungarian)
             who wrote what

      b.   Ki   mit   irt?
           who what wrote

      c.   Mit ki   irt?
           what who wrote
      (Intended:) 'Who wrote what?'   (Bošković 2007)

---

[16]Note that, as discussed in Bošković (1999), it is the same head, interrogative C, that has the relevant properties (Attract 1-wh, Attract All-focus) in Bulgarian. Given that the first wh-phrase that moves to SpecCP automatically satisfies the Attract 1-wh requirement (see Bošković 1999), the highest wh-phrase must move first, then the order of movement does not matter, since Attract All-focus does not care about the order of movement, as noted in fn. 15. (Note that, as standardly assumed, the order of fronted wh-phrases reflects the order of their movement, see Rudin (1988), Richards (2001) for different implementations of this, i.e. the wh-phrase that is first in the linear order is the one that moves first, hence the highest wh-phrase must move first when Superiority is in effect.)

(27) a. * Ki   irta   melyik levelet?   (Hungarian)
          who wrote which letter

   b.   Ki   melyik levelet irta?
          who which letter   wrote

   c.   Melyik levelet ki   irta?
          which letter   who wrote
      (Intended:) 'Who wrote which letter?'   (Bošković 2007)

Importantly, Hungarian MWF questions also disallow SP answers (see e.g. Surányi 2005) and do not show Superiority effects (see (26)), which is exactly the behavior expected of a true MWF counterpart of English (there are no Superiority effects since we are dealing only with an Attract-All fronting and SP answers are disallowed because the fronting is to SpecCP).[17] What appeared to be an exceptional behavior of Hungarian regarding Superiority and D-linking is thus explained, in fact in a uniform manner.

At any rate, the discussion from Section 2 is summarized below in table form (where the left column gives the relevant language types – there are two types for affixal article languages, depending on whether D-linked wh-phrases are also subject to MWF).

Table 8: Summary

	MWF	Superiority with MWF	SP with wh-fronting
Free-standing article	*	N/A	*
Affixal article		Yes	*
Affixal article + D-linking MWF		No	*
No article		No	

3 Conclusion

The paper has established correlations between MWF and other phenomena, in an attempt to understand what is behind MWF. In particular, MWF languages

---

[17]Horváth (1998), Puskás (2000), Lipták (2001), and É. Kiss (2002) suggest that the wh-phrase that is closest to the verb in Hungarian MWF questions undergoes focus-movement, other wh-phrases undergo movement that non-wh-quantifiers undergo, but see Surányi (2005) for arguments against this position.

have been shown to have a sub-wh indeterminate system, which was suggested to force MWF. In such a system, an inseparable affix is attached to the indeterminate, with the exact quantificational force of the indeterminate determined by the affix that merges with it. What is traditionally considered to be wh-phrases in the sub-wh indeterminate system are not really wh-phrases but bare indeterminates; they are not licensed *in situ* because they are bare – no licensing affix is attached to them – hence they must front to a position in the left periphery to get licensed, which in turn determines their interpretation. This yields MWF.

MWF languages are also characterized by a particular status regarding articles – they either lack articles or have affixal articles. It was argued that in these language types the possibility of wh-licensing *in situ* by interrogative C through unselective binding is blocked because such licensing is done through a null operator in SpecDP, which is absent in languages without articles and languages with affixal articles for a principled reason. The distinction between the lack of articles and affixal articles in MWF languages was, however, shown to have an effect on the presence/absence of Superiority effects. The exceptional behavior of D-linked wh-phrases regarding MWF (they don't need to undergo it) was also captured (including the Hungarian pattern, where D-linked wh-phrases are not exceptional in this respect – they are subject to MWF). Certain cases of non-wh indefinite interpretations of wh-phrases were also discussed.

All in all, the paper has established the following generalizations regarding MWF, where Hungarian was shown to be exceptional regarding (28c) but for a principled reason, which was tied to its exceptional behavior regarding D-linking; the reader should thus bear in mind that the way (28c) is deduced in the paper does leave room for principled exceptions.

(28) a. If a language has multiple wh-fronting, it has a sub-wh indeterminate system.
      b. MWF languages either lack articles or have affixal definite articles.
      c. MWF languages without articles do not show basic Superiority effects, those with affixal articles do.

At any rate, the main typological finding of this paper is that a confluence of independent factors, namely the sub-wh indeterminate system and a particular status regarding articles, is what is behind MWF.

Abbreviations

1	first person	PART	particle
2	second person	PASS	passive
ACC	accusative	PFV	perfective
CAUS	causative	PL	pair-list
CL	classifier	PRS	present
CTEMP	contemporaneous relative	PST	past
GEN	genitive	Q	question particle
HKSL	Hong Kong Sign Language	SC	Serbo-Croatian
INCH	inchoative	SG	singular
INS	instrumental	SP	single-pair
LOC	locative	SUBJ	subjunctive
NOM	nominative	TOP	topic
MWF	multiple wh-fronting		

Acknowledgments

References

Aldridge, Edith. 2009. Short wh-movement in Old Japanese. In Shoishi Iwasaki, Hajime Hoji, Patricia M. Clancy & Sung-Ock Sohn (eds.), *Japanese/Korean linguistics, Vol. 17*, 549–563. Stanford: CSLI Publications.

Baker, Mark C. 1996. *The polysynthesis parameter*. Oxford: Oxford University Press.

Belletti, Adriana. 2004. Aspects of the low IP area. In Luigi Rizzi (ed.), *The structure of CP and IP: The cartography of syntactic structures, Vol. 2*, 16–51. Oxford: Oxford University Press.

Bošković, Željko. 1999. On multiple feature-checking: Multiple wh-fronting and multiple head-movement. In Samuel Epstein & Norbert Hornstein (eds.), *Working minimalism*, 159–187. Cambridge: MIT Press.

Bošković, Željko. 2001. On the interpretation of multiple questions. *Linguistic Variation Yearbook* 1(1). 1–15. DOI: 10.1075/livy.1.03bos.

Bošković, Željko. 2002. On multiple wh-fronting. *Linguistic Inquiry* 33(3). 351–383. DOI: 10.1162/002438902760168536.

Bošković, Željko. 2007. A note on wh-typology. In Peter Kosta & Lilia Schürcks (eds.), *Linguistic investigations into formal description of Slavic languages: Contributions of the 6th European conference held at Potsdam University*, 159–170. Frankfurt/M.: Peter Lang.

Bošković, Željko. 2008. What will you have, DP or NP? In Emily Elfner & Martin Walkow (eds.), *NELS 37: Proceedings of the 37th annual meeting of the North East Linguistic Society*, 101–114. Amherst: GLSA Publications.

Bošković, Željko. 2012. On NPs and clauses. In Günther Grewendorf & Thomas Ede Zimmermann (eds.), *Discourse and grammar: From sentence types to lexical categories*, 179–242. Berlin: Mouton de Gruyter. DOI: 10.1515/9781614511601.179.

Bošković, Željko. 2016. On second position clitics crosslinguistically. In Franc Lanko Marušič & Rok Žaucer (eds.), *Formal studies in Slovenian syntax: In honor of Janez Orešnik*, 23–44. Amsterdam: John Benjamins. DOI: 10.1075/la.236.02bos.

Bošković, Željko. 2020. *Comparative syntax lectures, fall 2020*. University of Connecticut.

Cheng, Lisa Lai Shen. 1991. *On the typology of wh-questions*. Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Crnič, Luka. 2011. *Getting even*. Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Dadan, Marcin. 2019. *Head labeling preference and language change*. Storrs: University of Connecticut. (Doctoral dissertation).

Dench, Alan. 1981. *Panyjima phonology and morphology*. Canberra: The Australian National University. (MA thesis).

Dench, Alan. 1987. *Martuthunira: A language of the Pilbara region of Western Australia*. Canberra: The Australian National University. (Doctoral dissertation).

Diesing, Molly. 2003. On the nature of multiple fronting in Yiddish. In Cedric Boeckx & Kleanthes Grohmann (eds.), *Multiple wh-fronting*, 51–76. Amsterdam: John Benjamins. DOI: 10.1075/la.64.06die.

É. Kiss, Katalin. 2002. *The syntax of Hungarian*. Cambridge: Cambridge University Press.

Enç, Mürvet. 2003. *The syntax of specificity*. Ms. University of Wisconsin-Madison.

Erlewine, Michael Yoshitaka. 2016. Anti-locality and optimality in Kaqchikel agent focus. *Natural Language & Linguistic Theory* 34. 429–479. DOI: 10.1007/s11049-015-9310-z.

Erschler, David. 2012. From preverbal focus to preverbal ''left periphery'': The ossetic clause architecture in areal and diachronic perspective. *Lingua* 122(6). 673–699. DOI: 10.1016/j.lingua.2012.01.009.

Erschler, David. 2015. Embedded questions and sluicing in Georgian and Svan. *Languages of the Caucasus* 1(1). 38–74. DOI: 10.5070/L911029246.

Gan, Linghui. 2022. *Syntactic structure of argument wh-questions in Hong Kong Sign Language (HKSL)*. Presentation at the 28th Annual Conference of The International Association of Chinese Linguistics (IACL-28), Hong Kong. https://drive.google.com/file/d/1DFwkndEPovMonqPa0ibrhg_pw-CAF7HQ/view.

Giannakidou, Anastasia. 2007. The landscape of EVEN. *Natural Language & Linguistic Theory* 25. 39–81. DOI: 10.1007/s11049-006-9006-5.

Gillon, Carrie & Solveiga Armoškaitė. 2015. The illusion of the NP/DP divide. *Linguistic Variation* 15(1). 69–115. DOI: 10.1075/lv.15.1.03gil.

Grohmann, Kleanthes. 2006. Top issues in questions: Topics – topicalization – topicalizability. In Lisa Cheng & Norbert Corver (eds.), *Wh-movement moving on*, 249–288. Cambridge: MIT Press. DOI: 10.7551/mitpress/7197.003.0017.

Hagstrom, Paul. 1998. *Decomposing questions*. Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.

Hengeveld, Kees, Sabine Iatridou & Floris Roelofsen. 2023. Quexistentials and focus. *Linguistic Inquiry* 54(3). 571–624. DOI: 10.1162/ling_a_00441.

Horváth, Julia. 1998. Multiple wh-phrases and the wh-scope-marker strategy in Hungarian interrogatives. *Acta Linguistica Hungarica* 45. 31–60. DOI: 10.1023/A:1009656907847.

Izvorski, Roumyana. 1998. Non-indicative wh-complements of possessive and existential predicates. In Pius N. Tamanji & Kiyomi Kusumoto (eds.), *Proceedings of NELS 28*, 159–173. Amherst: GLSA Publications.

Kinjo, Kunio & Yohei Oseki. 2016. Wh-concord in Okinawan = syntactic movement + morphological merger. *University of Pennsylvania Working Papers in Linguistics* 22(1). 177–186. https://repository.upenn.edu/handle/20.500.14332/45087.

Kuroda, Sige-Yuki. 1965. *Generative grammatical studies in the Japanese language*. Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Lacerda, Renato. 2020. *Middle-field syntax and information structure in Brazilian Portuguese*. Storrs: University of Connecticut. (Doctoral dissertation).

Ledgeway, Adam. 2012. *From Latin to Romance: Morphosyntactic typology and change*. Oxford: Oxford University Press.

Lewis, Beccy. 2024. *Cross-linguistic variation in the expression of heterogeneous plural marking and its theoretical consequences*. Storrs: University of Connecticut. (Doctoral dissertation).

Lipták, Anikó. 2001. *On the syntax of wh-items in Hungarian* (LOT Dissertation Series 45). Utrecht: LOT.

MacWhinney, Brian. 1976. Hungarian research on the acquisition of morphology and syntax. *Journal of Child Language* 3(3). 397–410. DOI: 10 . 1017 / S0305000900007261.

Margolis, Rebecca. 2011. *Basic Yiddish: A grammar and workbook*. London: Routledge.

Mycock, Louise. 2007. Constituent question formation and focus: A new typological perspective. *Transactions of the Philological Society* 105. 192–251. DOI: 10.1111/j.1467-968X.2007.00188.x.

Oda, Hiromune. 2022. *The NP/DP-language distinction as a scale and parameters in minimalism*. Storrs: University of Connecticut. (Doctoral dissertation).

Pesetsky, David. 1987. Wh-in situ: Movement and unselective binding. In Eric J. Reuland & Alice G. B. ter Meulen (eds.), *The representation of (in)definiteness*, 98–129. Cambridge: MIT Press.

Puskás, Genoveva. 2000. *Word order in Hungarian: The syntax of A-bar positions*. Amsterdam: John Benjamins.

Richards, Norvin. 2001. *Movement in language: Interactions and architectures*. Oxford: Oxford University Press.

Rooth, Mats. 1985. *Association with focus*. Amherst: University of Massachusetts. (Doctoral dissertation).

Rudin, Catherine. 1988. On multiple questions and multiple wh fronting. *Natural Language & Linguistic Theory* 6. 445–501. DOI: 10.1007/BF00134489.

Shields, Rebecca. 2008. *What's so special about d-linking*. Paper presented at NELS 39, Cornell University. https://www.researchgate.net/publication/252605974_ What's_so_Special_about_D-Linking.

Šimík, Radek. 2011. *Modal existential wh-constructions* (LOT Dissertation Series 269). Utrecht: LOT.

Stepanov, Arthur. 1998. On wh-fronting in Russian. In Pius N. Tamanji & Kiyomi Kusumoto (eds.), *NELS 28: Proceedings of the 28th annual meeting of the North East Linguistic Society*, 453–467. Amherst: GLSA Publications.

Surányi, Balázs. 2005. Triggering wh-fronting. In Christopher Piñón & Siptár Péter (eds.), *Approaches to Hungarian Vol. 9: Papers from the Düsseldorf conference*, 231–259. Budapest: Akadémiai Kiadó.

Talić, Aida. 2017. *Functional and bare projections in the domain of N and A*. Storrs: University of Connecticut. (Doctoral dissertation).

Wachowicz, Krystyna A. 1974. Against the universality of a single wh-question movement. *Foundations of Language* 11. 155–166. https://www.jstor.org/stable/ 25000765.

Wang, Shuyan. 2019. An intermediate stage of spec-to-head reanalysis: Evidence from Mandarin and Cantonese. In Sae-Youn Cho (ed.), *Proceedings of the 12th Generative Linguistics in the Old World in Asia & the 21st Seoul international conference on generative grammar*, 594–603. Seoul: Hankook Munhwasa.

Zanon, Ksenia. 2022. *Wh-indefinites in Russian.* Paper presented at FDSL 15, HU Berlin.

Zwicky, Arnold M. 1984. "Reduced words" in highly modular theories: Yiddish anarthrous locatives reexamined. *Ohio State University Working Papers in Linguistics* 29. 117–126.

Chapter 6

A quantification-based approach to plural pronoun comitatives

Madeleine Butschety

University of Nova Gorica

Plural pronoun constructions (PPCs) consist of a plural pronoun and a comitative (i.e. *with*-) phrase. In sentences such as *My s Petej pojdëm domoj* (lit. 'We with Petja will-go home') from Russian, PPCs are ambiguous between a default interpretation according to which a plural referent *we* will go home with Petja (=ePPC), and an unexpected interpretation according to which Petja and the speaker will go home (=iPPC). I show that this ambiguity can be derived under the assumption that plural pronouns and (universal) quantifiers have some striking properties in common. In particular, I argue that the unexpected iPPC reading arises if the comitative phrase occurs inside the restrictor of the plural pronoun (which is similar to a quantifier's restrictor), and an ePPC reading arises if it occurs elsewhere in the structure. My account further offers an explanation regarding the availability and distribution of iPPC interpretations within and across Slavic languages.

1 Introduction

Plural pronoun comitatives (often also dubbed "Plural pronoun constructions" in the literature; henceforth PPCs) are complex expressions that consist of a plural pronoun and a comitative (i.e. *with*-) phrase. In many Slavic languages, PPCs can give rise to two different interpretations, paraphrased as (a) and (b) in the examples below.

(1) My s   Petej   pojdëm   domoj.
     we with Petja.INST go.FUT.1PL home

(Russian; Vassilieva & Larson 2005: 101)

    a. 'We will go home with Petja.'

    b. 'I and Petja will go home.'

(2) *pro* S   Mariju   smo   otišli u muzej.   (Torlakian BCMS)
      with Maria.INST AUX.1PL went in museum

    a. 'We went to the museum with Maria.'

    b. 'Maria and I went to the museum.'

The availability of an apparent singular ('I') interpretation of the (dropped) pronoun under reading (b) is unexpected given that its surface form is plural.[1] This reading (henceforth: iPPC; following Feldman's 2003 terminological distinction) contrasts with the default 'we' interpretation (henceforth: ePPC) of the plural pronoun under the reading paraphrased in (a).[2] In the light of the ambiguity between (a) and (b), it may seem tempting to assume that plural pronouns are ambiguous between a singular and a plural interpretation in general. As Vassilieva & Larson (2001) pointed out already, however, this cannot be the case. Constructions like (3) not involving a comitative phrase do not give rise to the ambiguity observed for the otherwise parallel example (1).[3]

(3) My pojdëm   domoj.   (Russian; Vassilieva & Larson 2001: 449)
    we go.FUT.1PL home

    a. 'We will go home.'

    b. *'I will go home.'

Common analyses have usually taken one of two explanations: either that the reference of a plural pronoun in a PPC is composed of its singular counterpart

---

[1]To talk about the pronoun's "surface form" is actually a bit misleading here. Example (2) from Torlakian BCMS involves *pro*-drop. So, strictly speaking, there is no such thing as a surface form of the plural pronoun in this sentence. Although this example is also felicitous in its version with an overt plural pronoun *mi* 'we', I omit it here and in the succeeding examples for reasons that will be discussed in Section 4. Furthermore, we can infer the "underlying" form, or rather the features of *pro* from the features on the verb – which is 1st person plural in both readings of (2).

[2]The distinction between "iPPC" and "ePPC" as made in Feldman (2003) refers to an "inclusive" and an "exclusive" interpretation of the plural pronoun, respectively. That is, under an "inclusive" interpretation of a PPC, the referent from the comitative phrase (e.g. *Petja* in (1)) is apparently included in the overall reference of the plural pronoun; whereas under an "exclusive" interpretation, the reference of the plural pronoun does not include the referent from the comitative phrase – i.e., under the ePPC reading of (1), the 1st person plural pronoun refers to the speaker and someone else (but Petja).

[3]The original transliteration of example (3) was altered to be in line with the scientific transliteration of Cyrillic.

and the referent of the comitative (cf. Vassilieva & Larson 2001, Vassilieva & Larson 2005), or that PPCs involve asymmetric coordination from a syntactic perspective, but have the same (or similar) semantics as symmetric coordination (cf. Dyła 1988, McNally 1993, Feldman & Dyła 2008). I will spell out the core assumptions of these and other approaches in Section 2.

A real challenge for pretty much any theory of PPCs (apart from apposition-based ones such as Cable 2017, perhaps) is that the plural pronoun and the comitative phrase can occur as a discontinuous constituent. And yet, they give rise to iPPC interpretations (at least in most Slavic languages). Consider the contrast between (4) and (5) from Torlakian, which is mainly syntactic but not semantic.

(4) *pro* S Mariju smo otišli u muzej. (Torlakian BCMS)
with Maria.INST AUX.1PL went in museum

   a. 'We went to the museum with Maria.' ePPC

   b. 'Maria and I went to the museum.' iPPC

(5) *pro* Otišli smo u muzej s Mariju. (Torlakian BCMS)
went AUX.1PL in museum with Maria.INST

   a. 'We went to the museum with Maria.' ePPC

   b. 'Maria and I went to the museum.' iPPC

To the best of my knowledge, no analysis has yet come up with a neat answer to the question of how an iPPC reading can arise for such "split" PPCs like (5). My attempt to explain this fact will be outlined in Section 3.2.1. In particular, I propose a novel analysis of PPCs in Slavic languages that is based on the assumption that plural pronouns and (universal) quantifiers behave alike in various respects. In particular, I argue that the difference between an iPPC and an ePPC interpretation can be derived in terms of whether the comitative phrase resides inside the restrictor of the plural pronoun (which is similar to a quantifier's restrictor) or not. My analysis is essentially based on data from Torlakian BCMS and from Bulgarian. These two languages are very suitable as a starting point, as they have different properties regarding available readings for split PPCs – and from these differences, I derive one of the core assumptions of my analysis.[4] Moreover, to the best of my knowledge, nothing has been said in the literature yet about the behavior of PPCs in Torlakian BCMS and Bulgarian.

---

[4]However, I also occasionally show how (and to what extent) the generalizations derived from those distinct features are applicable to other Slavic languages such as Russian or Polish.

This article proceeds as follows. In Section 2 I briefly summarize the main findings and claims of previous approaches to PPCs, and point out a few problems that arise from each. I present my proposal in Section 3. First, in Section 3.1, I present my analysis of the internal structure of plural pronouns in general. I claim that plural pronouns, just as quantifiers, select a restrictor argument and that it is the precise internal structure of this restrictor which determines the overall referential properties of the respective plural pronoun. In Section 3.2 I then show the implications of those assumptions for an analysis of PPCs. Specifically, I argue that the comitative phrase occurs inside of the plural pronoun's restrictor in iPPCs, but outside of it (that is, elsewhere in the syntactic structure) in ePPCs. I outline my explanation of why split PPCs can have iPPC interpretations in some Slavic languages in Section 3.2.1. Furthermore, I also offer an explanation as to why we find iPPC readings for split PPCs in precisely those Slavic languages in which we find them (such as Torlakian BCMS), but not in others (such as Bulgarian). In Section 3.2.2 and Section 3.2.3 I present data in favour of my analysis related to Subject Control constructions and binding. However, I want to point out in advance that the data from those two sections could probably also be correctly derived under other approaches to PPCs. Nevertheless, it is my intention to discuss them rather than to leave them as implicit evidence. So in the end, the virtue of my analysis is that it can properly predict when, how, and why iPPC readings for split PPCs arise – and thus, that it fills an explanatory gap that exists among previous analyses of PPCs.

2 Previous analyses of PPCs in Slavic

Existing approaches to PPCs can basically be divided into three categories: those assuming that PPCs have an underlying coordinative structure; those which treat PPCs as a kind of appositive construal involving ellipsis; and finally, those which are based on the idea that the comitative phrase acts as a complement of the plural pronoun in such constructions. In this section, I briefly illustrate the main claims of previous analyses as well as some issues that remain open in the light of the respective theories. My own analysis draws on the assumptions of analyses of the third category, i.e. on approaches which treat the comitative phrase as a complement of the plural pronoun. However, my proposal takes one step further in observing and implementing some parallels between plural pronouns and quantifiers.

2.1 Based on coordination

As has been observed in the literature on (primarily) Russian and Polish, "ordinary" or "regular" comitative constructions (i.e. those which are not headed by a plural pronoun) behave differently with regard to whether they trigger singular or plural agreement on the verb; see example (6). Traditional analyses (cf. Dyła 1988, McNally 1993, Feldman & Dyła 2008) anchor the alternation of verbal number agreement in diverging underlying syntactic structures. Specifically, those comitatives that trigger plural agreement on the verb are considered to be coordinative (that is, conjunctive) construals, whereas comitatives that trigger singular agreement on the verb are commonly treated as adjuncts to VP.[5]

(6) a. Maša   s   Dašej   xodjat v školu.
    Maša.NOM with Daša.INST go.1PL to school
    'Maša and Daša go to school.'   *s*-CONJUNCTION

   b. Anja   s   Vanej   pošla   v biblioteku.
    Anja.NOM with Vanja.INST went.SG.F to library
    'Anja went to the library with Vanja.'   *s*-ADJUNCTION
                                                (Russian; Feldman 2003)

I do not want to enter this debate to any extent and I will have little to say about such "regular" comitatives in this article. My account is based on the specific properties of plural pronouns and does not intend to make generalizations to any other kinds of DPs. Nonetheless, both structural options from (6) have been considered in the light of PPCs – although coordination-based accounts (such as the influential paper by McNally 1993) are more numerous than analyses that assume that PPCs have an adjunction structure (see Ladusaw 1989).

A recent example of a coordinative treatment of PPCs is Sokolová's (2019) analysis. It is claimed there that Slovak PPCs classify as "coordinate comitatives", i.e. the plural pronoun DP and the comitative phrase are assumed to have the same structural rank. Under Sokolová's (2019) approach, the denotation of a plural pronoun such as *my* 'we' in (7) consists of two referents (or "participants", as dubbed in the original article) – namely the speaker (= first participant), and a second participant that gets lexically specified by the referent of the comitative.

(7) My s   Evou   chodíme do rovnakej školy.
   we with Eva.INST go.1PL   to same   school
   'Eva and I go to the same school.'   (Slovak; Sokolová 2019: 101)

---

[5]Transliterations were adjusted here.

The iPPC interpretation of a sentence like (7) is assumed to arise on the basis of an "absorption" mechanism. The apparent double occurrence of the second participant conflates into one via the overall reference of the plural pronoun. Sokolová (2019) is not explicit about how this "absorption of a referent" (a notion based on Daniel 2000) is supposed to work out on a derivational level, however. Likewise, Daniel's (2000) account is not concerned with a precise formal underpinning. The original idea, though, is that plural pronouns are only explicit about the first referent in general (such as "speaker" or "addressee" in the case of 'we' and 'you$_{PL}$', respectively). Under some discourse conditions, an explication of the second participant, which otherwise remains implicit, is required. This explication can be resolved via different means in (morpho-)syntax; and so in the end, PPCs are just one (morpho-)syntactic conventionalization to feed these pragmatic needs for specification of the other referent of the plural pronoun (Mikhail Daniel, p.c.).[6]

Analyses such as McNally's (1993), on the other hand, assume an asymmetric kind of coordination. McNally (1993) does not deal with PPCs in particular, but suggests that PPCs such as (8a) and "regular" comitative constructions like (8b) share the same underlying structure illustrated in Figure 1.

(8) a. Oni   s Petej   pridut.
        they.NOM with Petja.INST come.3PL
        'He and Petja are coming.   (Russian; McNally 1993: 359)

     b. Anna   s Petej   napisali pis'mo.
        Anna.NOM with Petja.INST wrote.3PL letter
        'Anna and Petja wrote a letter.'   (Russian; McNally 1993: 347)
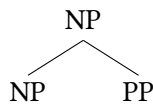
Figure 1: Asymmetric coordination structure (McNally 1993: 359)

---

[6]Note that an explanation along these lines, once incorporated into a formal discourse framework, would probably suffice the purpose of predicting when 2p-interpretations arise – namely if and only if the identity of the second participant otherwise remains unclear. Evidence in favor of this view comes from infelicitous sequences such as (i).

(i) Nie otidohme v muzeja. #Večerta nie s   Peter gledahme   star Disni film.
     we went.1PL in museum evening we with Peter watched.1PL old Disney movie
     'We$_i$ went to the museum. In the evening, #[I and Peter]$_i$ watched an old Disney movie.'
                                                                              (Bulgarian)

McNally (1993) claims that from a semantic point of view, the structure in Figure 1 has the same interpretation as a structure involving symmetric coordination.

Several problems arise from a coordination-based approach to PPCs. A salient challenge has to do with the fact that the plural pronoun can be dropped in a PPC (in Slavic languages that permit *pro*-drop, such as Torlakian BCMS, Bulgarian, Polish, or Slovak), but not in other (regular) forms of conjunction. But if we assume that a PPC is (structurally speaking) a conjunction of the plural pronoun and the comitative phrase, then we would expect that PPCs and regular coordination show the same syntactic pattern – contrary to the actual facts. Dropping the first conjunct (even if pronominal) in regular coordination is impossible, compare (9a) and (9b) from Polish.

(9)   a.   On/*pro* z   bratem poszli do kina.
          he   with brother went to cinema
          'He and his brother went to the cinema.' (Polish; Trawinski 2005: 385)

     b.   On/**pro* i   Maria poszli do kina.
          he   and Maria went to cinema
          (Intended): 'He and Maria went to the cinema.'
                                                      (Polish; Trawinski 2005: 384)

Moreover, it remains puzzling why the order of plural pronoun and the other DP involved is fixed in PPCs; i.e. why a construction like (10a) is ungrammatical. Because, as can be seen from (10b), such a reversed order is (although slightly deviant) sometimes possible in coordinated structures.

(10)   a.   * Petja   s   nami   tancevali.
             Petja.NOM with us.INST danced.PL
             Intended: 'Petja and us danced.'

      b.   ? Petja   i   ty   tancevali.
             Petja.NOM and you.NOM danced.PL
             'Petja and you danced.'   (Russian; Vassilieva & Larson 2005: 114)

Under the assumption that PPCs have an underlying structure like the one in Figure 1, we could potentially explain these differences to regular conjunction structures in terms of asymmetric coordination. But to the best of my knowledge, no such account has been presented yet, and thus, the exact reasons for those differences remain to be spelled out in detail.

2.2 Based on apposition

This line of approach originates from Cable (2017). The basic claim of Cable's (2017) analysis is that PPCs are appositive constructions involving ellipsis. In particular, he assumes that PPCs in their iPPC versions contain an elided instance of the respective plural pronoun's singular counterpart, such that a sentence like (11a) has the underlying structure in (11b).

(11) a. My   s   Petej   pojdëm   domoj.
          we.ɴᴏᴍ with Petja.ɪɴsᴛ go.ꜰᴜᴛ.1ᴘʟ home
          'Petja and I will go home.'   (Russian; Vassilieva & Larson 2005: 101)

    b. My   [⟨ja⟩   s   Petej]   pojdëm   domoj.
          we.ɴᴏᴍ   I.ɴᴏᴍ with Petja.ɪɴsᴛ go.ꜰᴜᴛ.1ᴘʟ home
          Approx.: 'We, I and Petja, will go home.'   (Russian; Cable 2017: 8)

A plural pronoun in a PPC is taken to denote the sum of the two referential expressions occurring in the apposition, i.e. the sum of *ja* ('I') and *Petja* in (11b).

While such an account seems appealing, it is not without complications. First of all, it raises the question of what licenses ellipsis in the apposition. It cannot be deletion under identity in a strict sense (see Lipták 2015 for an overview, and the references therein), because according to Cable's (2017) proposal, the plural pronoun denotes the sum of the elided element and the comitative referent. Hence, if anything, deletion should target both *ja* and *Petej* in (11b); and then, there would be no overt material left in the appositive apart from the comitative element *s* itself. At best, we could assume that only *ja* gets elided because ellipsis would have to target a discontinuous constituent otherwise, or because it bears the same grammatical case as the plural pronoun (whereas *Petej* bears instrumental case, assigned by the comitative element). It is hard to figure out which deletion mechanisms are supposed to be at stake in a structure like (11b) – especially as long as the assumptions concerning the appositive structure are not embedded in any tradition of analyzing appositions. However, Cable (2017) seems to be aware of this problem and mentions that it could also be the case that the comitative element itself has a meaning akin to appositive structure. How this meaning would need to be defined in particular is left as an open issue.

Furthermore, it seems that phonological deletion poses a problem to any apposition-based approach to (Slavic, at least) PPCs in general. Because as varied as theories of appositives may be, there is a syntactic property which appositions are uncontroversially assumed to have. Namely, that there is an anchor expression to which the appositive attaches. The anchor expression of the apposition ⟨*ja*⟩ *s Petej* from (11b) would (also quite undisputedly) have to be *my*, i.e. the plural

pronoun. While this is unproblematic for the Russian case at hand, we run into complications as soon as we want to apply this analysis to Slavic languages that allow *pro*-drop structures. Take example (12) from Torlakian BCMS, for instance. The version in (12a) with the overt plural pronoun could be treated along the lines of (11b), as shown in (12b). But for its *pro*-dropped version in (13a), this does not work out. Because we would have to assume that appositions can attach to silent anchor expressions, as pictured in (13b).[7]

(12) a. Juče   smo   mi   s   Mariju   otišli u bioskop.
          yesterday AUX.1PL we.NOM with Maria.INST went in cinema

     b. Juče   smo   mi   [⟨ja⟩ s   Mariju]   otišli u bioskop.
          yesterday AUX.1PL we.NOM   I.NOM with Maria.INST went in cinema
     Approx.: 'Yesterday we, I and Maria, went to the cinema.'

                                                        (Torlakian BCMS)

(13) a. Juče   smo   *pro* s   Mariju   otišli u bioskop.
          yesterday AUX.1PL   with Maria.INST went in cinema

     b. # Juče   smo   *pro* [⟨ja⟩ s   Mariju]   otišli u bioskop.
            yesterday AUX.1PL   I.NOM with Maria.INST went in cinema
     Intended: 'Yesterday we, I and Maria, went to the cinema.'

                                                        (Torlakian BCMS)

If it were indeed possible to adjoin an apposition or any kind of additional syntactic material (such as a PP, for instance) to an attachment site that is not overtly present in the syntactic structure we would expect to find constructions such as (14b) or (14d) regularly. But as we can see, this expectation is not met.

(14) a. Mi,   (naime) ja   s   Petra,   idemo sutra   u bioskop.
          we.NOM   namely I.NOM with Peter.INST go.1PL tomorrow in cinema
          'We, namely I and Peter, will go to the cinema tomorrow.'

     b. # *pro* Naime ja   s   Petra,   idemo sutra   u bioskop.
               namely I.NOM with Peter.INST go.1PL tomorrow in cinema
          Intended: 'We, namely I and Peter, will go to the cinema tomorrow.'

     c. Ona s   plavom kosom ide   sutra   u bioskop.
          she with blond hair   go.3SG tomorrow in cinema
          'She with the blond hair will go to the cinema tomorrow.'

---

[7]The hashtag # in (13b) is intended to indicate that it should be regarded with suspicion whether such a syntactic configuration is even possible.

    d. # *pro* S   plavom kosom ide   sutra   u bioskop.
             with blond   hair   go.3sg tomorrow in cinema
          'She will go to the cinema tomorrow with the blond hair.'
          Intended: 'She with the blond hair will go to the cinema tomorrow.'

                                              (Torlakian BCMS)

The sentence in (14b) is ungrammatical in an out-of-the-blue context as well as in any syntactic context that does not contain a salient anchor or anaphoric expression. More precisely, the only syntactic environment which could save (14b) is a preceding sentence like *We/The best of friends will go to the cinema – naime ja s Petra....* Example (14d) on the other hand is simply infelicitous. It cannot mean what it is supposed to, i.e. that he/she and the blond-haired he/she will go to the cinema tomorrow. Instead, the sentence only has a rather awkward interpretation according to which she (an individual salient from the discourse) will go to the cinema tomorrow taking the blond hair (literally) with her.

    Abstracting from (im)possible anchor expressions, an apposition-based analysis makes unwelcome predictions regarding the spell-out of an iPPC. Appositives are usually articulated with an intonational/phonological break or boundary. Thus, if for instance (11a) has the underlying structure in (11b), such a break or boundary should reflect in the pronunciation of such a sentence; see (15), where "(–)" signals the points where the breaks should occur.

(15)   My   (–) s   Petej   (–) pojdem   domoj.
       we.nom   with Petja.inst   go.fut.1pl home
       'We, with Peter, will go home.'

This prediction is not borne out, since a PPC (no matter which reading is intended) does not surface any kind of apposition-typical intonational breaks in an unmarked context. It rather has a regular pronunciation.

2.3 Comitatives as pronoun complements

The analyses in Vassilieva & Larson (2001) and Vassilieva & Larson (2005) treat (Russian) plural pronouns as incomplete expressions, comprising a singular nucleus and an unsaturated element △ in their meaning. These two components are elements of an ordered pair $\langle X, Y \rangle$. X is obligatorily taken by the plural pronoun's singular counterpart (in accordance with its person feature), and Y gets (per default) saturated by some σ from the context. The resulting (distributive)

semantics of Russian *my* 'we', *vy* 'you.PL' and *oni* 'they' can be represented as in (16) below.[8]

(16)   a.   VAL($\langle X, Y \rangle$, [$_D$ *my*], σ) iff $|(\{\sigma(a)\} \cup Y) - X| = 0$
         '(all of) speaker + others Y'

      b.   VAL($\langle X, Y \rangle$, [$_D$ *vy*], σ) iff $|(\{\sigma(b)\} \cup Y) - X| = 0$
         '(all of) addressee + others Y'

      c.   VAL($\langle X, Y \rangle$, [$_D$ *oni*], σ) iff $|(\{\sigma(i)\} \cup Y) - X| = 0$
         '(all of) he/she/it + others Y'

                                         (Vassilieva & Larson 2005: 119)

With regard to PPCs and their iPPC interpretations, the idea outlined in Vassilieva & Larson (2005) is that the comitative phrase occupies a complement position to the plural pronoun. The Y-slot in the meaning of the plural pronoun then gets filled by the referent of the comitative, i.e. as a matter of syntactic means. The underlying structure of the iPPC reading is given in Figure 2, and its semantics in (17).
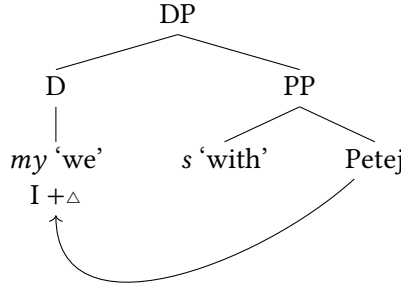
Figure 2: iPPC structure (Vassilieva & Larson 2005: 120)

(17)   VAL($\langle X$, *my s Petej*, $\sigma \rangle$ iff $|(\{\sigma(a)\} \cup \{\text{Petja}\}) - X| = 0$; i.e.
      VAL($\langle X$, *my s Petej*, $\sigma \rangle$ iff $|(\{\sigma(a), \text{Petja}\}) - X| = 0$
      '(all of) speaker + Petja'   (Vassilieva & Larson 2005: 120)

---

[8]The formalism that Vassilieva & Larson (2001) and Vassilieva & Larson (2005) make use of was adopted from Larson & Segal (1995), where it is assumed that the truth value assigned to a sentence is dependent on context sequences σ – such that $\sigma(a), \sigma(b), \sigma(c)$ and $\sigma(d)$ are associated with speaker, addressee, speaker time and speaker location, respectively, from the context. Indexical pronouns are claimed to get their values via these means. Other instances of such a sequence, dubbed $\sigma(n)$, determine the reference of non-indexical pronouns on the other hand. Since my analysis does not make use of this particular formal framework, I will not go into this topic any further here.

Hence, Vassilieva & Larson tie the availability of an iPPC interpretation to the specific syntactic configuration depicted in Figure 2. However, there are two pieces of data that they explicitly leave unaccounted for. Firstly, it remains unclear under this analysis why PPCs (in their iPPC meaning) can occur split, i.e. why the plural pronoun and the comitative phrase can be discontinuous and still an iPPC interpretation is available; see (18a). Secondly, it is puzzling why *wh*-questions such as (18b) do actually not give rise to an iPPC meaning.

(18) a. My   pojdëm   zavtra   s   Ivanom   v magazin i   vsë
          we.NOM go.FUT.1PL tomorrow with Ivan.INST to store   and all
          kupim.
          buy.FUT
          'Tomorrow, we will go to the store with Ivan and buy all we need.'
                                                                        ePPC
          'Tomorrow, Ivan and I will go to the store and buy all we need.' iPPC

     b. S   kem   my xodili v magazin?
          with whom we went to store
          'With whom did we go to the store?'   ePPC
          Unavailable iPPC reading: 'I and who went to the store?'
                                       (Russian; Vassilieva & Larson 2005: 122)

If the comitative phrase acts as a complement of the plural pronoun, we do not expect (18a) to give rise to an iPPC reading, because the plural pronoun and the comitative phrase do not form a constituent in this sentence – yet, an iPPC interpretation is available. This poses a serious challenge to Vassilieva & Larson's (2005) analysis since the availability of an iPPC reading for split PPCs is widespread across Slavic languages (with some exceptions). My analysis, to which we turn next, offers an explanation of these facts as well as a novel perspective on PPCs in general.

3 Proposal

In this section, I present my approach to PPCs in Slavic. My analysis is based on observations regarding similarities in the semantic and syntactic behaviour of plural pronouns and quantifiers; or rather, of plural pronominal DPs and quantificational (noun) phrases (henceforth: QPs).[9] Therefore I first outline my claims regarding the internal structure of plural pronouns in Section 3.1 and show how it

---

[9]It is important to note here that I do not claim that plural pronouns are quantificational elements (in a broad sense) by definition. While these two types of expressions exhibit some

is similar to the internal structure of QPs. On this basis, I derive the two readings of PPCs (i.e. iPPC and ePPC) in Section 3.2 and then illustrate in Section 3.2.1 why and how split PPCs give rise to iPPC interpretations in some Slavic languages, but not in others. Data related to Subject Control structures and binding are presented in Section 3.2.2 and Section 3.2.3, respectively.

3.1 Restrictor sets for plural pronouns

Plural pronouns have often been treated analogously to definite descriptions in that both denote pluralities of individuals (cf. Link 1983, Nunberg 1993, Elbourne 2008, Büring 2011, among many others). If we leave figurative uses aside, then this plurality obligatorily includes the speaker (1st person plural pronoun) or addressee (2nd person plural pronoun) of the utterance. In my analysis, I follow Link (1983) and others in assuming that plural pronouns denote pluralities of individuals. But the suggested way in which this plurality gets composed differs from previous approaches. I propose a slightly modified variant of the "plural pronouns as definite descriptions" view. In particular, I claim that plural pronouns are more similar to quantifiers after all. What I am arguing for in this and subsequent sections is that plural pronouns and (universal) quantifiers have some striking properties in common – syntactic, semantic, and pragmatic ones.

Let us start with a very basic yet deep parallel regarding the syntactic arguments these expressions take. Quantifiers are uncontroversially assumed to take two arguments: restrictor and scope. The former typically corresponds to the NP the quantifier forms a QP with. The latter typically corresponds to the predicate (i.e. the VP, roughly speaking). While it is probably also quite uncontroversial that a plural pronoun (in subject position) combines with a predicate (a scope argument so to speak), I claim here that the internal syntactic structure and semantic composition of such an expression also involves a restrictor argument. More specifically, the restrictor of a plural pronoun is argued to be the decisive factor in determining the plural pronoun's overall reference. So, what must the restrictor of a plural pronoun look like by analogy with the restrictor of a quantificational expression? And what semantic or pragmatic properties can we find?

There has been a vast debate in the literature on quantifiers concerning the question whether quantificational determiners presuppose that their restrictor

---

striking parallels, my analysis is not intended to treat them as completely analogous. That is, just because a (universal) quantifier has such and such properties or shows such and such behaviour (of a syntactic or semantic nature), this does not automatically also have to apply to plural pronouns. For the time being, I restrict the similarities between plural pronouns and quantifiers to what is explicitly diagnosed in the course of my analysis outlined in this article.

sets must not be empty (see Heim & Kratzer 1998, and Szabolcsi 2010 for an overview). By now, it is more or less commonly agreed that at least strong quantifiers such as *all, most,* or *each* presuppose non-emptiness of their restrictor's denotation. So if, as claimed here, requiring a restrictor argument is one of the properties that plural pronouns share with quantifiers, we would probably expect plural pronouns to trigger an equivalent presupposition regarding the denotation of their restrictor sets. While there seems to be nothing wrong in principle at first glance with the assumption that a plural pronoun like *we* presupposes that what it ranges over must not be the empty set, such a bare existential presupposition does not suffice under closer inspection. The reason is that non-emptiness alone does not account for referential properties related to a plural pronoun's person feature. Specifically, a felicitous use of *we* not only requires that the pronoun ranges over a plurality made up of whatever individuals, but rather over a plurality which obligatorily contains the speaker of the utterance. I will first and foremost tie these requirements directly to the lexico-semantic properties of plural pronouns.[10]

The lexical entries of *we* and *you*$_{\text{PL}}$ in (19) and those expressions' underlying structures in Figures 3–5 further illustrate the assumed parallel to a quantifier's restrictor.[11]

(19) a. $[\![we]\!]$, $[\![you_{\text{PL}}]\!] = \lambda P.\lambda Q.P \subseteq Q$
      b. $[\![\text{OP}_\cup]\!] = \lambda x.\lambda y.\lambda z.z \leq x \lor z = y$ \quad\quad such that $x \neq y$

The restrictor argument of a plural pronoun thus consists of three components: A silent instance of SPKR (i.e. reference to the speaker of the utterance) or ADDR (reference to the addressee of the utterance), a silent operator OP$_\cup$, and a contextual assignment function $g(i)$ which surfaces as a silent pronominal form *pro$_i$* (interpreted as $[\![pro_i]\!]^g$, i.e. $g(i)$ again) in the structure. The central idea is that

---

[10]So far, nothing hinges on this decision. I will not discuss any data that is concerned with presupposition projection or presupposition failure. Note, though, that presuppositional content of *we* that matches the conditions stated above has occasionally been suggested in the literature (see Stokke 2022 and the references therein).

[11]I concentrate on 1st and 2nd person plural pronouns only throughout this article. The reason for this restriction is twofold. On the one hand, 3rd person plural pronouns do not have a fixed first referent like 1st and 2nd person plural pronouns do (i.e. speaker or addressee, respectively). On the other hand, iPPC interpretations are often harder to obtain for PPCs in the Slavic languages I investigate, if they are available at all. And this, in turn, might be related to the lack of a fixed first referent. In particular, I found that iPPC readings are often judged infelicitous in out-of-the-blue contexts – and only in suitable contexts or in follow-up sentences to an explicit QUD highlighting which individual is intended to be the first referent, iPPC interpretations seemed more readily available.

DP

D

[1st.SG] / [2nd.SG]

OP$_\cup$ *pro$_i$*
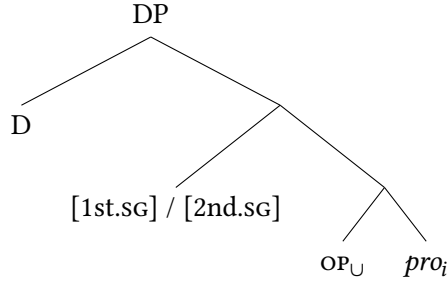
Figure 3: Underlying syntactic structure of 1st/2nd person plural pronoun

$\lambda Q.(\lambda z.z \leq g(i) \vee \text{SPKR}(z) \wedge \text{ATOMIC}(z)) \subseteq Q$

we

$\langle I \rangle$
SPKR

OP$_\cup$ *g(i)*
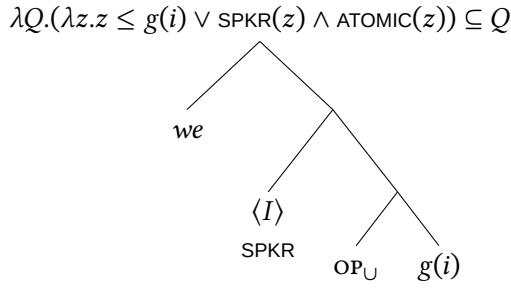
Figure 4: Semantic interpretation 1st person plural pronoun

$\lambda Q.(\lambda z.z \leq g(i) \vee \text{ADDR}(z) \wedge \text{ATOMIC}(z)) \subseteq Q$

you$_{\text{PL}}$

$\langle you_{\text{SG}} \rangle$
ADDR

OP$_\cup$ *g(i)*
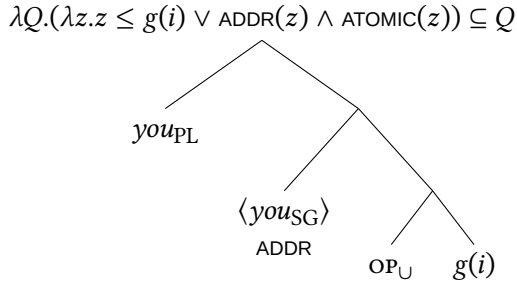
Figure 5: Semantic interpretation of 2nd person plural pronoun

the two referential instances within the restrictor are conflated by $\text{OP}_\cup$. Roughly speaking, the mechanism of $\text{OP}_\cup$ can be seen as set union. $\text{OP}_\cup$ takes two arguments x, y to form a z such that z is identical to y (i.e. the speaker SPKR or addressee ADDR) or z is less or equal to x (i.e. $g(i)$). Note that being less or equal to $g(i)$ will come out as plain $g(i)$ in any case. Actually, $\text{OP}_\cup$'s λz-part is just a formal workaround. Given how this function is defined it acts more or less like a type-shifter in that it takes two type e expressions as input and forming a set containing precisely and only those individuals. So what the 1st and 2nd plural pronouns refer to according to (19) can be paraphrased as 'the speaker and other(s)' and 'the addressee and other(s)', respectively.

3.2 Restrictors in PPCs

In the previous section, I suggested that plural pronouns have a (syntactically speaking) restrictor argument (and semantically speaking, introduce a restrictor set), and that this restrictor (or restrictor set) determines the plural pronoun's overall reference. The main hypothesis of my proposal regarding PPCs is that the difference between an iPPC and an ePPC interpretation boils down to whether the comitative phrase resides inside the restrictor of the plural pronoun or not – in the former case, an iPPC reading arises, whereas in the latter case, only an ePPC reading should be available. In particular, the structure of a PPC such as Russian (20) under its iPPC meaning is as pictured in Figure 6.

The crucial aspect in Figure 6 is that the comitative element *s* 'with' is assumed to have the very same semantics as $\text{OP}_\cup$ and that the referent of the comitative phrase occupies the position of the pronominal element interpreted as $g(i)$ in the default plural pronominal structure. That is to say that the comitative phrase *s Petej* 'with Petja.INST' as a whole acts as a spell-out of the more general $\text{OP}_\cup$ plus $g(i)$-part from Figure 3 – and consequently, *my* in the structural configuration (Figure 6) refers to just the speaker and Petja. Put differently, the comitative element serves the same purpose as $\text{OP}_\cup$ here, i.e. it forms a set from its two arguments, namely, the speaker and Petja.

(20) My   s   Petej   …
      we.NOM with Petja.INST
      Intended (iPPC): 'I and Petja …'

However, how to derive the ePPC interpretation of (20) is not entirely straightforward. Note that we cannot simply adjoin the comitative phrase to the (fully determined) plural pronominal DP – because on the one hand, it would not be straightforward how to integrate the comitative phrase into the scope argument

where $[\![s]\!] = [\![\text{OP}_\cup]\!]$

$\lambda Q.(\lambda z.z \leq \text{Petej} \vee \text{SPKR}(z) \wedge \text{ATOMIC}(z)) \subseteq Q$ ...

my

⟨I⟩
SPKR

s   *Petej*
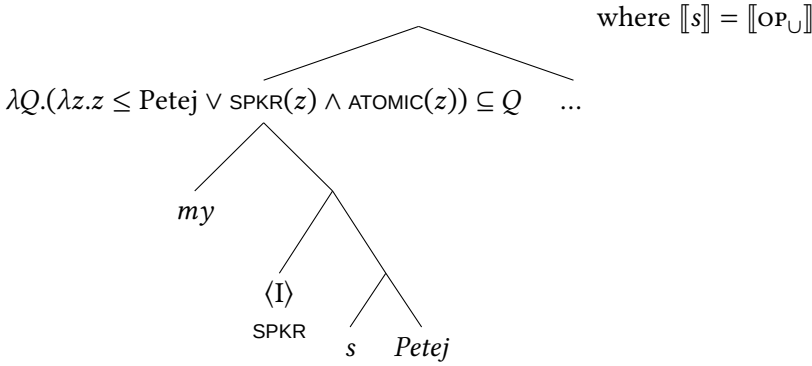
Figure 6: Semantic interpretation of iPPC structure

of the plural pronoun. On the other hand, we might run into incorrect predictions after all. The data that will be presented in the remainder of this section strongly suggests that the comitative phrase is actually rather an adjunct to VP in the ePPC cases. I thus claim at this point that we are dealing with adjunction to VP, and refer to the forthcoming sections for further evidence in favour of this view.

The predictions of this theory should now be straightforward: if and only if the comitative phrase occurs inside the restrictor of a plural pronoun, an iPPC interpretation arises. On the other hand, if the comitative phrase occupies a position adjacent to the VP, only an ePPC reading is predicted to be available. We turn to data from Torlakian BCMS and Bulgarian which support the suggested "quantificational" treatment of plural pronouns next. This choice regarding investigated languages is not due to arbitrary reasons. As will soon become clear, these two languages have very illustrative distinct properties when it comes to PPCs. Moreover, nothing has been said in the literature yet about PPCs in Bulgarian or Torlakian BCMS to the best of my knowledge – and hence, they are definitely worth a closer inspection.

To start with, we find a direct syntactic-semantic reflection of the predictions just stated in Bulgarian. In particular, a structure like (21a) in which the plural pronoun and the comitative phrase are "tied together" has an iPPC reading only. In contrast, a split PPC like (21b) exclusively gives rise to an ePPC interpretation.

(21) a. Nie/*pro* s   Peter otidohme v muzeja.   (Bulgarian)
          we.NOM with Peter went.1PL   in museum.DEF
          Unavailable ePPC reading: 'We went to the museum with Peter.'
          'Peter and I went to the museum.'   iPPC

b. Nie/*pro* otidohme v muzeja s Peter.
we.NOM went.1PL in museum with Peter
'We went to the museum with Peter.' ePPC
Unavailable iPPC reading: 'Peter and I went to the museum.'

In accordance with the comitative phrase's structural proximity to the plural pronoun, we could argue that *s Peter* occurs within the plural pronoun's restrictor in (21a), but not in (21b). But while Bulgarian exhibits a structure-meaning correspondence that perfectly matches the predictions of my theory, it happens to be the case that not all Slavic languages (and we might in fact be talking about a minority here) behave like Bulgarian in this respect. In Torlakian BCMS, for instance, both sentences from (22) are ambiguous between an iPPC and an ePPC interpretation. That is, no matter whether the plural pronoun and the comitative phrase appear structurally very close to one another or as separate constituents, both readings are available.

(22) a. *pro* s Mariju smo otišli u muzej. (Torlakian BCMS)
with Maria.INST AUX.1PL went in museum
'We went to the museum with Maria.' ePPC
'Maria and I went to the museum.' iPPC

b. *pro* otišli smo u muzej s Mariju.
went AUX.1PL in museum with Maria.INST
'We went to the museum with Maria.' ePPC
'Maria and I went to the museum.' iPPC

The crucial question is how to account for the iPPC interpretation of (22b) under the analysis proposed in this article. Because note that (22a)'s ePPC reading can be easily explained via the assumption that *s Mariju* is an adjunct to any suitable XP from the sentence except inside the plural pronominal DP, or directly attached to it.

3.2.1 Together vs. apart: Split PPCs as floated constructions

If a plural pronominal DP is indeed similar to a QP, then we would expect that these two phrasal expressions do not only share striking semantic properties, but also syntactic ones – such as the possibility to detach restrictor and head: Quantifier-floating (henceforth: Q-floating; for a general discussion see Sportiche 1988, Shlonsky 1991, Merchant 1996, Bošković 2004, Fitzpatrick 2006, among

many others).[12] I argue here that a similar movement operation is the source of (22b)'s iPPC interpretation, i.e. that the plural pronoun and the comitative phrase were base-generated within the same DP, but got detached via syntactic movement. But note that I remain intentionally vague in saying only that it is a "similar" movement operation – because strictly speaking, the floated element would be the restrictor in the case of split PPCs, not the head (as in Q-floating structures). I will set this issue aside for the time being and briefly return to it in Section 4.

However, notably, an explanation along the lines of an analogy to Q-floating cannot be applied to Bulgarian split PPCs such as (21b) from above since Bulgarian does not have "true" Q-floating structures.[13] This is shown by the ungrammatical transformation of (23a) into Q-floated (23b). What might look like a Q-floating structure in (23c) at first glance should rather be considered as topicalization since the clitic would not occur in the non-topicalized structure.

(23) a. Vsički bademi   sa na masata.
          all   almonds are on table.DEF
          'All almonds are on the table.'

     b. * Bademite   sa vsički na masata.
          almonds.DEF are all   on table.DEF
          Intended: 'The almonds are all on the table.'

     c. Knigite   gi   pročetoh vsičkite.
          books.DEF them.CL read.1SG all.DEF
          'As for the books, I read them all.' (Intended: 'The books I read all.')

                              (Bulgarian; Vulchanova & Giusti 1995: 55)

---

[12]At this point, one could object that this is not an accurate description of Q-floating. In some approaches, for example, it is assumed that a quantifier is not a syntactic head in its own right, but merely occupies the specifier position of the restrictor NP/DP. Yet other accounts suggest that quantifiers might be simply adjoined to their restrictor arguments. To enter such a debate is beyond the scope of this article and must be postponed to another occasion. What I assume, though, is a kind of generalized structure for which the crucial point is this: there is a syntactic element (i.e. the quantifier or the plural pronoun) which selects, or at least combines with a restrictor argument. These two expressions can be separated or dislocated from each other via a syntactic movement operation. Note that this is also to say that I assume with Sportiche (1988), Shlonsky (1991), Merchant (1996), or Bošković (2004) Q-floating to involve syntactic movement. There, too, is an ongoing debate in the literature concerning the question whether this is indeed the case, or whether the quantifier acts more like an adverbial in such constructions (see for instance Bobaljik 2003), or whether both are lively options (see especially Fitzpatrick 2006 for a detailed cross-linguistic investigation).

[13]By "true", I mean Q-floating structures that do not make use of a resumptive pronoun or any other resumptive linguistic device.

In contrast, Q-floating structures exist in Torlakian BCMS, as illustrated in (24a-24b) – and moreover, there seems to be a remarkable correlation in general between the availability of iPPC readings for split PPCs and whether the respective language also permits Q-floating. Take Russian as a further example, where we find the very same pattern as in Torlakian BCMS: split PPCs can give rise to iPPC interpretations and Q-floating structures are possible; see (25a) and (25b)–(25c), respectively.[14]

(24) a. Svi bademi su   na sto.   (Torlakian BCMS)
          all almonds AUX.3PL on table
          'All almonds are on the table.'

      b. Bademi su   svi na sto.
          almonds AUX.3PL all on table
          'The almonds are all on the table.'

(25) a. My pojdëm   zavtra   s   Ivanom   v magazin.
          we go.FUT.1PL tomorrow with Ivan.INST to store
          'Tomorrow, we will go to the store with Ivan.'   ePPC
          'Tomorrow, Ivan and I will go to the store.'   iPPC

      b. Prišli   vse deti.
          came.PL all children
          'All the children came.'

      c. Deti   prišli   vse.
          children came.PL all
          'The children all came.'   (Russian; Fitzpatrick 2006: 144)

The parallel between split PPCs and Q-floating does not only account for the availability of iPPC readings for split PPCs, but also for the other piece of data that remained puzzling under Vassilieva & Larson's (2005) approach – namely the lack of an iPPC interpretation of *wh*-questions like (18b), repeated in (26) for convenience.

---

[14]One could also mention Polish in this context, which seems to pattern more or less with Bulgarian. The picture is, however, not entirely clear. While it has been stated in Feldman & Dyła (2008) that the comitative phrase has to occur adjacent to the plural pronoun in order to obtain iPPC readings, my informants found iPPC interpretations somewhat available for split PPCs. Likewise, structures with floated *wszystkie* 'all' were occasionally (and given certain intonational circumstances) found acceptable. After all, it thus seems that Polish occupies an intermediate position.

(26) S   kem   my xodili v magazin?
with whom we went to store
Unavailable iPPC interpretation: 'With whom did I go to the store?'

(Russian; Vassilieva & Larson 2005: 122)

In the terms of my account, a structure like (26) would have to involve the reverse kind of Q-floating movement operation; and carrying out such a movement does simply not yield grammatical results. We would thus expect it to be unavailable for QPs and plural pronominal DPs alike. Examples analogous to the Russian (26) do also not give rise to iPPC readings in Torlakian BCMS:

(27) S   koj ste   prekjuče   išli u prodavnicu?
with who AUX.2PL the.day.before.yesterday gone in shop
'With whom did you_PL go to the store the day before yesterday?'
Unavailable: 'With whom did you_SG go to the store the day before
yesterday?'   (Torlakian BCMS)

The same holds, unsurprisingly, for Bulgarian (28a). The only *wh*-question structure for which iPPC interpretations arise are echo questions like (28b), i.e. syntactic constructions in which the *wh*-element remains *in situ*.

(28) a. S   koj vie   ste   bili v muzeja?   (Bulgarian)
with who you.PL are.2PL been in museum.DEF
'With whom have you_PL been to the museum?'

b. Vie   s   KOJ otidokhte v muzeja?
you.PL with who went.2PL in museum.DEF
'You_SG went to the museum with WHOM?'

Such cases are covered by the theory presented in this article, since there is nothing in the structure of (28b) that would suggest that the comitative phrase does not occur inside of the plural pronoun's restrictor – rather, the comitative-internal DP just got replaced by a respective wh-phrase.

3.2.2 Subject Control

Further support for the view put forth here comes from Subject Control (henceforth: SC) constructions: Consider the contrast between (29a) and (29b) from Torlakian BCMS with regard to (im)possible interpretations to start with.

(29) a. *pro* pokušali smo   juče   s   Mariju   da   PRO
           tried   AUX.1PL yesterday with Maria.INST COMP
      popravljamo krov.
      repair   roof
      'Yesterday, we tried to repair the roof with Maria.'   ePPC
      'Yesterday, Maria and I tried to repair the roof.'   iPPC

   b. *pro* pokušali smo   juče   da   PRO popravljamo krov s
           tried   AUX.1PL yesterday COMP   repair   roof with
      Mariju.
      Maria.INST
      'Yesterday, we tried to repair the roof with Maria.'   ePPC
      Unavailable iPPC interpretation: 'Yesterday, Maria and I tried to repair
      the roof.'

                                                              (Torlakian BCMS)

If, as in (29a), the comitative phrase occurs in the matrix, we get the standard
ambiguity between an iPPC and an ePPC reading. But if, on the other hand, the
comitative phrase occupies a syntactic position inside of the embedded complex
popravljamo krov 'repair the roof', as in (29b), an iPPC interpretation is not avail-
able. Given the observations from the previous section, i.e. that split PPCs usually
also give rise to iPPC interpretations in Torlakian BCMS, the question arises why
this reading is unavailable for (29b).

It is commonly assumed that PRO is a null pronoun that lacks any phonological
content. Thus, PRO is treated as an empty category (see Chomsky 1981 for the
origins of this notion). Other instances falling into this class are traces of moved
phrases, including wh-movement traces, and (dropped) *pro*. But in contrast to *pro*,
for instance, there is no kind of overt NP or DP that corresponds to PRO.[15] PRO is
simply a silent anaphoric element that gets bound by an antecedent expression.
To the best of my knowledge, no specific assumptions have been made in the
literature regarding any sort of internal structure of PRO. Hence we should not go
all wrong if we presume that there is none, i.e. if we take PRO just to be a syntactic
subject placeholder element that has to occupy SpecTP in SC constructions due
to the EPP (Extended Projection Principle).

So since PRO has no internal structure, the comitative phrase in (29b) cannot
occur inside of it – because PRO has no such thing as what I dubbed a restrictor
here. Consequently, *s Mariju* can only be part of the plural pronoun's scope argu-
ment in (29b). Given that we are dealing with an SC construction, no movement

---

[15]Whereas for *pro*, the corresponding overt expression would be the pronoun with the respective
number and person features, of course.

of the plural pronoun beyond *da* can be argued to be involved. Or, put differently, there is just no instance whatsoever (neither overt nor covert) of the plural pronoun inside of (29b)'s PRO-part such that we could assume the comitative phrase to be a leftover of.[16] On the one hand, we cannot claim that the comitative phrase was base-generated within PRO (for the reasons just mentioned) and that the two got separated via a syntactic movement operation akin to Q-floating. On the other hand, we also cannot claim that the (dropped) plural pronoun does, or at any point of the derivation did, appear inside of the *da*-complement of the matrix. So in sum, my analysis correctly rules out iPPC interpretations for sentences like (29b), but correctly predicts the availability of an iPPC reading for sentences such as (29a), where the PPC also occurs split, but the (dropped) plural pronoun and the comitative phrase are inside the same clause – and that is all the more important: Because if the movement operation behind split PPCs indeed resembles Q-floating, this is another reason why (29b) has no iPPC interpretation, even under an account to SC such as Hornstein's (1999). Q-floating is clause-bound (cf. Kayne 1981), i.e. the moved constituent cannot cross certain syntactic boundaries such as (full) clauses. In this sense, the complementizer *da*, or as we could also say, PRO, constitutes such a crucial border: a PPC that appears (linearly speaking) behind it cannot, and cannot have been part of the plural pronoun's restrictor. Thus, only ePPC readings are available for structures such as (29b). But a PPC that occurs in the matrix, may it be split or not (as long as both parts reside inside the matrix), can give rise to iPPC interpretations.

The significance of these SC examples can easily be overlooked, and what they indicate can just as easily be misunderstood. So I recap here why those data are relevant and what they show. First of all, one should bear in mind that split PPCs give rise to iPPC readings in Torlakian BCMS – nevertheless, this interpretation is not available for (29b). And, naïvely speaking, there is no reason why this should be the case. Hence if iPPC readings arise for split PPCs in Torlakian BCMS, what is (29b)'s special trait that blocks this interpretation? Or, asked the other way around, what would have to be the case in order to obtain an iPPC interpretation for (29b)? Well, the comitative phrase (inside the *da*- or PRO-part

---

[16]It has been suggested in Hornstein (1999), however, that SC does indeed involve movement of the subject from the infinitival clause to a higher syntactic position in the matrix. That is, PRO, in Hornstein's (1999) sense, would rather be a trace than a silent anaphoric pronominal. Under such an analysis, the argument concerning (29b) just made above might seem a bit misguided. However, even if the movement-based approach to SC constructions was the more suitable solution (for a discussion, see Landau 2003), there is another issue standing in the way of *s Mariju* originating in the restrictor of the plural pronoun in (29b) – namely clause-boundedness of floated quantifiers. I return to this issue in more detail below.

of the sentence) would need to have been base-generated within the DP of the (dropped) plural pronoun that occurs in the matrix. Is it a feasible assumption that at any stage of the derivation, this was the case in (29b)? I argue here that it is not. If we stick to the more common treatment of SC constructions in terms of PRO, then neither has the plural pronominal DP ever occurred within the *da*-part of the sentence, nor can the comitative phrase occur inside the restrictor of PRO itself since there is no such thing. If we follow Hornstein (1999) instead in assuming that SC constructions actually do involve syntactic movement, then still, an iPPC interpretation of (29b) is ruled out under my account – but for a slightly different reason. Under this view, we would have to say that the comitative phrase can nonetheless not have been base-generated within the plural pronoun's DP because Q-floating is a clause-bound operation. And if split PPCs involve a similar kind of movement, *s Mariju* must have originated somewhere else in the structure but within the plural pronoun's restrictor.

Nevertheless, one could object that the decisive difference between (29b) and (29a) lies elsewhere in the structure anyway. Specifically, in the fact that PRO is an anaphoric element that needs a binder, i.e. an antecedent expression from which it gets its reference. Such a binder expression can only occur somewhere higher in the syntactic structure, that is, in the matrix part of the sentence. Now the comitative phrase occurs in (29a), but not in (29b) within the matrix. The consequence is that PRO can be bound in (29a), but not in (29b) by the PPC – may it have an underlying ePPC or an underlying iPPC structure in (29a). And then, there is nothing in the structure of (29b) that would even suggest that we are dealing with a split iPPC. I would like to make two comments on such a line of reasoning to conclude this section: First of all, if that was indeed the reason for the difference in terms of available interpretations between (29a) and (29b), it does no harm to the theory proposed here – *s Mariju* would end up being a part of the plural pronoun's scope argument anyway. Consequently, we correctly predict only an ePPC reading to be available for (29b). It is fair to admit, however, that other analyses of PPCs would probably make similar predictions.[17]

---

[17]For example, if an iPPC had an underlying coordinative structure, then one could argue that the comitative phrase cannot be coordinated with PRO; no matter whether we are dealing with an asymmetric or symmetric kind of coordination. The precise predictions apposition-based theories would make in this respect are somewhat more difficult to calculate. But we could at least state the following: Appositives occur quite freely within a sentence in general, they need not occur strictly adjacent to their anchor expressions. A sentence-final position as the comitative phrase in (29b) occupies, however, is one of the positions that is usually readily available for appositions. Since many aspects have not been spelled out in Cable's (2017) approach, it can only be surmised that this theory (as it currently stands) would probably not preclude an iPPC reading of (29b).

But second, it has not yet been once and for all decided whether SC constructions actually involve a PRO, or whether Hornstein's (1999) approach involving movement is more practicable in the end – although one must of course say that there are many aspects about SC in favour of the PRO view. Should the latter option nevertheless be the case, my analysis has a suitable explanation prepared. Lastly, we must not forget that some analyses of PPCs face difficulties in explaining iPPC readings of split PPCs at all.

3.2.3 Binding data

My analysis in its current form makes the following predictions regarding the binding of non-independently referential expressions such as *-self* anaphors. If the plural pronoun's restrictor is made up of a silent instance of reference to the speaker as well as of a comitative phrase including a comitative referent (i.e. in the iPPC case), we expect that the two together can co-bind an anaphor. If, on the other hand, the plural pronoun's restrictor is fully determined by SPKR and *pro$_i$* (i.e. in the ePPC case), we expect that the comitative referent cannot participate in the binding of an anaphor. I use exclusively Russian data in this section to show that these predictions are borne out. The reason for this change in the languages discussed is the following: unlike Torlakian BCMS, Russian has the possessive SELF anaphor *svoj-*. This element can be bound by antecedents of various morpho-syntactic kinds without changing its surface form. That is to say that *svoj-* can, for instance, be bound by a referential expression in the singular, or by a referential expression in the plural without having to adapt its own morpho-syntactic shape – it will remain as plain *svoj-* either way. In Torlakian BCMS, pronouns such as *naše* 'our' would have to be used instead in analogous examples as the ones from Russian below. The virtue of discussing examples with *svoj-* is thus that we are on the safer side in excluding cases of mere co-reference (instead of binding).

The (un)available referential properties of *svoj-* under an iPPC and an ePPC reading of the Russian sentence in (30) show that our predictions are correct. Although *svoj-* could also be interpreted as *his* (in the sense of *Petja's*) in principle, this binding configuration is not available under the iPPC reading of (30). Namely, the comitative referent from inside the plural pronoun's restrictor is not available as a sole binder of *svoj-* – and neither is the silent instance of reference to the speaker. Rather, all referential instances from the restrictor of the plural pronoun jointly bind the SELF element here.

(30) My   s   Petej   čitaem   svoju   knigu.
     we.NOM with Petja.INST read.1PL POSS.REFL book
                                    (Russian; Vassilieva & Larson 2005: 112)
   a. 'We$_j$ are reading SELF$_{j/*i/*j+i}$'s book with Petja.'   ePPC
   b. 'Petja$_i$ and I$_j$ are reading SELF$_{*j/*i/j+i}$'s book.'   iPPC

As can be seen from the respective paraphrase in (30a), the comitative referent
is not at all involved in binding *svoj-* in the ePPC case. We find the same pattern
in "regular" comitatives where the comitative phrase is commonly analyzed as
an adjunct to VP, see (31).

(31) Mal'čik   s   kotënkom ušël   v svoju   komnatu.
     boy.NOM with kitten.INST went.3SG to POSS.REFL room
     'The boy$_j$ went to SELF$_{j/*i/*j+i}$'s room with the kitten$_i$.'
                                    (Russian; Vassilieva & Larson 2005: 109)

Therefore the conclusion suggests that the comitative phrase adjoins to VP in
ePPC structures as well.

4 Conclusion

The analysis presented in this article is based on the assumption that plural pro-
nouns and quantifiers have some crucial semantic and syntactic properties in
common. In particular, I suggested that plural pronouns select restrictor argu-
ments just like quantifiers. Analogous to instances of universal quantification, a
plural pronoun conveys that for all elements in its restrictor denotation, it holds
that Q (where Q is some predicate). Under my approach, the restrictor of a plural
pronoun contains the following three ingredients: A silent instance of the plu-
ral pronoun's singular counterpart (reference to the speaker/addressee), a silent
pronominal element whose reference is determined by means of a contextual
assignment function $g(i)$ (reference to other(s)), and a silent operator OP$_\cup$. The
function of this operator is to form a set consisting of exactly those individuals to
which the two referential instances from the restrictor refer. Thus, OP$_\cup$ basically
has a mechanism amounting to set union.

  I argued that in an iPPC structure, the comitative phrase acts as the spell-out
of the more general OP$_\cup$ plus $g(i)$ part which is present in the restrictor of a plural
pronoun anyway by default. Moreover, I claimed that this was the case since the
comitative element in these constructions has the very same semantics as OP$_\cup$
– that is to say that in an iPPC the comitative phrase occurs inside the plural

pronominal restrictor, i.e. within the same DP.[18] I suggested that we can account for the availability of iPPC interpretations of split PPCs by assuming that these constructions involve a kind of syntactic movement similar to the kind involved in Quantifier-floating structures. And indeed, we found an intriguing correlation: the very Slavic languages that allow iPPC readings of split PPCs also have Quantifier-floating structures. However, as already pointed out in Section 3.2.1, the wrong element is "floated" under my approach actually. That is, if we really want to assume that split PPCs are related to Q-floating structures, then we are faced with the question of why the restrictor gets floated in a split PPC, whereas the quantifier does in a Q-floating structure. I am aware of this problem, and of the fact that the analogy between these two cannot be 1:1 – but for the sake of explicitness, I stick with it for the time being; and leave it as an issue for further research which kind of syntactic movement could be involved in split PPCs instead, if any. An obvious candidate is so-called "Left Branch Extraction" (LBE; cf. Bošković 2008, and the references therein). But whether the generalizations made in relation to LBE in Slavic languages, and the generalizations we could derive in this article are actually compatible with each other remains to be seen. And in any case, it seems remarkable that the positions available for floated quantifiers within a sentence coincide with the positions available for the comitative phrase in a split PPC constructions – at least in Torlakian BCMS.[19]

Other issues remained open here as well. For example, it is puzzling why an iPPC reading in Torlakian BCMS is much more salient when the plural pronoun is dropped, and *vice versa* an ePPC interpretation is much more salient when the pronoun is pronounced; see (32).

(32)   a. Mi   smo   s   Mariju   otišli u muzej.
         we.NOM AUX.1PL with Maria.INST went in museum
         'We went to the museum with Maria.'   (more salient) ePPC
         'Maria and I went to the museum.'   (less prominent) iPPC

---

[18]As an anonymous reviewer mentions, this raises the question why (i)PPCs are not more widespread across the world's languages. But one of the crucial points of my analysis is indeed that the comitative element *s* and OP$_\cup$ have the same semantics – and there is no salient reason to assume that a particularly large number of languages should have a comitative element whose semantics matches those of OP$_\cup$. Moreover, other languages have found different ways of expressing the same meaning as an iPPC. Just consider Icelandic which has so-called Pro[NP] constructions (cf. Sigurðsson & Wood 2020). Those Pro[NP]s superficially differ from PPCs only in that they lack a comitative element.

[19]But to make Torlakian BCMS not to appear very exotic in this respect: The same correspondence of available positions can also be found in other languages (belonging to different language families) that have both Quantifier-floating and (split) PPCs – namely Finnish and Fenno-Swedish; see also Butschety (2023).

b. *pro* s   Mariju   smo   otišli u muzej.
   with Maria.INST AUX.1PL went in museum
  'We went to the museum with Maria.'   (less prominent) ePPC
  'Maria and I went to the museum.'   (more salient) iPPC
                                    (Torlakian BCMS)

I can only speculate here that this is a pragmatic effect. More precisely, one could assume that the pronunciation of the plural pronoun (which could otherwise be dropped) indicates that we are dealing with a fully determined instance of the plural pronoun and that the comitative phrase consequently does not occur inside its restrictor.

Moreover, it remains unclear why predicates such as *hate broccoli* or *believe in God* are felicitous with PPCs under an iPPC interpretation (but crucially not under an ePPC interpretation) in Russian (see (33a) and (33b)), but not in Torlakian BCMS, where analogous sentences are simply infelicitous no matter which reading is intended.

(33) a. My   s   Dašei   verim   v boga.
      we.NOM with Daša.INST believe.1PL in God
      Unavailable ePPC reading: 'We believe in God with Daša.'
      'Daša and I believe in God.'   iPPC
                                (Russian; Feldman 2003)

    b. My   s   Ivanom   nenavidim brokkoli.
      we.NOM with Ivan.INST hate.1PL   broccoli
      Unavailable ePPC reading: 'We hate broccoli with Ivan.'
      'Ivan and I hate broccoli.'   iPPC
                     (Russian; Vassilieva & Larson 2005: 112)

These topics need a much longer discussion, which I have to defer to another occasion.

To conclude, based on the observed analogies between plural pronouns and (universal) quantifiers, the two readings PPCs in Slavic can give rise to were derived in terms of whether the comitative phrase occurs inside the plural pronominal restrictor (=iPPC), or outside of it as an adjunct to VP (=ePPC). Assuming that split PPCs involve a similar kind of syntactic movement as floated quantifiers, my analysis could derive in which Slavic languages (namely precisely in those that have Q-floating structures) on the one hand, and in which syntactic configurations on the other hand iPPC readings arise for split PPCs. While this analogy turned out to be not as direct as intended, the correlation between the availability of floated (universal) quantifiers and iPPC readings for split PPCs nonetheless

represents a noteworthy datum. I leave it to future research to decide on whether and how these two phenomena can be put into a uniform picture.

Abbreviations

1	first person	F	feminine
2	second person	FUT	future tense
3	third person	INST	instrumental
ADDR	addressee	NOM	nominative
AUX	auxiliary	PL	plural
CL	clitic	POSS.REFL	possessive reflexive
COMP	complementizer	SPKR	speaker
DEF	definite	SG	singular

Acknowledgments

References

Bobaljik, Jonathan. 2003. Floating quantifiers: Handle with care. In Lisa Cheng & Rint Sybesma (eds.), *The second Glot International state-of-the-article book*, 107–148. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110890952.107.

Bošković, Željko. 2004. Be careful where you float your quantifiers. *Natural Language & Linguistic Theory* 22. 681–742. DOI: 10.1007/s11049-004-2541-z.

Bošković, Željko. 2008. What will you have, DP or NP? In Emily Elfner & Martin Walkow (eds.), *NELS 37: Proceedings of the 37th annual meeting of the North East Linguistic Society*, 101–114. Amherst: Booksurge Publishing.

Büring, Daniel. 2011. Pronouns. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, 971–996. Berlin: De Gruyter Mouton.

Butschety, Madeleine. 2023. *Inclusory constructions*. Graz: University of Graz. (Doctoral dissertation).

Cable, Seth. 2017. Some observations on the plural pronoun construction of Tlingit, Polish, and Russian. In Claire Halpert, Hadas Kotek & Coppe van Urk (eds.), *A pesky set: Papers for David Pesetsky*, 266–290. Cambridge: MIT Press.

Chomsky, Noam. 1981. *Lectures on government and binding.* Dordrecht: Foris. DOI: 10.1515/9783110884166.

Daniel, Mikhail. 2000. *Tipologija associativnoj množestvennosti.* Moscow: Russian State University for Humanities. (Doctoral dissertation).

Dyła, Stefan. 1988. Quasi-comitative coordination in Polish. *Linguistics* 26. 383–414. DOI: 10.1515/ling.1988.26.3.383.

Elbourne, Paul D. 2008. The interpretation of pronouns. *Language and Linguistics Compass* 2(1). 119–150. DOI: 10.1111/j.1749-818x.2007.00041.x.

Feldman, Anna. 2003. Comitative and plural pronoun constructions. In Yehuda N. Falk (ed.), *Proceedings of the 17th annual meeting of the Israel Association of Theoretical Linguistics.* https://www.iatl.org.il/?page_id=155.

Feldman, Anna & Stefan Dyła. 2008. On comitative constructions in Polish and Russian. In Lew Zybatow, Gerhild Zybatow, Luka Szucsich & Roland Meyer (eds.), *Formal description of Slavic languages: The fifth conference, Leipzig 2003.* Frankfurt am Main: Peter Lang.

Fitzpatrick, Justin M. 2006. *The syntactic and semantic roots of floating quantification.* Cambridge: MIT. (Doctoral dissertation). https://dspace.mit.edu/handle/1721.1/37420.

Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar.* Malden: Blackwell.

Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30(1). 69–96. https://www.jstor.org/stable/4179050.

Kayne, Richard. 1981. Binding, quantifiers, clitics and control. In Frank Heny (ed.), *Binding and filtering*, 191–211. Cambridge: MIT Press. DOI: 10.1515/9783111682228-005.

Ladusaw, William A. 1989. Group reference and the plural pronoun construction. In *Papers on the plural pronoun construction and comitative coordination*, 1–7. Santa Cruz: UCSC Syntax Research Center.

Landau, Idan. 2003. Movement out of control. *Linguistic Inquiry* 34(3). 471–498. http://www.jstor.org/stable/4179243.

Larson, Richard K. & Gabriel Segal. 1995. *Knowledge of meaning: An introduction to semantic theory.* Cambridge: MIT Press. DOI: 10.7551/mitpress/4076.001.0001.

Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretic approach. In Paul Portner & Barbara H. Partee (eds.), *Formal se-*

mantics: The essential readings, 127–147. Malden: Blackwell. DOI: 10 . 1515 / 9783110852820.302.

Lipták, Anikó. 2015. Identity in ellipsis: An introduction. *Lingua* 166(B). 155–171. DOI: 10.1016/j.lingua.2015.08.003.

McNally, Louise. 1993. Comitative coordination: A case study in group formation. *Natural Language & Linguistic Theory* 11. 347–379. DOI: 10.1007/BF00992917.

Merchant, Jason. 1996. Object scrambling and quantifier float in German. In Kiyomi Kusumoto (ed.), *NELS 26: Proceedings of the North East Linguistic Society*, 179–193. Amherst: GLSA.

Nunberg, Geoffrey. 1993. Indexicality and deixis. *Linguistics and Philosophy* 16(1). 1–43. https://www.jstor.org/stable/25001498.

Shlonsky, Ur. 1991. Quantifiers as functional heads: A study of quantifier float in Hebrew. *Lingua* 84(2–3). 159–180. DOI: 10.1016/0024-3841(91)90069-H.

Sigurdsson, Halldór Ármann & Jim Wood. 2020. *We Olaf*: Pro[(x-)NP] constructions in Icelandic and beyond. *Glossa: a journal of general linguistics* 5. 1–16. DOI: 10.5334/gjgl.1054.

Sokolová, Jana. 2019. On comitative constructions in Slovak. *Skase Journal of Theoretical Linguistics* 16(3). 90–107.

Sportiche, Dominique. 1988. A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry* 84(3). 425–449. https://www.jstor.org/stable/25164903.

Stokke, Andreas. 2022. Features of referential pronouns and indexical presuppositions. *Inquiry: An Interdisciplinary Journal of Philosophy* 65(8). 1083–1115. DOI: 10.1080/0020174x.2020.1801500.

Szabolcsi, Anna. 2010. *Quantification* (Research Surveys in Linguistics). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511781681.

Trawinski, Beata. 2005. Plural comitative constructions in Polish. In Stefan Müller (ed.), *Proceedings of the 12th international conference on Head-Driven Phrase Structure Grammar*, 375–395. Stanford: CSLI Publications.

Vassilieva, Maria B. & Richard K. Larson. 2001. The semantics of the plural pronoun construction. In Rachel Hastings, Brendan Jackson & Zsofia Zvolenszky (eds.), *Proceedings of SALT 11*, 449–465. Ithaca: CLC Publications. https : / / journals . linguisticsociety . org / proceedings / index . php / SALT / article / view / 2853/2593.

Vassilieva, Maria B. & Richard K. Larson. 2005. The semantics of the plural pronoun construction. *Natural Language Semantics* 13. 101–124. DOI: 10 . 1007 / s11050-004-1031-5.

Vulchanova, Mila & Guiliana Giusti. 1995. Quantified noun phrase structure in Bulgarian. *University of Venice Working Papers in Linguistics* 5(1). 43–65.

Chapter 7

Equatives and two theories of negative concord

Mojmír Dočekal

Masaryk University

This article reports the results of an experiment targeting the acceptability of Czech neg-words and strong NPIs under Neg-Raising predicates and in the complement clauses of equatives. The theoretical consequences of the results are discussed and range from the support of non-standard negative concord theories to the support of non-standard degree semantics for the equative constructions.

1 Introduction

In this article, I explore expressions that are polarity-dependent. The evidence comes from Czech, a strict negative-concord language. I will focus on one recent experiment in the paper.[1] The theoretical ambition of this paper is to examine the distribution of neg-words and strong NPIs in two environments: Neg-Raising predicates and equatives. The acceptability pattern of these two kinds of negative dependent expressions is challenging for standard theories of neg-words but also for the current degree theories of equatives. The data are subtle, and therefore I report results of an acceptability judgment task experiment on Czech native speakers; in this way, I add to the experimental research on Negative Polarity Items (NPIs), like Chemla et al. (2011), Gajewski (2016), Alexandropoulou et al. (2020), a.o., more specifically to the experimental research on cross-linguistic variation in NPI licensing as found in Djärv et al. (2018), Schwarz et al. (2020), Chierchia (2019). Empirically, the experimental data concern the Czech strong

---

[1]But the experiment is a continuation of many previous experimental works which incrementally changed the nature of questions and research goals reflected in the current experiment.

NPIs, like *ani jeden* 'even one,' and neg-words, like *žádný* 'no', as exemplified in (1).

(1) Petr nepotkal {ani jednoho / žádného}   studenta.
    Petr NEG.met   STRONG.NPI   NEG.WORD student
    'Petr didn't meet {even one / any} student.'

Both polarity-sensitive items are in the majority of contexts interchangeable, but their meaning differs, and the experiment was focused on the environments where the meaning difference is detectable. In previous works, the acceptability of NPIs was reported as varying between speakers (Homer 2021, Mayer et al. 2018). In the neg-words research, the variation of acceptability was found to correlate with demographic factors such as age or education (Burnett et al. 2015, 2018). For these reasons, I also included demographic variables in my experimental research to see whether a more complex picture (that incorporates both grammatical and demographic factors) can explain the distribution of neg-words and strong NPIs more successfully.

The article is structured as follows: In Section 1.1, I introduce the theoretical background for the experiment. In Section 2, I describe the experiment. In Section 3, I discuss the results, and in Section 4, I conclude the article.

1.1 Theoretical background

1.1.1 Polarity-dependent expressions in equatives and under Neg-Raising predicates

The theoretical background for the experiment is the contrast between strong NPIs and neg-words in two environments, under Neg-Raising predicates and in equatives. As a baseline against the two environments, I used simple unembedded sentences with negated verbs. The baseline is important since it is the only environment where both strong NPIs and neg-words are grammatical under any theory of neg-words and NPIs. The baseline is illustrated in (1). In the current section I discuss the theoretical background for the constructions (equatives and Neg-Raising predicates), in Section 1.1.2 Czech strong NPIs and neg-words are discussed, and Section 1.1.3 is dedicated to the theoretical background for the licensing of strong NPIs and neg-words.

As for Neg-Raising, we can adapt any current theory of Neg-Raising, be it the presuppositional version of Gajewski (2007) or the scalar implicature version of Romoli (2013). Both share the insight that Neg-Raising predicates bear the excluded middle inference: for *believe*: $Bel(p) \lor Bel(\neg p)$, adding the negated

assertion ¬*Bel*(*p*) results in the deductively valid conclusion where the negation scopes in the embedded clause, *Bel*(¬*p*). Non-Neg-Raising predicates then come without the excluded middle inference, resulting in the surface interpretation of the negation. Since the scope of negation ends in the embedded clause, the negation is local, and therefore, the strong NPIs are licensed, as demonstrated with Spanish *ni un* 'not even one' in (2).

(2)   No creo   que ni un   solo soldado pueda lograrlo.
       not believe.1SG that not even one soldier can   achieve
       'I don't believe that not even one soldier can achieve it.'

The second environment which was tested in the experiment was equatives. The first thing to note is that the standard theory of equatives is built on the ">" analysis of comparatives (Beck 2019, von Stechow 1984) where the core operation is the relation > comparing two maxima: (i) the maximum of the set of degrees from the main clause, (ii) the maximum of the set of degrees from the complement of the comparative clause; see (3) as an illustration.

(3)   The dog is taller than the cat.
       a.   MAX(*d*|the height of the dog ≥ *d*) > MAX(*d*|the height of the cat ≥ *d*)

The standard theory of equatives (Beck 2019, von Stechow 1984, Rullmann 1995) then follows the ">" analysis of comparatives, just replacing > with ≥ which is in most contexts pragmatically strengthened to "="; see (4) for an illustration.

(4)   The dog is as tall as the cat.
       a.   MAX(*d*|the height of the dog ≥ *d*) ≥ MAX(*d*|the height of the cat ≥ *d*)

Comparatives are then theoretically expected to license NPIs in their complement clauses, since the degree argument is downward-monotonic, therefore if some degree $d > d'$ and there is another degree d'', such as $d' > d''$, then by transitivity $d > d''$. Intuitively, if the dog from (3) is taller (or of the same height) than the cat from (3), then he is taller than any cat smaller than that cat. The literature on comparatives (von Stechow 1984, Rullmann 1995, Gajewski 2008) agrees on the empirical verification of this prediction. Weak NPIs (like English *any*) are licensed in the complement clauses of the comparative. In the case of strong NPIs, the empirical situation is less clear. Still, at least empirically, it is claimed for Germanic languages that strong NPIs appear in the complement clause of comparatives felicitously; see Hoeksema's (2012) Dutch example in (5). (5) contains the Dutch expression *ook maar* 'even,' which is taken as a standard example of a strong NPI (see Zwarts 1998).

(5) Zij was beter dan ook maar iemand verwacht had.
    she AUX better than STRONG.NPI   expected AUX
    'She was better than anyone could have expected.'

Under the premise that equatives are built on the ">" analysis of comparatives, the standard theory of equatives predicts that NPIs should be licensed in the complement clause of equatives. This prediction works for English and supports the standard theories of equatives (see von Stechow 1984, Beck 2019, a.o.), since NPIs in English equatives are licensed; see (6) from Seuren (1984).

(6) Paris is as quiet as ever.

The theories of Neg-Raising and equatives are general and their aim is to model the meaning of the construction. The interaction of the constructions with various classes of polarity dependent expressions is not their primary concern. Nevertheless, as discussed in this section at least for strong NPIs, the predictions of the standard theories of Neg-Raising and standard theories of equatives are clear. The strong NPIs should be licensed. But if we want to apply the predictions of the standard theories of Neg-Raising and equatives to both strong NPIs and neg-words, we have to introduce the theories of neg-words and strong NPIs. This is the topic of the next section.

1.1.2 Czech strong NPIs and neg-words

Let us introduce some background information and intuitions concerning both classes. Starting with strong NPIs (for a theoretical framework, see Gajewski 2011), Czech strong NPIs of the *ani* sort bear the unlikelihood presupposition, discussed concerning English stressed *ANY* (see Krifka 1995, a.o.), Hindi *ek bhii* (see Lahiri 1998, a.o.) or English *even one* (see Crnič 2014b, a.o.). But unlike the English or Hindi strong NPIs, the Czech strong NPIs are much more limited in distribution, requiring clause-mate negation in most of their occurrences. Nevertheless, this requirement is not obligatory, as will be demonstrated, and Czech *ani* strong NPIs can appear embedded under negated Neg-Raising predicates without any overt clause-mate negation. But in all contexts, *ani* presupposes that its prejacent (a proposition which *ani* modifies) entails all the relevant alternatives. By way of example, *ani jeden* 'even one' in (7) is acceptable since not scoring one goal entails not scoring two, three, etc. goals ($\neg \text{SCORE}(1) \vDash \neg \text{SCORE}(2 : \infty)$), the relevant alternatives. But *ani deset* 'even ten' is much less acceptable, since not scoring ten goals is entailed by not scoring 9, 8, etc. goals ($\neg \text{SCORE}(1 : 9) \vDash \neg \text{SCORE}(10)$) and therefore the prejacent does not entail all the relevant alternatives. In this

respect, *ani jeden* belongs to the same class of strong NPIs as English *even one*, which yields the scalar presupposition (in (7) the focus alternatives have to be less probable and entailed by the prejacent).

(7) FC Barcelona nedala   {ani jeden / #ani   deset} gól/ů.
     FC Barcelona NEG.gave even one   even ten   goal(s)
     'FC Barcelona didn't score {even one/#ten} goal(s).'

Turning now to neg-words, Czech (and generally Slavic) neg-words are similar to Italian neg-words (such as *niente*, e.g., see Ladusaw 1992). In contrast to strong NPIs like *ani jeden* in (7), or English *even one*, neg-words do not bear any scalar or additive presupposition. In addition, neg-words have strong syntactic requirements on their licensing, and in Czech, as in all Slavic languages, which are strict negative-concord languages (see Zeijlstra 2004, a.o.), Czech neg-words in the majority of contexts require verbal negation (in the same clause), the requirements being more strict than in the case of Czech strong NPIs; see (8).

(8) a. Petr nedal   žádný   gól.
        Petr NEG.scored NEG.word goal
        'Petr didn't score any goal.'

     b. Nikdo   {nepřišel   / #přišel}.
        NEG.WORD NEG.came   came
        'Nobody came.'

     c. *Petr neřekl, že nikdo   přišel.
        Petr NEG.said that NEG.WORD came
        'Petr didn't say that anybody came.'

Unlike strong NPIs, neg-words do not yield the scalar presupposition but their licensing is more locality-constrained. Therefore Czech neg-words are degraded under negated Neg-Raising predicates (see Dočekal & Dotlačil 2016a,b for details, and (23b) for an example from the experiment).

   The most influential current analysis of neg-words is the syntactic approach of Zeijlstra (2004), a.o. (the standard theory/Zeijlstra 2004 hereinafter). It claims for strict negative-concord languages that all neg-words (and the verbal negation) carry a [uNeg] feature and are checked against an [iNeg] (covert) operator with the semantics of ¬. Part of this paper is dedicated to providing experimental support for an alternative semantic theory of neg-words (see Ovalle & Guerzoni 2004, Kuhn 2022, a.o.; I will refer to this theory as the alternative theory/Ovalle & Guerzoni 2004), which will be explained in detail later; see Section 1.1.3. The

empirical point concerns equatives, one of the contexts where the distribution of strong NPIs and neg-words diverges. It was noted for Polish equatives at least as early as Błaszczak (2001) that neg-words are surprisingly grammatical in them. Czech equatives are similar; they seem not to license strong (and weak) NPIs, resembling German and many other non-English equatives (see Krifka 1992, a.o.). Nevertheless, neg-words are very much acceptable in the complement clauses of Czech equatives; see (9).

(9) Petr je tak vysoký jako {#ani jeden / žádný} jiný student.
Petr is so tall how STRONG.NPI NEG.WORD other student.
'Petr is as tall as any other student.'

The acceptability of neg-words in Czech equatives is also surprising according to the standard theory of neg-words/Zeijlstra (2004), since there is no plausible overt or covert operator with the interpretable [iNeg] feature in the complement clause of Czech equatives. The standard theory of neg-words/Zeijlstra (2004) predicts the ungrammaticality of neg-words in Czech equatives, which is empirically wrong. Part of the experimental work reported in this paper is to test the acceptability of neg-words in Czech equatives, such as (9), and to compare it with the acceptability of strong NPIs.

Concerning Neg-Raising predicates, Czech (like Spanish in (2)) allows licensing of strong NPIs in the embedded clause; see (10).

(10) Nechci, aby ani jeden student odešel.
NEG.want.1SG that even one student left
'I don't want even one student to leave.'

As for neg-words, previous experimental research reported their decreased acceptability under Neg-Raising predicates (see Dočekal & Dotlačil 2016a). Such a pattern is expected in the standard theory of neg-words/Zeijlstra (2004), since the negation is syntactically localized in the root clause. In terms of the syntactic approach, the root negation bearing a [iNeg] feature is too far away from eventual neg-words in the embedded clause to license them.

To this end, the experiment reported below scrutinizes the contrast between strong NPIs and neg-words in equatives (and under Neg-Raising predicates). First, Slavic literature observed the acceptability of neg-words in equatives, and second, NPI literature noticed the unacceptability of strong NPIs in Germanic. Nevertheless, the contrast was neither experimentally researched nor theoretically explained. Moreover, equatives are one of the environments where the contrast between Czech neg-words and strong NPIs is most robust, but still, there

seems to be speaker variation involved. In simple terms, some speakers treat *ani* as a neg-word and therefore do not accept it as much in equatives, unlike the speakers who use *ani* as a strong NPI. In a bit broader picture, the speaker variation resembles the variation of English NPIs vs. negative quantifiers, e.g., as studied first by functional linguists (see Tottie 1991, a.o.), and more recently in the formal syntactic tradition (see Burnett et al. 2015, 2018, a.o.). Burnett et al. (2015, 2018) show that formal constraints explain the English speaker variation with higher success than historical and social factors (discovered in the functional tradition before). According to Burnett et al. (2018), the English negative quantifiers are replaced by NPIs in lower syntactic domains.[2] This process overrules any demographic factors, like age or education. In a similar vein, Burnett et al. (2015) describe the variable negative concord in Québec French as explainable by the interplay of grammatical and demographic factors, where the first type of factors is decisive.

The speaker variation mentioned above is an intriguing and hard-to-pin-down phenomenon, and one of the reasons for using experimental methods, since it certainly resists any simple intuition-based methods for data collecting. The emerging picture is that speaker variation concerning negation, negative concord, negative quantifiers, and NPIs comes both from social and grammatical sources, and only experimental work can give some reasonable answers as to their respective strength. In this respect, the experimental work reported below is the first tiny step in explaining Slavic neg-words vs. NPIs speaker variation due to the possible interplay between demographic and grammatical factors.

The following section introduces the licensing conditions for strong NPIs and neg-words in a formal way. At the end of the section the predictions of the theories of neg-words, strong NPIs, Neg-Raising and equatives are summarized.

1.1.3 Assumptions concerning licensing of (strong) NPIs

Let us assume a standard approach to NPIs and strong NPIs licensing. For the general framework, the so-called *even*-theory of NPIs licensing is naturally the most attractive candidate (see Krifka 1995, Lahiri 1998, Crnič 2014b, a.o.), since *ani* bears the unlikelihood presupposition similar to English *even*. And for strong NPIs, let us follow Gajewski's formalization of strong NPIs (Gajewski 2011). According to Gajewski (2011), strong NPIs are licensed in downward-entailing (DE)

---

[2]Consider a contrast like *There were no jobs to be had* – higher syntactic domain vs. *I can't have any form of gluten*, where in the first sentence, the negative quantifier is used, while in the second sentence, an NPI occurs. Examples come from Burnett et al. (2018), where it is claimed that while this constraint is soft in contemporary English, it is a hard one in the Scandinavian language family.

environments. But the downward entailments are checked both in Truth Conditions (TC), the at-issue part of the meaning, and in the non-at-issue meaning, presuppositions and implicatures being the most pertinent non-at-issue meaning components. Weak NPIs, on the other hand, require DE environments only in the TC part of the meaning. The conditions for weak and strong NPIs are summarized in (11).

(11) An NPI is licensed in the environment γ
      $[_\alpha exh[_\beta ... [_\gamma \text{NPI}] ...]]$:
      a. the environment γ is DE in β   weak NPIs
      b. the environment γ is DE in α   strong NPIs

The standard exhaustifier from (11) is the formalization of the *only*-kind of focus operator which works very well for weak NPIs like unstressed English *any*. But for other weak or strong NPIs with the unlikelihood-presupposition meaning, another kind of exhaustifier, a covert counterpart of English *even*, was proposed (see Crnič 2011, 2014a). The same mechanism is used in formal approaches to focus particles (see Panizza & Sudo 2020). The *even* exhaustifier, like its overt version, then comes with two presuppositions. The first is scalar, demonstrated in (12a) – the sentence is acceptable in such contexts where a dancing Pope is very unlikely (compatible with the actual world). The second is additive, exemplified with (12b).

(12) a. Even the Pope$_F$ danced.
      b. Even one$_F$ cat will make the Pope happy.

The sentence is true if two, three, …cats will make the Pope happy as well. The placement of focus determines the nature of alternatives used in presuppositions. Let us follow the formalization of both presuppositions by Panizza & Sudo (2020); see (13).

(13) 'Even ϕ' presupposes:
      a. that ϕ is relatively unlikely to be true among Alt(ϕ); and
      b. that there is $\psi \in$ Alt(ϕ) that is not entailed by ϕ and is true.

For monotonic scales, likelihood from (13) translates into entailment (after Crnič 2011), therefore the predictions of traditional downward-entailing approaches like (Ladusaw 1992) and *even*-theories of NPIs collapse for downward-monotonic contexts.

Assuming this standard approach to strong NPIs, its predictions are clear for simple negated sentences like (14), where *even* associates with the weak scalar item (the numeral *one*). The scalar presupposition of *even* has to scope over negation, schematically [$_{even}$¬[one student arrived]], and since this logical form entails all other alternatives ([$_{even}$¬[*n* students arrived]], where $n > 1$), the prejacent is both strongest and the least likely from the alternatives and the scalar presupposition (see (13)) is fulfilled. In the experiment, I used baseline sentences of similar form for both strong NPIs and neg-words. Both expressions were unsurprisingly well-accepted in the baseline.

(14)   Even one student didn't arrive.

For Neg-Raising predicates like *want* from (15), then the schematic scope configuration in the embedded clause is covert (*even*) > ¬ > [$_\alpha$... one ...] – the scalar presupposition of *even* is fulfilled. Moreover, the licensing condition for strong NPIs (see (11)) requires that the local domain (α in (15)) is DE, meaning after we factor in all non-at-issue meaning components, which is the case for Neg-Raising predicates.

(15)   The director doesn't want [$_\alpha$ even one student to depart].

As for equatives, they are theoretically expected to license NPIs, which seems to be the case for English weak NPIs, as illustrated by (6), repeated for convenience below as (16).

(16)   Paris is as quiet as ever.

But it was noticed before that this does not hold cross-linguistically; see Krifka (1992) for German and Penka (2016) for German and Romance languages. But at least in the comparative/equative ">" theories, if comparatives license strong NPIs, the expectation is that equatives will behave similarly. Turning now to Slavic equatives, there are many factors at play here, though. First, Slavic equatives are different from English equatives, and their morpho-syntax is very similar to correlatives (like German and Romance equatives). And since it is known at least from Jacobson (1995) that correlatives are bad licensors of NPIs, the expectation is that both weak and strong NPIs will be much worse in Slavic equatives (compared to Germanic languages).[3] To summarize, the standard degree theory

---

[3]To address this issue, another experiment targeting both weak and strong NPIs in comparatives and equatives is in preparation.

of equatives – the standard degree theory/Rullmann (1995) hereinafter (and assuming the standard theory of NPIs licensing introduced above) – predicts that NPIs (weak and strong) should be licensed in equatives. I was unable to locate any scholarly discourse pertaining to strong NPIs and equatives, but the following example (17) from Chisholm (2010) can be seen as an empirical approval of the standard degree theory prediction for English. In (17), there is a strong NPI *until recently* in the equative standard; the verb *changed* is telic. Therefore the strong NPI should be licensed by the DE logical properties of the English equative, which seems to be the case. But as will be demonstrated in the next section, the situation is quite different in Czech, which confirms the observations concerning German and Romance equatives (Krifka 1992, Penka 2016) and in alternative theories of equatives (the alternative degree theories/Penka 2016 hereinafter).

(17)   Under the party system in Canada cabinets changed as often as, until recently, they did in France.

1.1.4 Assumptions concerning licensing of neg-words

As for the licensing of neg-words, I will now introduce the syntactic theory of Negative Concord (NC) developed by Zeijlstra (2004), Penka (2007), and Zeijlstra (2022) in detail. The standard theory/Zeijlstra (2004) is the syntactic tool for dealing with negative concord both in strict and non-strict negative concord languages. Since Czech (like all Slavic languages) is an example of strict negative concord, I will focus on the part of the theory that deals with strict NC. The basic assumption for strict NC languages in Penka/Zeijlstra's syntactic theory is that all morphologically-negated words come without semantic negation. Neg-words and sentential negation carry a so-called uninterpretable [uNeg] feature, which is in agreement with the logical operator (propositional negation) that has an interpretable [iNeg] feature. Sentential negation is a signal of propositional negation, but propositional negation is located higher in the syntactic tree than sentential negation. The syntactic theory then treats neg-words as indefinites, and their negation is purely syntactical (the uninterpretable feature). The purpose of the uninterpretable feature is then to signal the presence of the propositional negation operator. Let us illustrate the mechanism used in the syntactic theory with a Czech example in (18a). The sentence contains three morphological negations, but according to the syntactic theory, none of them bears semantic force, which is delegated to the abstract logical operator with the semantics of classical propositional negation, see (18a-i). The final logical form is in (18b).

(18) a. Nikdo   neviděl nic.
          NEG.person NEG.saw NEG.thing
          'Nobody saw anything.'

  i. Op$_{\neg\text{[iNeg]}}$[Nikdo$_{\neg\text{[uNeg]}}$ neviděl$_{\neg\text{[uNeg]}}$ nic$_{\neg\text{[uNeg]}}$.]

  ii. ⟦nikdo⟧ $= \lambda P \exists x[\text{PERSON}(x) \wedge P(x)]$

  iii. ⟦nic⟧ $= \lambda P \exists x[\text{THING}(x) \wedge P(x)]$

  iv. ⟦neviděl⟧ $= \lambda y.\lambda x.\text{SEE}(x, y)$

  b. $\neg \exists x \exists y[\text{PERSON}(x) \wedge \text{THING}(y) \wedge \text{SEE}(x, y)]$

The syntactic theory is well-equipped to deal with the locality constraints on negative concord and, of course, easily explains the baseline kind of example like (18a), where the neg-words and verbal negation appear in a root clause. The logical operator has to be local, around the level of the TP projection (of the clause where neg-words or verbal negation appears).

As for Neg-Raising, the predictions of the syntactic theory are the following: Since the inferential process by which the scope of negation ends on the embedded predicate (schematically: ¬ NegRaisingVerb [Predicate] ⤳ NegRaisingVerb [¬ Predicate]) is pragmatic in nature, the excluded middle inference is (depending on theory) treated either as a presupposition or as an implicature. The valid scope of the invisible operator is the root sentence (Op$_{\neg}$[NegRaisingVerb [Predicate]]). Therefore, neg-words in the embedded sentence are too far away for agreement between the uninterpretable feature and the interpretable feature of the operator.

Nevertheless, in the case of equatives, the standard theory/Zeijlstra (2004) simply predicts the ungrammaticality of neg-words (in the case of the positive main predicate), which is empirically wrong. Let us start with some empirical observations. According to Sketch Engine (Kilgarriff et al. 2014), in their csTenTen19 (the most representative Czech corpus in Sketch Engine), there are 28 occurrences of neg-words in the standard clause of equatives.[4] One example sentence from the query is in (19). This is in contrast to strong NPI: in Sketch Engine, there is no occurrence of strong NPI *ani* in the standard clause of equatives. This asymmetry is also verified by the intuitions of native speakers, as will be reported in the experiment. The empirical inadequacy of the syntactic theory follows from the standard theory of equatives as the $\geq$ relation between two maxima of two sets of degrees – there is no place for negation in the semantics of equatives, neither in the standard theory nor in the alternative theories of equatives (see Penka

---

[4]The CQL used for the search was: `[lemma="tak"]` `[tag="k2.*"]` `[lemma="jak[o]?"]` `[lemma="žádný"]`.

2016, a.o.). And for this reason, I will now introduce the alternative, non-standard theory of neg-words.

(19) Ve zbarvení   je pstruh obecný tak variabilní jako žádná   naše ryba.
      in coloration is trout   brown as variable   like NEG.WORD our   fish
      'The brown trout is as variable in coloration as any of our fish.'

The alternative theory of neg-words was formulated in Ovalle & Guerzoni (2004), and a more recent reformulation can be found in Kuhn (2022). It shares some assumptions with the syntactic theory, though. First, both theories agree on the indefinite description status of neg-words. Therefore neg-words denote sortally existential quantifiers like in (20a) in the alternative theory too. The negative force, which in the syntactic theory is carried by the covert operator (the bearer of the classical logical semantic of ¬), is in the semantic/pragmatic theory reformulated as a presupposition of empty reference in the original version; see (20b). Or in the dynamic reformulation as a test on the cardinality of discourse referents, like in (20c).

(20) a. $[\![\text{neg-word}]\!] = \lambda P.\exists x[\text{SORT}(x) \wedge P(x)]$ TC
      b. $[\![\text{neg-word}]\!] = \neg\exists x[\text{SORT}(x) \wedge P(x)]$ non-at-issue
      c. after Kuhn (2022): $\wedge 0_x$ ...postsupposition (highest scope)

In this article, we can abstract away from the formal implementations and work with the core assumption: The emptiness of reference is a presupposition with the usual projection properties of presuppositions. One of the main differences concerns the interpretation of verbal negation in strict negative concord languages though. While in the standard theory/Zeijlstra (2004), the verbal negation is just an agreement negation with the active covert logical operator, which carries the logical negation, in the alternative theory, the verbal negation has its semantic interpretation, the classical propositional logic ¬.

The original version of the semantic/pragmatic theory/Ovalle & Guerzoni (2004) does not come with any locality constraints on the neg-word licensing, which is a problematic assumption since negative concord is, in most cases, limited to the clause-internal dependency between neg-words and verbal negation. This is also one of the reasons why the syntactic approach is so successful and remains the standard theory of neg-words today. Kuhn (2022) improves in many aspects over the original version of the semantic/pragmatic theory; one of them is the delimitation of the emptiness of reference presupposition in terms of previous contexts, and also in tying it to discourse referents and therefore

making the presupposition more specific. But most importantly, Kuhn (2022) brings some syntactic constraints into the game. He formalizes neg-words' syntax via split scope around their licensor (prototypically verbal negation). Since the split scope is realized via quantifier raising, some locality constraints on the neg-word emerge. More specifically, Kuhn's (2022) empirical claim is that the locality constraints on neg-word licensing should correspond to the locality of quantifier raising in the particular language and construction. Whether this is the right theoretical solution is a separate question, which is not answerable in this article. Still, it is definitely a step in the right direction, including some form of syntactic sensitivity for locality into the semantic/pragmatic theory.

Let us go through the predictions the semantic/pragmatic theory/Ovalle & Guerzoni (2004) makes concerning the baseline (simple root sentences with negated verbs), Neg-Raising sentences, and equatives. For the first environment, the predictions of the standard syntactic approach/Zeijlstra (2004) do not differ from the alternative one/Ovalle & Guerzoni (2004). Both approaches agree on the indefinite and positive at-issue meaning of neg-words. The syntactic theory delegates the negative property into the uninterpretable features; the alternative theory explains the negative force as a presupposition. In simple cases, like (18a), both theories predict grammaticality (either via feature checking or by the verification of the emptiness of the reference presupposition).[5]

---

[5]More interesting is how both theories account for the sentences like Spanish (i), where the neg-word c-commands a positive predicate. Such configurations are ungrammatical in strict negative concord languages, though. The syntactic approach does not have a straightforward answer for the ungrammaticality of such [Neg-word positive-V] sentences since both neg-words and verbal negation are posited to bear uninterpretable features, so it is not clear why one such feature is not enough to signal the covert Op$_\neg$. A way out is offered by Penka (2007) in (ii), but as she herself admits, the principle is not anything else than restating the problem. For the alternative theory, the answer for the ungrammaticality of (i) in the strict NC languages is straightforward: The presupposition of neg-word clashes with the assertion of the sentence, leading to a contradiction. Nevertheless, the alternative theory has to use more machinery to account for non-strict negative concord languages like Spanish, exemplified in (i). The solution, in a nutshell, lies in the accommodation of the emptiness of the reference presupposition, which can happen in specific circumstances. The technical details and extensive discussion can be found in Kuhn (2022).

(i) Nadie   vino.
     NEG.WORD came
     'Nobody came.'

(ii) Principle for the expression of negation:
      Mark sentential negation on the finite verb, unless this results in a different meaning.

The predictions of the standard theory/Zeijlstra (2004) concerning Neg-Raising were already introduced. The alternative theory/Ovalle & Guerzoni (2004) requires quantifier rising of the neg-word over its licensor (negation) in syntax, but since the scope of negation in Neg-Raising predicates ends on the embedded verb (but in the pragmatic part of the derivation), the alternative theory can predict somehow decreased acceptability of neg-words. Moreover, the emptiness of reference presupposition can be relativized to the belief or other possible worlds. Nevertheless, a full comparison of both theories with respect to Neg-Raising would have to take into account also non-Neg-Raising predicates and islands. Such configurations were not tested in the current experiment, though.

Finally, concerning the equatives, only the alternative theory of neg-words can reasonably explain why neg-words are licensed in the standard clause of equatives. First, the emptiness of reference presupposition can be satisfied in equative sentences like (19): It would require that no other fish (with the exception of brown trout) has the particular degree (on the scale of coloration) which is compatible with the truth conditions of the equative. Moreover, the split-scope part of the mechanics would need to quantifier-raise the neg-word over the given operator (MAX), and also, the dynamic properties of equatives would have to be checked off. Precise derivation of this must wait for future work, but the alternative approach has at least a good chance to derive the empirical asymmetry: Strict negative concord languages seem to allow the neg-words in the standard clauses of equatives but do not allow (strong) NPIs there. There are a couple of other environments studied before where such licensing of neg-words goes beyond negation: the complement of prepositions like *without*, and licensing of Spanish neg-words under verbs like *forbid, doubt* and *deny* (see Herburger 2001).

I will end this section via recapitulation of the predictions. As is clear, the predictions are very much theory-dependent, and for many patterns, the non-standard theories (either in the polarity or in the degree theories) are more promising than the established ones. Table 1 represents the predictions for three conditions: BAS(eline), N(eg-)R(aising) and EQ(uatives). We can expect that baseline will be acceptable for all speakers. Non-standard theories of neg-words/Ovalle & Guerzoni (2004) predict acceptance of neg-words in Neg-Raising predicates. Alternative neg-word theories/Ovalle & Guerzoni (2004) and alternative degree theories of equatives/Penka (2016) predict acceptance of neg-words in equatives and rejection of strong NPIs in equatives. On the other hand, standard syntactic and degree theories predict no neg-word licensing in equatives (rejection) and licensing of strong NPIs (acceptance), due to the downward-entailing environment of the equative clause. And likewise for neg-raising predicates.

Table 1: Expected acceptability (Czech speakers)

Condition	BAS	NR	EQ
strong NPIs (standard NPIs theories)	High	High	High
neg-words (standard neg-words theories/ Zeijlstra 2004)	High	Low	Low
strong NPIs (non-standard equative theories/ Penka 2016)	High	High	Low
neg-words (non-standard neg-words theories/Ovalle & Guerzoni 2004)	High	Low	High

1.2 Research questions

We will tackle two questions. The first question, in (21), is the main empirical question behind the experiment and, more generally, the search for the distinction between Czech strong NPIs and neg-words focused on one particular environment.

(21)   Question 1: Are Czech equatives acceptable with neg-words and unacceptable with strong NPIs?

The question is theoretically important since the current standard theories of equatives (like von Stechow 1984, Beck 2019) build upon the analysis of the *as*-clause of the equatives as downward-monotonic, therefore predicting at least grammaticality of weak NPIs and unacceptability of negation, negative quantifiers (and neg-words in languages with negative concord). This is the empirical pattern of English, but as suggested above, exactly the opposite is true for Slavic (as well as for German and other non-English) equatives. The experiment also scrutinizes Neg-Raising. For Neg-Raising the acceptability pattern is expected to be reversed (compared to equatives): Strong NPIs should be more acceptable than neg-words. But since both the standard/Zeijlstra (2004) and the alternative theory of neg-words/Ovalle & Guerzoni (2004) predict the same pattern in the case of Neg-Raising, the first research question is focused on equatives only. The theoretical consequence of the positive answer to the first research question is empirical support for non-standard theories of neg-words and equatives.

The second question, in (22), concerns the factors of the variation in the acceptability of strong NPIs.

(22)   Question 2: Is speaker variation of Czech strong NPIs caused by grammatical or demographic factors?

As introduced above, previous works on variation in polarity-sensitive expressions revealed that both grammatical and demographic factors could play various roles in the speaker variation of the negative-dependent expressions. Burnett et al. (2015) convincingly show that next to grammatical (syntactic) factors, a proper analysis of variation should control for age and education level, since these demographic factors explain some portion of the speaker variation for negative concord (absence or presence of negative concord in Montréal French in the case of Burnett et al. 2015). Since variation in speakers and their interpretation of strong NPIs was detected in previous research (Dočekal & Dotlačil 2017, Dočekal 2020), I included age, reading time (as a measure of education level or aspiration), and region as demographic questions in my experiment. The second research question in (22) phrases exactly this research agenda targeting demographic factors. And as will be shown, the experimental data give us precise-enough results to give some answers to both questions.

2 Experiment

The experiment aimed at answering the two research questions, (21) and (22): to test acceptability of neg-words and strong NPIs in Neg-Raising and in equatives. Next, the speaker variation was tested as well.

2.1 Methods

2.1.1 Participants & fillers

The experiment was run online on the L-Rex platform (Starschenko & Wierzba 2023). The participants were students of Masaryk University (Brno) and Charles University (Prague), and the majority of the students received credit for their participation. 105 participants filled out the experiment. The experiment included practice items to help subjects familiarize themselves with the acceptability judgment task, which was then used in the experiment itself. The experiment also included 64 fillers, half of them grammatical Czech sentences and half clearly ungrammatical sentences. Both halves of the fillers were complexity-wise similar to the items; the ungrammatical fillers included unlicensed anaphors and neg-words unlicensed by constituent negation, a.o. The exclusion rate was 66%

success. 82 of the participants passed the fillers, and their data points were included in the analysis.[6]

2.1.2 Materials & procedure

Each questionnaire consisted of 64 items, and there were 48 randomized lists generated from the items by L-Rex. The questionnaire started with three demographic-related questions: (i) the age of the participant, (ii) the region of the participant during their first language acquisition, and (iii) their daily reading time (explained as reading time of books and/or journals, not looking at the screen of phones, etc.). Each participant filled out 128 trials (half items, half fillers) in the acceptability part of the experiment, which is the part reported in the present article.[7]

The experiment consisted of two parts: (i) an acceptability judgment task where sentences were judged without context, and (ii) an acceptability judgment task where sentences were judged against a probability/scalarity manipulated context (see footnote 7). In both parts, participants judged the acceptability of sentences on a 1 to 7-point Likert scale (1 the worst – the least acceptable, 7 the best – the most acceptable). In both parts, all conditions were crossed with two conditions: (i) NEG-WORDS, (ii) STRONG NPIS.

An example item from the experiment is in (23). There were three conditions: (i) BAS(ELINE), (23a), (ii) Neg-Raising, NR, (23b), and (iii) equative, EQ, (23c). All three conditions were crossed with two types of negative polarity expression: (i) neg-words *žádný*, z, and (ii) strong NPIs *ani*, A. Therefore the experiment was a 3x2 design. A mnemonic for crossed conditions for baseline are BASA for strong NPIs, and BASZ for neg-words.

(23)   a.   V království nezůstal   {žádný   / ani jeden} zloděj.
           in kingdom   NEG.remained   NEG.WORD   NPI   thief
           'No thief remained in the kingdom.'

---

[6]The criterion was whether the subject was more than 66% successful in fillers or not. One of the two anonymous reviewers asked about the details of the exclusion rate and also why the most-used standard rate, 75%, was not used. The exclusion criterion was measured as follows: For each participant the difference between answers to good and bad fillers was computed – since the scale is 7-point, the difference was in the interval 0 to 6. The success rate was then computed on the difference scale for each subject. As a sanity check, I ran the analysis with 75% exclusion rate during the revisions of the article. The descriptive and inferential statistics remained the same as in the original analysis modulo changes in the second digit to the right of the decimal point; also the strength of the effects remained the same.

[7]The other part of the experiment was an acceptability judgment task with probability/scalarity manipulated. The results of the second part are not reported in the present article due to space reasons.

b. Král nechce,   aby v království zůstal   {žádný   / ani jeden}
   King NEG.wants that in kingdom   remained NEG.WORD   NPI
   zloděj.
   thief
   'The king doesn't want any thief to remain in the kingdom.'

c. Zloděj ze   souostroví Qwghlm je tak šikovný jako {žádný   /
   thief   from archipelago Qwghlm is so   clever   how NEG.WORD
   ani jeden} zloděj.
   NPI   thief
   'The thief from the Qwghlm archipelago is as clever as any other
   thief.'

2.2 Predictions

All the discussed theories predict that neg-words and strong NPIs will be accepted in the baseline condition, BAS. The condition is present in the experiment to check how much worse the other two conditions will be compared to the baseline.

On the contrary, the standard theory of neg-words/Zeijlstra (2004) and the alternative theory of neg-words/Ovalle & Guerzoni (2004) differ in their predictions for equatives, EQ: The standard theory predicts neg-words to be not acceptable in EQ while the alternative theory is compatible with their acceptability in EQ. The standard theory of equatives/Rullmann (1995) predicts the acceptability of strong NPIs in EQ while the alternative theory of equatives/Penka (2016) predicts their unacceptability.

Finally, both theories of neg-words predict the decreased acceptability of neg-words in NR and much higher acceptability of strong NPIs in NR.

Concerning the speaker variation, there are no theory-specific predictions. But since the speaker variation was observed in previous research, it is expected that the variation will be observed in the current experiment too. The demographic factors were included in the experiment to test whether the variation is related to grammatical or demographic factors.

2.3 Results

The descriptive-statistics results can be seen in the graph of acceptance, including error bars, in Figure 1.[8] As can be seen with the naked eye, both expressions

---

[8]The graph of acceptance in Figure 1 uses the standard Cartesian coordinate system with the *y*-axis origin at 0. That does not mean that the response scale was 0 to 7 but is simply the default behavior of the GGPLOT2 R package (see Wickham 2016).

are nearly at the ceiling in BAS, but their acceptability in the two other conditions is reversed. While neg-words are much more acceptable in equatives (EQZ), strong NPIs are preferred in Neg-Raising contexts (NRA).[9]
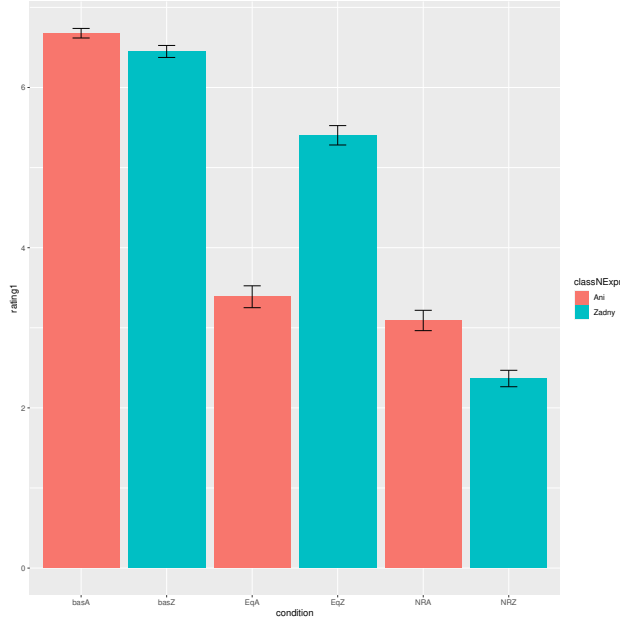
Figure 1: Graph of acceptance (+error bars) conditions: BAseline, EQuative, NegRaising expressions: Ani (strong NPI), zádný (neg-word)

2.3.1 Inferential statistics

The Bayesian hierarchical random-effects model with default priors was fit using the R package RSTANARM (Goodrich et al. 2022): The dependent variable was the subject's response; the independent variables were: (i) environment (BAS, EQ, NR), (ii) type of the polarity-dependent expression (A, Z), and their interaction; the reference level was BAS, A. The baseline was selected as the condition which should be uncontroversially accepted by speakers, which indeed was the case since no main effect was positive against the baseline. The model included random effects for both subject and item intercepts.

---

[9] One of the two anonymous reviewers asked whether unacceptable fillers were judged worse than neg-words (or strong NPIs) in Neg-Raising contexts. The answer is yes, the ungrammatical fillers' median acceptability was 1, while the median acceptability of neg-words and strong NPIs was 2.

The model was fit to the data and we found that (i) the baseline was very well accepted (Intercept = 6.67, 95% C(redibility) I(nterval)= [6.38, 6.95]); there is no distinction between neg-words and strong NPIs in it and sine qua non, both expressions are acceptable to the same extent (posterior main effect in the form of median and 95% CI: $\hat{\mu} = -0.20$, CI $= [-0.47, 0.08]$); (ii) neg-words were much better accepted in equatives than strong NPIs (the positive interaction of EQ by z: $\hat{\mu} = 2.18$, CI $= [1.81, 2.58]$ – against the reference level); (iii) strong NPIs were preferred in Neg-Raising (the negative interaction of NR by z: $\hat{\mu} = -0.53$, CI $= [-0.91, -0.14]$ – against the reference level).[10] The results are also supported by the results of R(egion) O(f) P(ractical) E(quivalence), ROPE: Only z is not significant, since it is 23% in ROPE. For all medians, confidence intervals, and ROPE percents, see Table 2; all percents of ROPE are computed for the interval $[-0.10, 0.10]$. Medians, 95% credibility intervals, and ROPE are also visually represented by the graph in Figure 2. Notice that in Figure 2, as is usual in Bayesian modeling, the reference level condition (BAS, A) is coded as 0 of the x-axis and each condition (or interaction of condition) has its own y-line (with the distribution and median); the credibility of a condition can be visually inspected via observation of the condition including 0 (e.g. clearly z) or differing from it either positively (e.g. the interaction between EQ by z) or negatively (e.g. NR).

2.3.2 Demographic factors

Next, three Bayesian generalized mixed linear models were fitted to detect the effects of demographic factors inhibiting or prohibiting acceptability. This was important since previous work (see Burnett et al. 2015 and Burnett et al. 2018, a.o.) revealed that both grammatical and demographic factors are at play when negative polarity variation is linguistically studied. As a reminder, the experiment included three demographic questions: region, age, and daily reading time. The last factor was used as a proxy for investigating educational level. The selection of factors was influenced by the previous work on variation in negative dependent expressions: Burnett et al. (2015) have shown that age, education level, and location of the speakers can have an impact on the variation. And since in

---

[10] As one of the two anonymous reviewers correctly points out, NR (whether it includes strong NPI or neg-word) is surprisingly poorly accepted – see the row of NR in Table 2. I agree, but this seems to be the case generally; the results of my experiment resonate with the experimental finding from Dočekal & Dotlačil (2016a), where the strong negative effect between the acceptable baseline and tested Neg-Raising was observed. The reasons for this negative effect are unclear, but see Alexandropoulou et al. (2020) for some other environments where including NPIs leads to a strong decrease in acceptability even if such an effect is theoretically unexpected.

Table 2: Bayesian model and its posterior distribution for the experiment

Parameter

	Median	CI	% in ROPE
Intercept	6.67	$[6.38, 6.95]$	0%
EQ	−3.27	$[−3.53, −3.00]$	0%
NR	−3.57	$[−3.86, −3.30]$	0%
Z	−0.20	$[−0.47, 0.08]$	23.08%
EQ:Z	2.18	$[1.81, 2.58]$	0%
NR:Z	−0.53	$[−0.91, −0.14]$	0%

Random effects

	Name	SD
subject	Intercept	0.57
item	Intercept	0.35
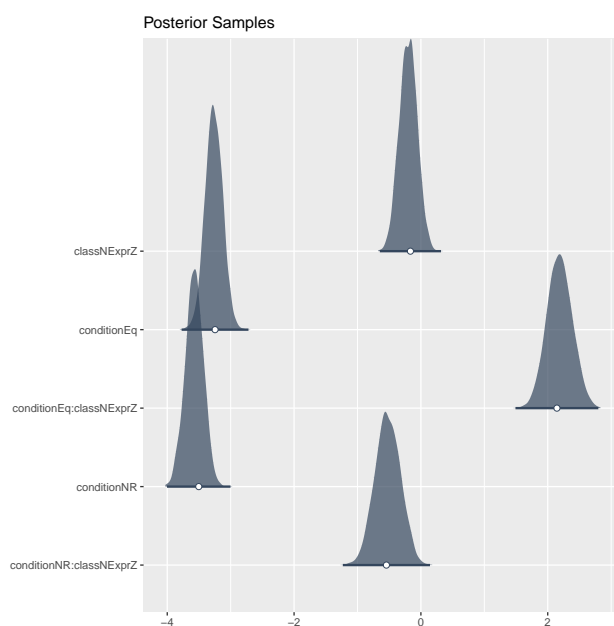
Figure 2: Graph of posterior samples with ROPE (for the experiment)
conditions: BAseline, EQuative, NegRaising
expressions: Ani (strong NPI), zádný (neg-word)

the previous experimental work it was revealed that there are idiolects of Czech speakers interpreting the strong NPI *ani* (Dočekal & Dotlačil 2017, Dočekal 2020), I included the three mentioned demographic factors to detect whether the variation can be traced to some extralinguistic sources eventually. Nevertheless, it has to be said that the pool of subjects was rather homogeneous, consisting mainly of university students. Therefore, at least the education-level results should be taken with a grain of salt and mainly as a first step in the general description of polarity items' variation in Slavic languages. Also, for this reason, I move the inferential statistics into footnotes and describe the main outcomes in terms of descriptive statistics.

Let us start with AGE. Descriptively, AGE ranged from 19 to 71, with a median = 23, mean = 25.59, and sd = 9.47. The age was first z-transformed and then plugged in as the third interaction variable in the Bayesian model (next to the two conditions, Z and BAS/NR/EQ environment).[11] But the model did not confirm any effect on overall acceptability or any particular age-related inhibition or prohibition of any construction or negative-dependent expression.

Next, REGION was more varied than AGE, where the data points from the first to the third quantile were in the range of 21 to 25. But since I did not control for the specificity of the values entered into the form, the answers ranged from city-specific to region-specific. For this reason, I aggregated all the answers into a discrete factor with two levels: MORAVIAN and NONMORAVIAN. 67% of subjects entered as their region NONMORAVIAN; the remaining 33% identified themselves as being from Moravia. Again the factor REGION was used as the third interaction variable in the Bayesian model.[12] Overall, the REGION did not increase or decrease acceptability, but there is some anecdotal evidence for higher acceptance of Neg-Raising in the Czech (non-Moravian) part of the population. Nevertheless, the interaction effect is so weak that I doubt there is any genuine linguistic Neg-Raising isogloss between Czech dialects.

---

[11] The model revealed that acceptability overall was not affected by age at all (main effect of AGE: $\hat{\mu} = 0.01$, CI = $[-0.25, 0.28]$, ROPE: 58.00% for the $[-0.10, 0.10]$ interval). There was also no significant interaction with any single condition or pair of conditions. The lowest ROPE was 30.55% for the three-way interaction between EQ:Z:AGE. All other interactions had an even bigger portion in ROPE and were also less significant.

[12] The main effect of the region was not credible ($\hat{\mu} = 0.33$, CI = $[-0.18, 0.88]$, ROPE: 14.79% for the $[-0.10, 0.10]$ interval) but this time there was very weak evidence coming from interactions. Namely, there seemed to be a slight tendency for higher acceptance of Neg-Raising in the non-Moravian part of the Czech Republic (the interaction NR:MORAVIAN: $\hat{\mu} = -0.61$, CI = $[-1.28, 0.04]$, ROPE: 3.89% for the $[-0.10, 0.10]$ interval). All other interactions with REGION were less significant.

The last demographic factor was reading time. As hinted above, the factor was used to get information about education or education aspirations. The answers (converted to hours) ranged from 0 to 10 hours, with 1 hour as the median, 1.43 hours as the mean, and the range of first and third quantiles being 1 hour and 2 hours, respectively. Similarly to AGE, data points are centered around the mean with a small standard deviation, 1.26, and few outliers. As in the case of REGION, I recorded the continuous variable as a factor READINGTIME with two levels: OVER1HOUR,UNDER1HOUR dividing the sample according to the median value of reading time. The result was two nearly proportional halves: 52% of the subjects claimed that their daily reading time is under 60 minutes, and the remaining 48% entered that they read more than one hour. The third demographic factor (READINGTIME) was plugged into the Bayesian model as an independent (interaction) variable.[13] The modeling results show that there is some weak evidence for the positive correlation between reading time and the acceptance of the Neg-Raising construction: Subjects who claimed to read more were more accepting of the Neg-Raising construction. Such a tendency is intuitively plausible but does not say anything linguistically important about the constructions and polarity-dependent expressions tested in the experiment.

Let us summarize: The design of the experiment and three demographic questions did not reveal any important information concerning the demography-related variation in polarity constructions of Czech speakers. Two weak effects can be interpreted as clues about region and education-level variation concerning Neg-Raising. Still, there seems to be nothing significant in the variation of *ani* vs. *žádný* in the studied constructions. So, whatever speaker variation (in the usage of *ani*) we will discuss further, it seems not to be related to age, region, or education level as revealed by the sample of the experiment (in this respect, the results of the experiment are different from the previous work on speaker variation in polarity dependent expressions, like Burnett et al. 2015, 2018).

2.3.3 Correlations

The Bayesian model revealed that both non-baseline environments (NR and EQ) were less accepted by speakers, but there was no difference between *ani* and

---

[13]And again, as with two previous demographic factors, the main effect of READINGTIME was not credible ($\hat{\mu} = -0.13$, CI = $[-0.63, 0.39]$, ROPE: 28.89% for the $[-0.10, 0.10]$ interval). And similarly to REGION, there was one weakly-credible interaction: subjects claiming to read more than average (over 60 minutes daily) were accepting Neg-Raising constructions more (NR:OVER1HOUR interaction: $\hat{\mu} = 0.66$, CI = $[0.04, 1.27]$, ROPE: 1.24% for the $[-0.10, 0.10]$ interval). All other interactions were much less credible.

žádný in terms of main effects. Nevertheless, speakers accepted in equatives many more neg-words than strong NPIs (the strong effect, and the only one which yielded positive interaction). But speakers are also inclined to reject neg-words in Neg-Raising against strong NPIs (the negative interaction effect between z and nr). The theoretical consequences of these findings will be discussed below, but let us turn to another kind of variation, this time not demographic.

The first important thing to note is that all speakers agreed on their high acceptance of baseline, and in this condition, they accepted neg-words and strong NPIs indistinguishably. But this acceptance of both polarity expressions diverged in the two other conditions. Namely, some speakers rate *ani* high in equatives (unlike the main thrust of speakers; recall the strong positive interaction between z and Eq) but also reject it in NegRaising (again going against the overall acceptance of strong NPIs there: the negative interaction between z and nr). And vice versa, subjects who reject strong NPIs in equatives (behaving according to the negative interaction effect) accept strong NPIs in Neg-Raising (again verifying the negative interaction effect). In both cases, we observe a negative correlation between the acceptability of neg-words/strong NPIs in the two environments, equatives and Neg-Raising. One way to understand this reversed correlation is to assume that the first kind of speaker (those who accept *ani* in equatives) treats *ani* more like a neg-word and not like a strong NPI. The rest of the sample (the majority, in fact) treats *ani* as a strong NPI and therefore accepts it under Neg-Raisers and rejects it in equatives.

Therefore, post-hoc correlation statistics were run and are reported below. Notice, though, that the correlation is post-hoc in the sense of interpreting sub-clusters of speakers (let us say idiolects in linguistic terms) but not in the sense of avoiding the Type I error (mistakenly rejecting the null hypothesis) since the null hypothesis is not an important part of Bayesian statistics; instead the bulk of the inference statistics in Bayesian framework is posterior distribution, which represents the probability of the parameters of a model given the data and which was reported here in Section 2.3.1. The motivation for the correlation analysis comes both from the observed variation introduced above and from the previous work on Czech strong NPIs (Dočekal & Dotlačil 2017, Dočekal 2020), where it was observed that there are idiolects of Czech speakers concerning their strong-NPI interpretation.

The way the correlations were checked statistically is the following. First, the acceptance of conditions was z-transformed (by subject). Then, such z-transformed variables were checked for correlations across conditions. Indeed, there is a strong negative correlation between the acceptability of *ani* in equatives and its acceptability under Neg-Raising predicates (Pearson's product-

moment correlation: $t = -5.93$, $p < 0.001$). The correlation graph is in Figure 3. This means that we can identify two groups of speakers: (i) speakers who accept *ani* under Neg-Raisers and reject it with equatives (top left section in Figure 3), and (ii) speakers who accept *ani* in equatives and reject it under Neg-Raisers (the bottom right part of Figure 3). But crucially, no speakers are accepting both conditions (the empty top right corner) nor speakers who would reject both conditions (the empty space in the bottom left part). And also, there is no correlation between the acceptability of *ani* in the baseline and equatives, just pure noise, as can be seen in Figure 4. This correlation of *ani* between NR and EQ resonates with the previous work (Dočekal & Dotlačil 2017) where similar correlations were found (for *ani*) in the case of probability-manipulated conditions and Neg-Raising.
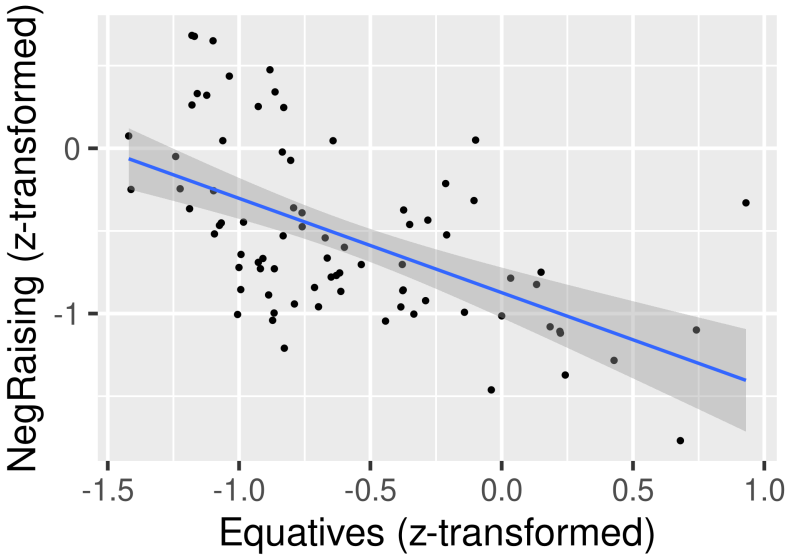
Figure 3: Correlations between Equatives and Neg-Raising for *ani*

2.3.4 Discussion

The reported experimental results suggest that Czech strong NPIs are (for most speakers) accepted under Neg-Raising predicates (NR showed a credible negative effect but the negative interaction effect NR by z shows that participants prefer strong NPIs in NR) and rejected in equatives (the main effect of EQ is negatively strong, but there is a very strong positive interaction EQ by z indicating robust
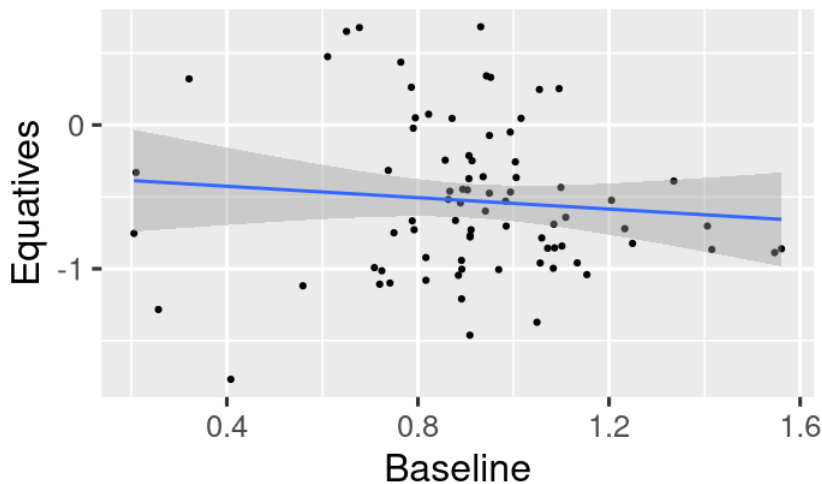
Figure 4: Correlations between Equatives and Baseline for *ani*

preference for neg-words in equatives). For Czech neg-words, the opposite is true: Most speakers accept them in equatives but reject them under Neg-Raising predicates. In the case of *ani* (strong NPI), there is speaker variation, and some subset of speakers treats it more like a neg-word; nothing similar was found for neg-words. The correlation discussed in Section 2.3.3 suggests that there are two kinds of speakers: (i) speakers accepting *ani* in NR and rejecting it in EQ (strong NPI treatment of *ani*), and (ii) speakers accepting *ani* in EQ and rejecting it in NR (the speakers who use *ani* more like a neg-word). Moreover, the speaker variation does not seem to be derivable from demographic factors (or at least not from the demographic factors controlled in the experiment).

3 Theoretical consequences

The nature of this article is mainly experimental. For this reason and the obvious constraints of space, the consequences of the analyzed experimental data will be discussed only to a limited extent.

3.1 Application: strong NPIs

Now, let us demonstrate how the approaches introduced in Section 1.1 can be applied to the data gathered in the experiment. First, concerning the strong NPIs

(Czech *ani*), I will show how the *even*-approach fits the Czech data from the experiment, starting with the baseline. For a reminder, the Table 2 summarizes predictions of the theories for the experiment. The results of the experiment confirm more the non-standard theories of neg-words/Ovalle & Guerzoni (2004) and non-standard theories of equatives/Penka (2016).

As discussed above, *ani* comes with the *even*-presuppositions, namely the scalar and the additive. In formal terms, this is translated as an association of *ani* with covert *even* scoping at the propositional level; see a schematic representation of the baseline in (24) and its logical form in (24a). Since the sentence is negated, the entailment between numerals is reversed by negation: ¬($[\![$ one thief $]\!]$ …) ⊨ ¬($[\![$two thieves$]\!]$ …). The alternatives to the prejacent come from the focused numeral, since *ani* can associate with nouns, clauses, etc., but as usual in Slavic languages, it associates mostly with its sister node (the numeral in (24)). Because the entailment is reversed, the *even*-approach predictions agree with the downward-entailing explanation. Moreover, since *ani* is a strong NPI, it requires DE/*even*-presuppositions to be satisfied both in truth conditions (configurationally β in (24a)) but also at the level of non-at-issue meaning (where the exhaustifier, silent *even*, scopes: α in (24a)). Next, we have to check both presuppositions as schematically formalized in (24d) and (24e), which are also fulfilled. The theoretical prediction of the standard account/Gajewski (2011) then agrees with the high baseline acceptability of *ani*: Even the inferential statistic baseline median intercept was 6.67 on the 7-point Likert scale.

(24) Ani [one]$_F$ thief neg-remained in the kingdom.

   a. $[_\alpha$ (*even*) $[_\beta$ ¬$[_\gamma$ ani one thief remained in the kingdom $]]$ $]$
   b. TC (in β) DE: ✓
   c. non-at-issue (in α) DE: ✓
   d. scalar presupposition of (even): → ¬(two thieves remained), ¬(three thieves remained), …: ✓
   e. additive presupposition: ¬(two thieves remained) ∨ ¬(three thieves remained), …: ✓

Continuing now to the Neg-Raising condition, neg-words were in NR less accepted than strong NPIs; although the effect was not particularly strong, it was still significant. The theories of Neg-Raising (Gajewski 2007, Romoli 2013) predict that for the Neg-Raising predicates the scope of root negation ends (via presupposition or implicature calculation) in the embedded clause. Schematically we can formalize the important ingredients of the LF for the experiment NR conditions as in (25).

(25) The king does ~~not~~ want $[_\alpha$ (*even*) $[_\beta \neg$ [ani [one]$_F$ thief remained in the kingdom]]].

From the point of view of the standard theory of strong NPIs and the *even*-presuppositions, the logical form of the embedded clause is the same as in the case of the baseline, therefore the explanation of the NR acceptability is the same as in the baseline case. What differs is the actual acceptance by speakers (much lower in the case of Neg-Raising than in the baseline), but as discussed in footnote 6, in this respect, our experiment nearly replicates the previous findings with respect to Neg-Raising and NPIs generally. The second difference concerns the diverging acceptability of strong NPIs and neg-words in Neg-Raising contexts (both kinds of expressions were indistinguishably well-accepted in the baseline). Still, this point will be discussed in Section 3.2.

Finally, the standard theory of equatives predicts that strong NPIs can be licensed in the standard clause. This was argued before to be wrong for German and Romance languages (Krifka 1992, Penka 2016) and also clashes with the intuition of Czech speakers, for whom strong NPIs were much less accepted than neg-words in the standard clauses of equatives. A proper investigation of Slavic equatives must wait for future work, but let us take the first steps in this direction. If we take seriously the morpho-syntax of Slavic equatives and follow the theoretical hints from Penka (2016), it is possible to model the equative conditions from the experiment as in (26).

(26) The thief from the Qwghlm archipelago is so^2 clever $[_\alpha$ (*even*) $[_\beta$ MAX$_{inf}$ how$_1$ ani [one]$_F$ thief t_1 ~~clever is.~~]]^2

First, since the morpho-syntax of Czech (and Slavic, Romance, German) equatives is built upon the correlatives, anaphoric *so* (Czech *tak* 'so', see (23c) for full glosses) picks up the referent of the definite degree description (index 2). The definite degree description is yielded by the MAX$_{inf}$ operator, similar to free relatives. And since free relatives are known not to license NPIs (Jacobson 1995), the correlative standard equative clause is expected not to be a good environment for NPIs. Therefore we can predict, if we assume this non-standard but well-motivated theory of equatives, why in Czech equatives, strong NPIs are not accepted: Since β is most probably not DE, even if the presuppositions of silent *even* in α were satisfied (MAX$_{inf}$ makes the standard of equatives most probably non-monotonic), the licensing of strong NPIs is not satisfied. The much better acceptance of neg-words in equatives will be discussed in Section 3.2.

3.2 Neg-words

In this section, I will present the application of theories introduced in Section 1.1.4 to the results of the experiment.

The baseline can be explained easily both in the standard theory of neg-words/ Zeijlstra (2004) and in the alternative theory of neg-words/Ovalle & Guerzoni (2004). I discuss just the alternative theory explanation here. In a negative sentence, like schematic (27), the truth conditions (indefinite descriptions) of the neg-word and its presupposition agree: The indefinite description is under the scope of negation, and the presupposition of the emptiness of the discourse referent is compatible with the truth conditions. But in a positive minimal pair sentence, like (28), the existential quantification over the discourse referent and the presupposition of its reference emptiness would clash. Therefore, the positive minimal pair sentence is predicted to be unacceptable. This is exactly what we observe in the experiment: The baseline is very well accepted, but the positive minimal pair is rejected.

(27) neg-word thief neg-remained in the kingdom.

    a. $[\neg[\exists x[\textsc{thief}(x) \land \textsc{remained}(x)]\,]\,] \land 0_x$

(28) neg-word thief remained in the kingdom.

    a. $[\exists x[\textsc{thief}(x) \land \textsc{remained}(x)]\,] \land 0_x$ $\bot$

Turning now to Neg-Raising, neg-words were less accepted in NR than strong NPIs. From the perspective of the alternative theory of neg-words/Ovalle & Guerzoni (2004), we should expect that this would align with the locality constraints: The neg-words have to quantifier-raise over their licensors, creating a split-scope logical form. But since the root negation ends in the embedded clause in the case of Neg-Raising predicates, as discussed above, the schematic logical form for Neg-Raising conditions can be rendered as (29). This reasoning explains the difference between relatively freely-licensing (both for neg-words and strong NPIs) Neg-Raising predicates and much worse non-Neg-Raising predicates (verbs of causation or communication) in such configurations like (29) – see Dočekal & Dotlačil (2016a) for experimental data and analysis. Recall that experimental results showed a slight preference for strong NPIs in this condition. The difference can be captured as follows: Since the split scope relies on syntactic mechanisms and the negation ends in the embedded clause in the pragmatic part of the derivation, there is a slight timing issue which maybe can be coerced. A similar kind of explanation can be retold in the standard approach/Zeijlstra (2004), too: The syntactic licensing of features should proceed before the pragmatic mechanisms

like presupposition and implicature calculation. But in both kinds of explanation, a lot hinges upon the assumptions about the architecture of grammar and many other assumptions. Moreover, it is not totally clear how to linguistically interpret the effects from the experiment – as Table 2 shows, both kinds of expressions (strong NPIs and neg-words) lead to a dramatic acceptability decrease (−3.57), but there is the slight preference for strong NPIs (the negative interaction of neg-words with Neg-Raising: −0.53).

(29)   The king wants ~~not~~ that [neg-word thief$_2$ ¬ [t_2 remained in the kingdom]].

Finally, concerning the equatives: (i) As tested in the experiment, neg-words are, but strong NPIs are not, acceptable in the complement clause of the Czech equatives; (ii) adding to this, verbal negation is not acceptable either – see (30). The high acceptability of neg-words is especially surprising from the perspective of English since negative quantifiers are distinctly odd in this position (in comparatives, but as discussed in 1.1.3, the expectations are – in the standard theories – similar for comparatives and equatives), see (31) from Gajewski (2008).

(30)   Petr je tak chytrý jak {nikdo   jiný / *Marie ne / *ani   jeden}.
       Petr is so smart as   NEG.WORD else Mary not STRONG.NPI
       'Petr is as smart as anyone.'

(31)   *Mary is taller than no boys are.

The ambition of this article is not to solve the above-mentioned theoretical puzzles. But let us at least indicate where a possible solution can be. The experimental results show that Czech neg-words are very much accepted in the complement clause of equatives, while strong NPIs are degraded there. Moreover, intuitions and preliminary results from the follow-up experiment suggest that weak NPIs are not acceptable in the equatives either, following the German and Romance data discussed in Krifka (1992) and Penka (2016). One possible explanation is that Czech complement clauses of equatives are not downward-monotonic in either truth conditions or non-at-issue meaning. But for some reason, the emptiness of the neg-words' discourse referents' presupposition can be easily satisfied in this environment. Consider (32): The emptiness of discourse referents' presupposition here would be that there is no such thief with the degree of cleverness *d* which would exceed the degree of the Qwghlm archipelago thief. Such a presupposition is plausible, and more generally, it can be said that Czech equatives

are one of the rare environments where the neg-words can be licensed by expressions not including negation (similarly to Spanish verbs like *forbid, doubt* or *deny*).

(32) The thief from the Qwghlm archipelago is so^2 clever [MAX$_{inf}$ how$_1$ neg-word thief t_1 ~~clever is~~.]^2

This can be compatible with Penka's (2016) suggestion to replace the MAX operator in analyzing English equatives with a different relation on the degrees, MAX$_{inf}$ discussed in Section 3.1. But so far, I consider the evidence to be inconclusive regarding the monotonic properties of Czech complement clauses of equatives; they are not downward-entailing for sure, but the resulting two possibilities, upward entailing or non-monotonic, are still open. From a theoretical standpoint, I agree with Penka (2016) that current degree theories of equatives do not hold the cross-linguistic water. And in the same direction, it is clear that a purely syntactic approach to neg-words faces big trouble when posed with the acceptability of neg-words in equatives. No matter how the cross-linguistically feasible degree theory of equatives will look like (e.g., using MAX$_{inf}$ as suggested by Penka 2016), there is clearly no room for a sentential negation operator in its version for Slavic (and Romance) equatives, since then the weak or strong NPIs would be admissible there, contrary to facts. Concluding this section, merging the non-standard theory of neg-words/Ovalle & Guerzoni (2004) and the non-standard theory of equatives/Penka (2016) is promising as a theoretical explanation for the high acceptability of neg-words in the standards of Czech equatives and the very low acceptance of strong NPIs there.

4 Summary

The findings of the current study provide an answer to research question 1, repeated below as (33). The experiment confirmed the robust acceptability of neg-words in the standard clause of equatives. This can be explained as deriving from the neg-words presupposition relativized to the set of degrees introduced in the main clause if we follow the alternative theory of neg-words/Ovalle & Guerzoni (2004). Neg-words in examples like (30) are accepted since, in this configuration, the presupposition does not require total emptiness of reference, just the emptiness of reference for such discourse referents whose degree would exceed the degree of the subject. Next, the strong NPI unacceptability in Czech equatives is a direct consequence of the Czech equatives complement clauses not being downward-monotonic. The sub-answers to research question 1 are in (34).

The results of the experiment bring empirical support for alternative theories of neg-words/Ovalle & Guerzoni (2004) and alternative theories of equatives/Penka (2016).

(33) Question 1: Are Czech equatives acceptable with neg-words and non-acceptable with strong NPIs?

(34) The non-standard theories of negative concord and equatives give promising answers:

    a. The semantic/pragmatic theory of neg-words allows the presupposition of discourse emptiness to be satisfied (relativized to degrees of the main clause).
    b. The complement clause of the Czech equatives is not downward-entailing.

Now, research question 2 is repeated below as (35). Some speaker variation was observed. Recall that for some speakers *ani* behaved more like a neg-word. Nevertheless, it is not likely that the variation can be related to demographic factors such as age, region, or daily reading time. But there is one way to theoretically explain the variation: We can assume that both neg-words and strong NPIs in Czech come with presuppositions, the emptiness of discourse referents for neg-words, and the scalar presupposition for strong NPIs. Then the flux from strong NPIs to neg-words can be theoretically cashed out as follows: Speakers substitute the scalar presupposition with the emptiness of discourse referents' presupposition. Speculatively, we can try to explain the one-way direction in terms of economy: The scalar presupposition needs a covert exhaustifier, but the emptiness presupposition does not. Therefore it is less costly and more attractive for speakers who oscillate between the two presuppositions. For this reason, there is no speaker variation concerning neg-words: Adopting the scalar presupposition would mean a less economical logical representation. Why the flux is unrelated to the demographic factors is an issue for future work. The answers are summarized in (36).

(35) Question 2: Is speaker variation of Czech strong NPIs caused by grammatical or demographic factors?

(36) The speaker variation is explainable as shifting from the scalar to the emptiness of the DR presupposition (in the case of *ani jeden* 'even one').
    a. Social factors don't seem to play a role in this shift.

Let us end this section with some open questions. The first of them concerns the locality constraints on neg-words' licensing. The alternative theory of neg-words/Ovalle & Guerzoni (2004) predicts that the neg-word locality should approximate the quantifier raising. Only syntactic islands (such as relative clauses) should be hard limits for both neg-word licensing and quantifier raising. The syntactic literature on the topic of Slavic quantifier raising seems to argue for a possibility of overt movement (out of non-island clauses) but obligatory reconstruction (see Neeleman & Titov 2009, a.o.). Still, the experimental research in this direction seems limited to mono-clausal conditions (see, e.g., Ionin & Luchkina 2018). So there is space for future research in this direction, and only then can we conclude whether the locality constraints between quantifier raising and neg-word licensing coincide. Another open question concerns the cross-linguistic variation in neg-word licensing: In Romance languages, neg-words are licensed in *before*-clauses and under *doubt*-type predicates; in Slavic languages, this is not the case. The alternative theory of neg-words/Ovalle & Guerzoni (2004) predicts that this should follow from the different presupposition projection properties in the two types of languages. Whether this is true remains again a question for future work.

Abbreviations

1	first person	NEG	negation
AUX	auxiliary	NPI	Negative Polarity Item
EQ	equative	NR	Neg-Raising
COMP	comparative	SG	singular

Acknowledgments

References

Alexandropoulou, Stavroula, Lisa Bylinina & Rick Nouwen. 2020. Is there *any* licensing in non-DE contexts? An experimental study. In Michael Franke, Nikola Kompa, Mingya Liu, Jutta L. Mueller & Juliane Schwab (eds.), *Proceedings of Sinn und Bedeutung 24*, 35–47. University of Konstanz: Gesellschaft für Semantik. DOI: 10.18148/sub/2020.v24i1.851.

Beck, Sigrid. 2019. Comparison constructions. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: Lexical structures and adjectives*, 415–476. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110626391-013.

Błaszczak, Joanna. 2001. *Investigation into the interaction between the indefinites and negation*, vol. 51 (Studia Grammatica). Berlin: Akademie Verlag.

Burnett, Heather, Hilda Koopman & Sali A. Tagliamonte. 2018. Structural explanations in syntactic variation: The evolution of English negative and polarity indefinites. *Language Variation and Change* 30(1). 83–107. DOI: 10.1017/s0954394517000266.

Burnett, Heather, Mireille Tremblay & Hélène Blondeau. 2015. The variable grammar of negative concord in Montréal French. *University of Pennsylvania Working Papers in Linguistics* 21(2). 3. https://repository.upenn.edu/handle/20.500.14332/45068.

Chemla, Emmanuel, Vincent Homer & Daniel Rothschild. 2011. Modularity and intuitions in formal semantics: The case of polarity items. *Linguistics and Philosophy* 34(6). 537–570. DOI: 10.1007/s10988-012-9106-0.

Chierchia, Gennaro. 2019. Factivity meets polarity: On two differences between Italian versus English factives. In Daniel Altshuler & Jessica Rett (eds.), *The semantics of plurals, focus, degrees, and times*, 111–134. Cham: Springer. DOI: 10.1007/978-3-030-04438-1_6.

Chisholm, Hugh (ed.). 2010. *Encyclopaedia Britannica*, vol. 5. Urbana: Project Gutenberg. https://www.gutenberg.org/files/33052/33052-h/33052-h.htm.

Crnič, Luka. 2011. *Getting even.* Cambridge: MIT. (Doctoral dissertation).

Crnič, Luka. 2014a. Against a dogma on NPI licensing. In Luka Crnič & Uli Sauerland (eds.), *The art and craft of semantics: A Festschrift for Irene Heim*, vol. 1, 117–145. Cambridge: MIT Press. https://semanticsarchive.net/Archive/jZiNmM4N/Crnic.pdf.

Crnič, Luka. 2014b. Non-monotonicity in NPI licensing. *Natural Language Semantics* 22(2). 169–217. DOI: 10.1007/s11050-014-9104-6.

Djärv, Kajsa, Jérémy Zehr & Florian Schwarz. 2018. Cognitive vs. emotive factives: An experimental differentiation. In Robert Truswell, Chris Cummins, Caroline Heycock, Brian Rabern & Hannah Rohde (eds.), *Proceedings of Sinn und Bedeutung*, vol. 21, 367–386. University of Konstanz: Gesellschaft für Semantik. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/143.

Dočekal, Mojmír. 2020. N-words and NPIs: Between syntax, semantics, and experiments. In Franc Marušič, Petra Mišmaš & Rok Žaucer (eds.), *Advances in formal Slavic linguistics 2017*, vol. 3, 15–49. Berlin: Language Science Press. DOI: 10.5281/zenodo.3763740.

Dočekal, Mojmír & Jakub Dotlačil. 2017. *Strong NPIs vs. n-words: Acceptability experiment in Czech.* Presentation at Sinn und Bedeutung 22. Berlin. https://sinnundbedeutung22.wordpress.com/wp-content/uploads/2017/09/all_abstracts.pdf.

Dočekal, Mojmír & Jakub Dotlačil. 2016a. Experimental evidence for neg-raising in Slavic. *Linguistica* 56(1). 93. DOI: 10.4312/linguistica.56.1.93-109.

Dočekal, Mojmír & Jakub Dotlačil. 2016b. When is not-believing believing that not? In Robert Truswell, Chris Cummins, Caroline Heycock, Brian Rabern & Hannah Rohde (eds.), *Proceedings of Sinn und Bedeutung*, vol. 21, 387–398. University of Konstanz: Gesellschaft für Semantik. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/144.

Gajewski, Jon R. 2007. Neg-raising and polarity. *Linguistics and Philosophy* 30(3). 289–328. DOI: 10.1007/s10988-007-9020-z.

Gajewski, Jon R. 2008. More on quantifiers in comparative clauses. In Tova Friedman & Satoshi Ito (eds.), *Semantics and linguistic theory*, vol. 18, 340–357. New York: Linguistic Society of America. DOI: 10.3765/salt.v0i0.2494.

Gajewski, Jon R. 2011. Licensing strong NPIs. *Natural Language Semantics* 19(2). 109–148. DOI: 10.1007/s11050-010-9067-1.

Gajewski, Jon R. 2016. Another look at NPIs in definite descriptions: An experimental approach. In Pierre Larrivée & Chungmin Lee (eds.), *Negation and polarity: Experimental perspectives*, 307–327. Cham: Springer. DOI: 10.1007/978-3-319-17464-8_13.

Goodrich, Ben, Jonah Gabry, Imad Ali & Sam Brilleman. 2022. *Rstanarm: Bayesian applied regression modeling via Stan.* R package version 2.21.3. https://mc-stan.org/rstanarm/.

Herburger, Elena. 2001. The negative concord puzzle revisited. *Natural Language Semantics* 9(3). 289–333. DOI: 10.1023/A:1014205526722.

Hoeksema, Jack. 2012. On the natural history of negative polarity items. *Linguistic Analysis* 38(1). 3–33.

Homer, Vincent. 2021. Domains of polarity items. *Journal of Semantics* 38(1). 1–48. DOI: https://doi.org/10.1093/jos/ffaa006.

Ionin, Tania & Tatiana Luchkina. 2018. Focus on Russian scope: An experimental investigation of the relationship between quantifier scope, prosody, and information structure. *Linguistic Inquiry* 49(4). 741–779. DOI: 10.1162/ling_a_00288.

Jacobson, Pauline. 1995. On the quantificational force of free relatives. In Emmon Bach, Eloise Jelinek, Angelika Kratzer & Barbara Partee (eds.), *Quantification in natural languages*, 451–486. Dordrecht: Kluwer. DOI: 10.1007/978-94-017-2817-1_15.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1. 7–36. DOI: 10.1007/s40607-014-0009-9.

Krifka, Manfred. 1992. Some remarks on polarity items. In Dietmar Zaefferer (ed.), *Semantic universals and universal semantics*, 150–189. Berlin: Foris. DOI: 10.1515/9783110870527-009.

Krifka, Manfred. 1995. The semantics and pragmatics of polarity items. *Linguistic Analysis* 25(3-4). 209–257.

Kuhn, Jeremy. 2022. The dynamics of negative concord. *Linguistics and Philosophy* 45(1). 153–198. DOI: 10.1007/s10988-020-09318-3.

Ladusaw, William A. 1992. Expressing negation. In Chris Barker & David Dowty (eds.), *Proceedings of the 2nd Semantics and Linguistic Theory conference, held May 1-3, 1992 at the Ohio State University*, 237–260. Columbus: Ohio State University. DOI: 10.3765/salt.v2i0.3030.

Lahiri, Utpal. 1998. Focus and negative polarity in Hindi. *Natural Language Semantics* 6(1). 57–123. DOI: 10.1023/A:1008211808250.

Mayer, Erika, Susanne Wurmbrand & Jon Sprouse. 2018. An experimental investigation of NPI licensing under DE flip-flop. In Sherry Hucklebridge & Max Nelson (eds.), *NELS 48: Proceedings of the forty-eighth annual meeting of the North East Linguistic Society*. Amherst: GLSA.

Neeleman, Ad & Elena Titov. 2009. Focus, contrast, and stress in Russian. *Linguistic Inquiry* 40(3). 514–524. DOI: 10.1162/ling.2009.40.3.514.

Ovalle, Luis Alonso & Elena Guerzoni. 2004. Double negatives, negative concord and metalinguistic negation. In Mary Androni, Erin Debenport, Anne Pycha & Keiko Yoshimura (eds.), *Proceedings from the main session of the thirty-eighth meeting of the Chicago Linguistic Society: Volume 38-1*, 15–31. Chicago: Chicago Linguistic Society Publications.

Panizza, Daniele & Yasutada Sudo. 2020. Minimal sufficiency with covert *even*. *Glossa: a journal of general linguistics* 5(1). 98. DOI: 10.5334/gjgl.1118.

Penka, Doris. 2007. *Negative indefinites*. Tübingen: Universität Tübingen. (Doctoral dissertation). DOI: 10.1093/acprof:oso/9780199567263.001.0001.

Penka, Doris. 2016. *Degree equatives: The same as comparatives?* Presentation at the workshop *Equative Constructions*, Universität zu Köln. https://ling.sprachwiss.uni-konstanz.de/pages/home/penka/Pubs/equatives_koeln.pdf.

Romoli, Jacopo. 2013. A scalar implicature-based approach to neg-raising. *Linguistics and Philosophy* 36(4). 291–353. DOI: 10.1007/s10988-013-9136-2.

Rullmann, Hotze. 1995. *Maximality in the semantics of Wh-constructions*. Amherst: University of Massachusetts. (Doctoral dissertation).

Schwarz, Florian, Kajsa Djärv & Jérémy Zehr. 2020. Do Italian factives entail their presupposition? Yes, but… In Rajesh Bhatt, Ilaria Frana & Paula Menéndez-Benito (eds.), *Making worlds accessible: Essays in honor of Angelika Kratzer*, 150–166. Amherst: University of Massachusetts Amherst. https://scholarworks. umass.edu/ak_festsite_schrift/1/.

Seuren, Pieter A.M. 1984. The comparative revisited. *Journal of Semantics* 3(1). 109–141. DOI: 10.1093/jos/3.1-2.109.

Starschenko, Alexej & Marta Wierzba. 2023. *L-Rex Linguistic rating experiments [software], version 1.0.2*. GNU General Public License v3.0. https://github.com/ 2e2a/l-rex/.

Tottie, Gunnel. 1991. *Negation in English speech and writing: A study in variation* (Quantitative Analyses of Linguistic Structure 4). San Diego: Academic Press.

von Stechow, Arnim. 1984. Comparing semantic theories of comparison. *Journal of Semantics* 3(1-2). 1–77. DOI: 10.1093/jos/3.1-2.1.

Wickham, Hadley. 2016. *Ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. https://ggplot2.tidyverse.org.

Zeijlstra, Hedde. 2004. *Sentential negation and negative concord*. Amsterdam: University of Amsterdam. (Doctoral dissertation).

Zeijlstra, Hedde. 2022. *Negation and negative dependencies*. Oxford: Oxford Academic. DOI: 10.1093/oso/9780198833239.003.0003.

Zwarts, Frans. 1998. Three types of polarity. In Fritz Hamm & Erhard Hinrichs (eds.), *Plurality and quantification*, 177–238. Dordrecht: Springer. DOI: 10.1007/ 978-94-017-2706-8_5.

Chapter 8

The meaning of Czech response particles

Kateřina Hrdinková & Radek Šimík
Charles University

This article deals with the semantics and interpretation of Czech response particles *ano* 'yes' and *ne* 'no'. Based on two experiments involving responses to negative polar questions, we argue that *ano* 'yes' encodes the relative feature [AGREE] and *ne* 'no' encodes the absolute feature [−], adopting the parlance of Roelofsen & Farkas's (2015) feature model. This contrasts with the proposal of Gruet-Skrabalova (2016), who argues, following previous work on English, that both of the Czech response particles are ambiguous between a relative and an absolute reading. We also find some tentative evidence for context affecting the interpretation of response particles, in line with the predictions of Krifka (2013).

1 Introduction

Response particles like *yes* and *no* are a common way to respond to polar questions. They exhibit anaphoric behavior in that their interpretation crucially depends on previous context and, more specifically, on the form and interpretation of the polar question they respond to. While responses to affirmative questions are largely unproblematic, responses to negative questions give rise to ambiguities (Kramer & Rawlins 2011, Espinal & Tubau 2019, Roelofsen & Farkas 2015, Krifka 2013; etc.); see (1) and (2), respectively. (The translations in (2B) are tentative and will be rectified in view of the experimental results.)

(1) A: Zalil   Petr květiny?
         watered Petr flowers
         'Has Petr watered flowers?'

B: Ano. (= Zalil.)   / Ne. (= Nezalil.)
yes   watered   no   NEG.watered
'Yes. (= He has.) / No. (= He hasn't.)'

(2) A: Nezalil   Petr květiny?
NEG.watered Petr flowers
'Hasn't Peter watered flowers?'

B: Ano. (= Zalil   / Nezalil.)   / Ne. (= Zalil   / Nezalil.)
yes   watered NEG.watered no   watered NEG.watered
'Yes. (= He has. / He hasn't.) / No. (= He has. / He hasn't.)'

(translations tentative; to be rectified)

If the polar question is negative, as in (2), both *ano* 'yes' and *ne* 'no' can in principle correspond to a positive or a negative answer. They can, however, differ in naturalness and likelihood. To give an example from German, Claus et al. (2017) found out that it is more natural to confirm negative questions by *ja* 'yes' than by *nein* 'no'.

Using a version of the truth-value judgment task, we investigate the meaning of the two Czech response particles *ano* 'yes' and *ne* 'no', hoping to contribute to related recent literature on Slavic languages (e.g. Gruet-Skrabalova 2016, Esipova 2021, Geist & Repp 2023). A more specific goal is to evaluate the adequacy of two types of existing accounts of response particle meaning: the feature model of Roelofsen & Farkas (2015), in which response particles have a lexically specified range of meanings, and the saliency account of Krifka (2013), in which the meaning is expected to be more context-dependent. We also discuss our results in the light of Gruet-Skrabalova's (2016) analysis of Czech response particles, which is couched in the feature model. We conclude that our data primarily support a particular version of the feature model, though not the one proposed by Gruet-Skrabalova (2016). More specifically, we see a very clear tendency for *ano* 'yes' to express agreement (the feature [AGREE]) with its antecedent, be it positive or negative, and *ne* 'no' to express a negative proposition (the feature [−]), independently of the polarity of the antecedent. What counts as the "antecedent" is crucially modulated by the interrogative strategy used: negative polar questions with an interrogative syntax (verb-first) primarily contribute a positive antecedent (i.e., the negation is, by hypothesis, "pleonastic"), while negative polar questions with a declarative syntax (non-verb-first) contribute a negative antecedent (negation is semantic/propositional). Even though the feature model appears to be more suitable for modelling our results, we also observe – in a subset of our data – a statistically significant result predicted by Krifka's (2013) saliency theory.

The article is structured as follows. Section 2 briefly introduces the two approaches under consideration – the feature model (Roelofsen & Farkas 2015) and the saliency theory (Krifka 2013). We also discuss Gruet-Skrabalova's (2016) particular application of the feature model to Czech. Section 3 reports on the experiments we have conducted: experiment 1, in which we investigated responses to negative polar questions with interrogative syntax (V1), and experiment 2, in which we looked at negative polar questions with declarative syntax. In Section 4 we discuss the results and propose a new implementation of the feature model which is consistent with the results. Finally, Section 5 concludes the paper.

2 Approaches to response particle meaning

2.1 Feature model

2.1.1 Roelofsen & Farkas (2015)

The influential FEATURE MODEL of Roelofsen & Farkas (2015) is based on the assumption that a response particle like 'yes' or 'no' has a lexically specified range of meanings, defined in terms of two types of features – absolute and relative polarity features. The ABSOLUTE FEATURES [+] and [−] correspond to the polarity of the response. The relative features [AGREE] and [REVERSE] indicate a relation between the response and its propositional antecedent (derived from a polar question or an assertion that antecedes the response): the former expresses agreement with the polarity of the antecedent, the latter reverses its polarity. The semantics of the features is presuppositional (see Roelofsen & Farkas 2015: 385f. for details). In the lexicon, a response particle can either be specified for a single feature or for a combination of features. Additional complexity may arise in the process of feature realization (spellout), where Roelofsen & Farkas (2015) assume that a feature bundle can be realized by a particle which matches only its proper subset.[1]

Table 1 shows the assumed lexical entries and corresponding realization patterns of the English particles *yes* and *no* and the German particle *doch*. By hypothesis, the English particles encode single features, but are lexically ambiguous – they either encode the respective absolute or relative features. If a feature bundle is generated in the syntax (and interpreted in the semantics), it is realized by a particle whose lexical makeup matches a proper subset of that bundle. In two cases – [AGREE, −] and [REVERSE, +] – both *yes* and *no* provide a good match,

---

[1]For a recent experimental evaluation of Roelofsen & Farkas's (2015) model, see Maldonado & Culbertson (2023).

giving rise to an ambiguity which must be resolved pragmatically.[2] An example of a bundle-encoding particle is German *doch*, which responds to negative antecedents and at the same time reverses their polarity, whence [REVERSE, +].

Table 1: Feature bundles in the feature model

	Lexically encoded by	Realized by
[+]	*yes*	*yes*
[−]	*no*	*no*
[AGREE]	*yes*	*yes*
[REVERSE]	*no*	*no*
[AGREE, +]	n.a.	*yes*
[AGREE, −]	n.a.	*yes* or *no*
[REVERSE, +]	*doch*	*yes* or *no* / *doch*
[REVERSE, −]	n.a.	*no*

Thus, the English *yes* can signal that the answer has a positive polarity [+] or it can agree with its propositional antecedent [AGREE]. In contrast, *no* can serve either to signal negative polarity [−] or to reverse the polarity of its antecedent [REVERSE]. These double properties of *yes* and *no*, according to Roelofsen & Farkas (2015: 383), explain why response particles are generally clear after a positive question / positive statement (3), while a double interpretation is possible after a negative question / negative statement (4).

(3) Amy left.   (positive antecedent)
     Agreement: Yes, she did. / *No, she did.
     Reversal: *Yes, she didn't. / No, she didn't.

(4) Amy didn't leave.   (negative antecedent)
     Agreement: Yes, she didn't. / No, she didn't.
     Reversal: Yes, she did. / No, she did.

2.1.2 The feature model applied to Czech: Gruet-Skrabalova (2016)

Gruet-Skrabalova (2016) adopts Roelofsen & Farkas's (2015) feature model and adapts it to Czech. Gruet-Skrabalova assumes that Czech response particles *ano*

---

[2]Roelofsen & Farkas (2015) employ a set of additional markedness-based rules which nudge the likelihood in one or the other direction.

'yes' and *ne* 'no' exhibit the same ambiguity as the English particles *yes* and *no*, i.e., they can either realize the absolute features ([+] and [−], respectively) or the relative features ([AGREE] and [REVERSE], respectively).

Gruet-Skrabalova (2016) further modulates her analysis relative to the form of the question which antecedes the response. She assumes that in interrogative questions, i.e., questions with the verb in clause-initial position (V1), the polarity is neutralized. Response particles used in reaction to V1 questions therefore realize their absolute features. This is illustrated for the case of negative V1 questions in (5), where there is no ambiguity in the response: *ano* 'yes' indicates positive polarity and *ne* 'no' negative polarity. In declarative questions, i.e., questions with the verb in a non-initial position (non-V1), the polarity is salient and the response particles realize their relative features. As a result, response particles are also not ambiguous in this case, but have opposite truth-conditions; see (6).

(5) Negative interrogative question (with V1)

    A: Nenapsala Jitka esej?
           NEG.wrote Jitka essay
           'Hasn't Jitka written an essay?'

    B: Ano. (= Napsala.) / Ne. (= Nenapsala.)
           yes   wrote   no   NEG.wrote
           'Yes. (= She has.) / No. (= She hasn't.)'

(6) Negative declarative question (with non-V1)

    A: Jitka esej   nenapsala?
           Jitka essay NEG.wrote
           'Hasn't Jitka written an essay?'

    B: Ano. (= Nenapsala.) / Ne. (= Napsala.)
           yes   NEG.wrote   no   wrote
           'Yes. (= She hasn't.) / No. (= She has.)'

2.2 Saliency account: Krifka (2013)

Krifka's (2013) SALIENCY ACCOUNT takes an additional factor into account, namely the role of contextual and more generally pragmatic considerations, co-determining which proposition is selected as the antecedent for the response particle.

In Krifka's (2013) theory, response particles are propositional anaphors, not unlike pronouns.[3] If a response particle is preceded by a question which con-

---

[3]In this respect, Roelofsen & Farkas's (2015) and Krifka's (2013) theories are similar. Both crucially build on an analogy with pronouns – the former via pronominal-like presuppositions (not discussed here), the latter via the anaphoric potential of pronouns.

tains negation, there are in principle two possible antecedents for the response particle: either the negative proposition or the negation's prejacent, i.e., the corresponding positive proposition. This is illustrated in (7) (adapted from Krifka 2013: 14).[4]

(7) [$_{ActP}$ did REQUEST [$_{NegP}$ Ede not [$_{TP}$ t$_{Ede}$ t$_{did}$ steal the cookie]]]?
          ↳ *d*   ↳ *d′*

Response particles used in reaction to a question like (7) can thus be interpreted as in (8), capturing the ambiguity discussed above.

(8) a. *Yes.* ⇝ ASSERT($d′$) ≈ *Yes, he did!*   (rejecting accent, with clause)
     b. *Yes.* ⇝ ASSERT(d) ≈ *Yes, he didn't.*   (natural, but with clause)
     c. *No.* ⇝ ASSERT($\neg d′$) ≈ *No (he didn't).*   (natural, clause not necessary)
     d. *No.* ⇝ ASSERT($\neg d$) ≈ *Well, he did!*   (rejecting accent, with clause)

What is of interest to us is how the antecedent of the response particle is selected, i.e., whether the response particle denotes *d* (the negative proposition) or *d′* (the positive proposition). Krifka (2013) assumes that the saliency of the propositions – and hence the likelihood of their antecedent status – can be modulated contextually. Example (9) (adapted from Krifka 2013: 14) and the matching example (10) (created by us) illustrate this point. These examples differ in the question under discussion put on the table by A: in (9), the issue is negatively defined, in (10), the issue is positively defined. Although the default antecedent for the response particles in both cases will be the negative proposition asserted by B, the context is assumed to modulate the availability of the positive antecedent, which leads to an increased likelihood of the truth-conditionally opposite responses in (10).[5]

(9) Negative context (italicized)
     A: Which of the mountains on this list *did Reinhold Messner not climb*?
     B: Well, let's see... He did not climb Mount Cotopaxi in Ecuador.

---

[4]For questions with high negation (*Didn't Ede steal the cookie?*), Krifka (2013: 14) assumes only one possible antecedent, namely the positive *d′*. The negation in this case is applied outside of the scope of the proposition, making it unavailable for anaphoric pickup.

[5]An anonymous reviewer points out that the positive interpretation ('He climbed it') of response (10A$_1$) might be contingent on *yes* being pronounced with a specific intonation. This is indeed what Goodhue & Wagner (2018) confirmed experimentally; they call the fall rise intonation used in these cases "contradiction contour", following Liberman & Sag (1974).

A$_1$: Yes.
  Likely: 'He didn't climb it.'
  Unlikely: 'He climbed it.'

A$_2$: No.
  Likely: 'He climbed it.'
  Unlikely: 'He didn't climb it.'

(10) Positive context (italicized)

A: Which of the mountains on this list *did Reinhold Messner climb*?

B: Well, let's see...He did not climb Mount Cotopaxi in Ecuador.

A$_1$: Yes.
  Likely: 'He didn't climb it.'
  More likely than in (9): 'He climbed it.'

A$_2$: No.
  Likely: 'He climbed it.'
  More likely than in (9): 'He didn't climb it.'

One of our experiments (experiment 1) will tap not only into the basic meaning of Czech response particles, which can be formulated in terms of the feature model, but also into the influence of the context in determining the antecedent of the response particles.

3 Experiments

The aim of our experiments was to find out the preferred meaning of *ano* 'yes' and *ne* 'no' in response to polar questions and test the above-mentioned approaches, in particular Gruet-Skrabalova's (2016) version of the feature-based analysis and Krifka's (2013) idea that the choice of the response particle antecedent is context-dependent. The experimental design was inspired by Kramer & Rawlins (2012) and Claus et al. (2017), who investigate the meaning of response particles in English and German.

Our results suggest a relative ([AGREE]-based) semantics for *ano* 'yes' and absolute ([−]-based) semantics for *ne* 'no'. This can be easily modeled using the feature model. The particular predictions of Gruet-Skrabalova (2016) were, however, not borne out: we do not see evidence for relative ([REVERSE]-based) semantics for *ne* 'no'. In addition, in a particular corner of our data, we see a pattern which is predicted by Krifka's (2013) saliency account.

We used 2 experiments combined in a single setup, such that each of the two experiments provided fillers for the other one. This setup makes it possible to draw inferences cross-experimentally. The more complex and powerful experiment 1 uses a $2 \times 2 \times 2$ design and investigates responses to syntactically interrogative negative polar questions. Experiment 2 uses a 2×2 design and investigates responses to syntactically declarative negative polar questions.

We first describe aspects common to the two experiments (see Section 3.1) and then turn to the individual experiments (Section 3.2–Section 3.3).

3.1 Aspects common to both experiments

As detailed in Table 2, our experimental setup consisted of two experiments with 16 and 8 items, respectively, and an additional set of 16 filler items, giving a total of 40 items. The number matches the number of stimuli seen by each participant.

Table 2: Overall experimental setup

Experiment 1	16
Experiment 2	8
Fillers	16
Total	40

3.1.1 Task, procedure, and dependent variables

The participants were exposed to written stimuli which consisted of a short narrative (a few sentences) followed by a short dialogue between two people (A and B for ease of reference), in which A opens the dialogue with an assertion associated with the narrative, B asks a relevant polar question, and A responds by saying either 'yes' or 'no'. The narrative and the dialogues contained mildly colloquial elements, in order to simulate an informal setting – a dialogue between two friends. The participant's task was twofold: (i) to determine whether A's 'yes'/'no' response is consistent with the information provided in the narrative (i.e., a truth-value judgment task) and (ii) to rate the naturalness of that response given the preceding narrative and dialogue (on a scale from 1 = completely unnatural to 7 = completely natural). In this paper, we analyze the responses from task (i) and leave (ii) aside. This is mainly because there was a strong correlation between the two in the sense that responses which were judged as consistent with

the information provided were also rated as natural and conversely – responses judged as inconsistent were rated as unnatural.

3.1.2 Participants

Data from 66 adult native speakers of Czech (43 women, 23 men) entered the analyses. We used convenience sampling, recruiting participants from an extended social network of the first author. Most of the participants (44) were 18–29 years old and most (43) had university education. All participants were informed about the purpose of the experiment and all gave informed consent to participate in the experiment and the subsequent processing and anonymous publication of the collected data.

The analyzed sample consists of participants who have passed a preset quality measurement, namely scoring in the expected way on variable (i) in at least 75 % of the 8 filler items, where the relation between the information provided (in the narrative) and the 'yes'/'no' response was particularly transparent.

3.1.3 Software and administration

The experiments were prepared and administered using the L-Rex software (Starschenko & Wierzba 2023). The stimuli from each experiment were distributed on lists using the Latin Square design, so that one participant saw only a single stimulus from each item. The lists from each experiment were then combined and the order of presentation was pseudo-randomized in such a way that two stimuli from a single experiment never directly followed one another and two stimuli from a single condition were always interspersed by at least one stimulus from a different condition.

The experiment was distributed online by sending a link. Participants took part at their own personal computers and most of them needed 25–40 minutes to complete the experiment.

3.2 Experiment 1: Syntactically interrogative negative polar questions

3.2.1 Design and manipulated variables

This experiment focuses on the most complex and problematic case: responses to negative polar questions which are syntactically interrogative, which means that the finite verb is located in the clause-initial position (V1 for short); see (11). Since negation is obligatorily prefixed to the verb in Czech, there is no reliable formal

difference between high and low negation (cf. Ladd 1981) and its semantic correlate outer (extra-propositional) vs. inner (propositional) negation (AnderBois 2019, Goodhue 2022). Yet there is a general consensus that negation in V1 polar questions in Czech corresponds to high negation, which is either pleonastic (expletive) or applies at an illocutionary level (Repp 2013; for a discussion on Czech, see Staňková 2023, Staňková & Šimík forthcoming, and the references cited therein).

(11) Neprodala Jitka ty   staré boty?
      NEG.sold   Jitka DEM old   shoes
      'Didn't Jitka / Did Jitka not sell the old shoes?'

In a factorial 2 × 2 × 2 design, we manipulated three variables (all within items and subjects): INFORMATION, CONTEXT, and RESPONSE. The INFORMATION variable, with values positive (i_pos) and negative (i_neg) is manipulated in the lead-in narrative and fixes the factual state of affairs relative to which the participant judges the consistency of the response. The CONTEXT variable, likewise with values positive (c_pos) and negative (c_neg), was manipulated in the first utterance of the dialogue, which is then followed by the polar question. Finally, the RESPONSE variable has the values *ano* (yes) and *ne* (no) and is manipulated in the final utterance of the dialogue. For purposes of visualization and statistical analysis, we have found it useful to include an auxiliary variable, namely the ACCORDANCE between INFORMATION and RESPONSE, yielding the value accord for the cases where the positive information is matched by a yes response and negative information by a no response, and the value discord where this is not so. Table 3 provides an overview of all the 8 unique conditions of experiment 1.

3.2.2 Materials

Example (12) provides an example of an item (particularly item 14) in all eight conditions. The values of the INFORMATION variable is set in small caps, the CONTEXT in italics, and the RESPONSE in boldface. The parts that remained constant across the manipulations – including the negative polar question – are set in ordinary typeface. The value of the INFORMATION variable was located in the lead-in narrative, specifically in the position indicated by [...].[6]

---

[6]All the experiment materials, results, and outputs of statistical models are available at Open Science Framework under the following link: https://doi.org/10.17605/OSF.IO/9VXJS.

Table 3: Conditions in the factorial design of experiment 1

	INFORMATION	CONTEXT	RESPONSE	ACCORDANCE
a	i_pos	c_pos	yes	accord
b	i_pos	c_pos	no	discord
c	i_pos	c_neg	yes	accord
d	i_pos	c_neg	no	discord
e	i_neg	c_pos	yes	discord
f	i_neg	c_pos	no	accord
g	i_neg	c_neg	yes	discord
h	i_neg	c_neg	no	accord

(12) Eva a   Lída se   zúčastnily vánočního plesu ve svém rodném městě.
Eva and Lída REFL took.part Christmas ball in their birth   town
Lída, která se   velmi zajímá o společenský život ve svém rodišti, [...]
Lída who REFL a.lot interest in social   life in her   birthplace
Když se   po nějaké době potkají, probírají spolu   proběhlý ples.
when REFL after some time meet   discuss together passed ball
'Eva and Lída attended a Christmas ball in their hometown. Lída, who is
interested in the social life in her hometown very much, [...] When Eva
and Lída meet after a while, they discuss the ball together.'

 a. ví,   že HLAVNÍ ORGANIZÁTORKOU BYLA JEJICH BÝVALÁ
knows that main   organizer   was their former
SPOLUŽAČKA A   ZKUŠENÁ   ORGANIZÁTORKA PLESŮ ALICE.
classmate   and experienced organizer   balls Alice
'knows that the main organizer was their former classmate and an
experienced ball organizer Alice.'

Lída: *Ten ples se   jim moc povedl.*
DEM ball REFL them much worked.out
'The ball worked out really well.'

Eva: Neorganizovala ho Alice?
NEG.organized it Alice
'Didn't Alice organize it?'

Lída: **Ano.**
yes
'Yes.'

b. ví,   že   HLAVNÍ ORGANIZÁTORKOU BYLA JEJICH BÝVALÁ
knows that main   organizer   was their former
SPOLUŽAČKA A   ZKUŠENÁ   ORGANIZÁTORKA PLESŮ ALICE.
classmate   and experienced organizer   balls Alice
'knows that the main organizer was their former classmate and an
experienced ball organizer Alice.'

Lída: *Ten ples se   jim   moc   povedl.*
        DEM ball REFL them much worked.out
        'The ball worked out really well.'

Eva: Neorganizovala ho Alice?
       NEG.organized   it Alice
       'Didn't Alice organize it?'

Lída: **Ne.**
        no
        'No.'

c. ví,   že   HLAVNÍ ORGANIZÁTORKOU BYLA JEJICH BÝVALÁ
knows that main   organizer   was their former
SPOLUŽAČKA A   ZKUŠENÁ   ORGANIZÁTORKA PLESŮ ALICE.
classmate   and experienced organizer   balls Alice
'knows that the main organizer was their former classmate and an
experienced ball organizer Alice.'

Lída: *Ten ples se   jim   moc   nepovedl.*
        DEM ball REFL them much NEG.worked.out
        'The ball worked out really well.'

Eva: Neorganizovala ho Alice?
       NEG.organized   it Alice
       'Didn't Alice organize it?'

Lída: **Ano.**
        yes
        'Yes.'

d. ví,   že   HLAVNÍ ORGANIZÁTORKOU BYLA JEJICH BÝVALÁ
knows that main   organizer   was their former
SPOLUŽAČKA A   ZKUŠENÁ   ORGANIZÁTORKA PLESŮ ALICE.
classmate   and experienced organizer   balls Alice
'knows that the main organizer was their former classmate and an
experienced ball organizer Alice.'

Lída: *Ten ples se jim moc nepovedl.*
  DEM ball REFL them much NEG.worked.out
  'The ball worked out really well.'

Eva: Neorganizovala ho Alice?
  NEG.organized it Alice
  'Didn't Alice organize it?'

Lída: **Ne.**
  no
  'No.'

e. VÍ, ŽE JEJICH BÝVALÁ SPOLUŽAČKA A ZKUŠENÁ
 knows that their former classmate and experienced
 ORGANIZÁTORKA PLESŮ ALICE SE TENTOKRÁT NA ORGANIZACI
 organizer balls Alice REFL this.time in organization
 NEPODÍLELA.
 NEG.was.involved
 'knows that their former classmate and an experienced ball organizer
 Alice wasn't involved in the organization this time.'

Lída: *Ten ples se jim moc povedl.*
  DEM ball REFL them much worked.out
  'The ball worked out really well.'

Eva: Neorganizovala ho Alice?
  NEG.organized it Alice
  'Didn't Alice organize it?'

Lída: **Ano.**
  yes
  'Yes.'

f. VÍ, ŽE JEJICH BÝVALÁ SPOLUŽAČKA A ZKUŠENÁ
 knows that their former classmate and experienced
 ORGANIZÁTORKA PLESŮ ALICE SE TENTOKRÁT NA ORGANIZACI
 organizer balls Alice REFL this.time in organization
 NEPODÍLELA.
 NEG.was.involved
 'knows that their former classmate and an experienced ball organizer
 Alice wasn't involved in the organization this time.'

Lída: *Ten ples se jim moc povedl.*
  DEM ball REFL them much worked.out
  'The ball worked out really well.'

Eva: Neorganizovala ho Alice?
NEG.organized it Alice
'Didn't Alice organize it?'

Lída: **Ne.**
no
'No.'

g. VÍ, ŽE JEJICH BÝVALÁ SPOLUŽAČKA A ZKUŠENÁ
knows that their former classmate and experienced
ORGANIZÁTORKA PLESŮ ALICE SE TENTOKRÁT NA ORGANIZACI
organizer balls Alice REFL this.time in organization
NEPODÍLELA.
NEG.was.involved
'knows that their former classmate and an experienced ball organizer
Alice wasn't involved in the organization this time.'

Lída: *Ten ples se jim moc nepovedl.*
DEM ball REFL them much NEG.worked.out
'The ball worked out really well.'

Eva: Neorganizovala ho Alice?
NEG.organized it Alice
'Didn't Alice organize it?'

Lída: **Ano.**
yes
'Yes.'

h. VÍ, ŽE JEJICH BÝVALÁ SPOLUŽAČKA A ZKUŠENÁ
knows that their former classmate and experienced
ORGANIZÁTORKA PLESŮ ALICE SE TENTOKRÁT NA ORGANIZACI
organizer balls Alice REFL this.time in organization
NEPODÍLELA.
NEG.was.involved
'knows that their former classmate and an experienced ball organizer
Alice wasn't involved in the organization this time.'

Lída: *Ten ples se jim moc nepovedl.*
DEM ball REFL them much NEG.worked.out
'The ball worked out really well.'

Eva: Neorganizovala ho Alice?
NEG.organized it Alice
'Didn't Alice organize it?'

> Lída: **Ne.**
> no
> 'No.'

16 items like (12) were created, meaning that each participant was exposed to each unique condition twice (following the Latin Square distribution we used; see Section 3.1). This number – admittedly not great judging by current standards (Häussler & Juzek 2017) – resulted from a compromise between statistical power considerations and the significant cognitive load imposed by the task on the participants. The introductory narrative was always presented in stylistically neutral language and the dialogues occasionally contained colloquial expressions. The information at issue (above: whether Alice organized the ball) is known to the first dialogue participant (above: Lída), but not to the second one (above: Eva). The first participant makes a claim relevant to the information at issue (above: how the ball worked out), but does not reveal its value. The first participant's utterance stands in a particular relation to the information: it makes it more likely or less likely. The second participant asks a question about the information, followed by a response from the first participant.

Consider (12a) for illustration. In this condition it is the case that Lída (the first dialogue participant) knows that Alice organized the ball. Also, as the narrative implies, both Lída and Eva (the second dialogue participant) are aware that Alice is a good ball organizer. Lída's first utterance in the dialogue – that the ball worked out really well – implies that the ball was organized by Alice (a case of evidential bias). Eva then asks a polar question, in order to verify or falsify the implication. Lída responds *ano* 'yes'. Setting the naturalness rating aside, the participant had the option of either saying that Lída's response is consistent with the information provided (i.e., the response is true), or saying that Lída's response is not consistent with the information provided (i.e., the response is false). In the former case, we assume that the participant either interprets the response absolutely (feature [+] / positive polarity = 'Alice organized the ball.') or relatively (feature [AGREE] / agreement with the antecedent 'Alice organized the ball' – made available by the prejacent of the negative polar question asked). In the latter case, the participant interpreted Lída's response as involving the relative feature [AGREE], agreeing with the negative antecedent – 'Alice didn't organize the ball' – made available by the prejacent of the polar questions, including the negation. Whether the positive or the negative form of the antecedent is more salient (and hence whether the response is considered as true or false, provided it is interpreted relatively) is, by hypothesis (Krifka 2013), co-determined by the context – Lída's first utterance (more technically: evidential bias).

3.2.3 Results

Figure 1 shows the raw results of experiment 1, in particular of the ratings of consistency between the RESPONSE (yes vs. no) and the INFORMATION provided (i_pos vs. i_neg). Figure 2 provides the corresponding 95% confidence intervals (computed with the emmeans function of the emmeans package of R; Lenth 2024). The results are visualized – and also statistically analyzed – using the auxiliary ACCORDANCE variable. The value accord combined with the value i_pos equals the value yes, combined with the value i_neg equals no, and conversely for discord. The values of the response variable (yes vs. no) are indicated in the top left corner of each of the four panes for clarity. The reason for using ACCORDANCE rather than RESPONSE is that from the perspective of the results, the levels of the former variable form more of a natural class than the levels of the latter variable. The results are thus easier to evaluate and interpret.[7]
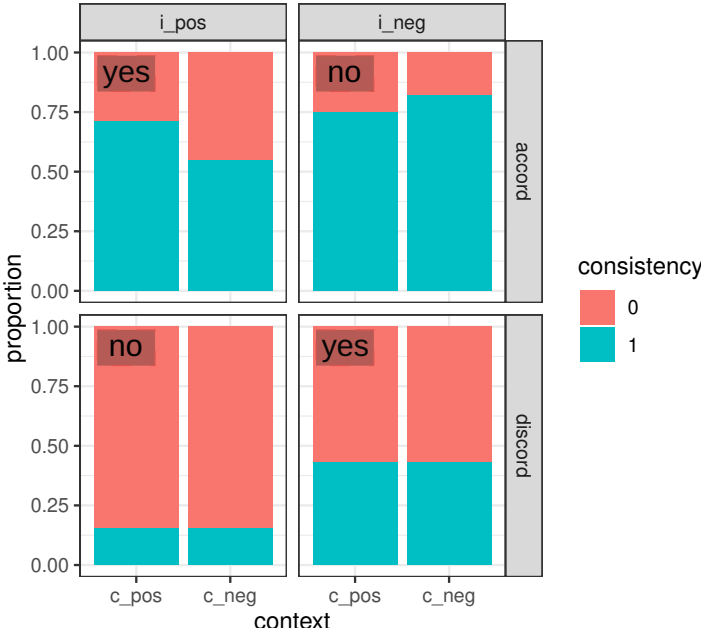
Figure 1: Experiment 1: Response–information consistency ratings

Looking at the dependent variable, we note that consistency value 1 indicates that the participant considered the response to be consistent with the information provided, or, in other words, true relative to the information provided. Consistency value 0 indicates a judgment of falsity.

---

[7]See the general discussion (Section 4) for a visualization using the RESPONSE variable.
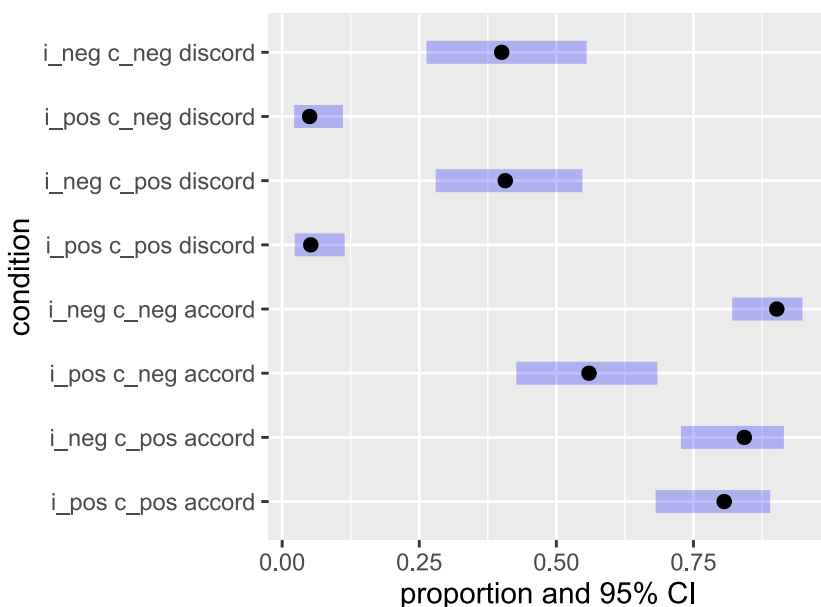
Figure 2: Experiment 1: 95% confidence intervals of consistency ratings

To give a particular example, the top right panel shows that a no response after negative information (corresponding to accord / conditions f/g of our design; see Table 3 and (12f/g)) was considered consistent with the information in about 78 % of the cases; on the other hand, the bottom left panel shows that a no response after positive information (corresponding to discord / conditions b/d) was considered consistent with the information in only about 15 % of the cases.

We fitted a generalized linear mixed model, using the glmer function of the lme4 package (Bates et al. 2015) of the R software (R Core Team 2021), to estimate the effect of INFORMATION, CONTEXT, ACCORDANCE, and their mutual interactions on the consistency rating. We included random intercepts and slopes for both items and participants; INFORMATION and CONTEXT were sum-coded, ACCORDANCE treatment-coded (using accord as the reference level).[8] The model confirms the naked-eye-visible effect of ACCORDANCE: responses which were

---

[8]The particular formula used was: CONSISTENCY RATING ~ INFORMATION * CONTEXT * ACCORDANCE + (1 + INFORMATION + CONTEXT + ACCORDANCE | participant) + (1 + INFORMATION + CONTEXT + ACCORDANCE | item). Treatment coding was used for ACCORDANCE because it has a natural reference level (accord) at which we expected high consistency (as compared to the discord level). Such a clear relationship was absent in the other factors, for which reason we applied sum coding to them.

in accord with the information (the top row in Figure 1) were rated as true much more often than responses which were in discord with the information ($z = -6.580$, $p < .001$). For instance, a positive response after positive information (see (12a)) was rated as more consistent than after negative information (see (12e)). In addition, the model revealed a significant main effect of INFORMATION ($z = -3.660$, $p < .001$; not easily interpretable), an interaction between INFORMATION and ACCORDANCE ($z = -3.849$, $p < .001$), between INFORMATION and CONTEXT ($z = 3.627$, $p < .001$), and a three-way interaction between all factors ($z = -2.399$, $p = .016$). The interaction between INFORMATION and ACCORDANCE indicates that the effect of INFORMATION is more pronounced in the discord level of ACCORDANCE. The interaction between INFORMATION and CONTEXT is only visible if the response was in accord with the information (the top pane in Figure 1), which is also indicated by the significant three-way interaction. In order to see the effect CONTEXT in a clearer way, we fitted a model onto the accord data subset, including random intercepts and slopes for both items and participants and sum-coding for both predictors – INFORMATION and CONTEXT.[9] This model confirmed the aforementioned interaction ($z = 3.584$, $p < .001$), and a further statistical analysis (nesting CONTEXT within the levels of INFORMATION) revealed that its source is both in i_pos and i_neg: if the information was positive, responses were rated as true more often if the context was also positive, as in (12a) as opposed to (12c) (simple effect of CONTEXT within i_pos; $z = 1.999$, $p = .046$), and if the information was negative, responses were rated as true more often if context was also negative, as in (12h) as opposed to (12f) (simple effect of CONTEXT within i_neg; $z = -3.389$, $p < .001$).[10]

3.2.4 Discussion

Experiment 1 clearly reveals that particle responses to negative syntactically interrogative questions are judged to be true if their polarity is in accordance with the polarity of the information provided: *ano* 'yes' is judged as true if the information is positive and *ne* 'no' is judged as true if the information is negative. This result is consistent with the absolute feature analysis, under which *ano* 'yes' encodes [+] and *ne* 'no' encodes [−]. If the negation is pleonastic, as is commonly assumed for interrogative (V1) questions (Gruet-Skrabalova 2016), then the results also follow under the relative feature analysis, under which *ano* encodes [AGREE] and *ne* [REVERSE], because in both cases the antecedent is positive.

---

[9]The formula used was: CONSISTENCY RATING ~ INFORMATION * CONTEXT + (1 + INFORMATION * CONTEXT | participant) + (1 + INFORMATION * CONTEXT | item).

[10]The formula used for the last model was: CONSISTENCY RATING ~ INFORMATION / CONTEXT + (1 + INFORMATION * CONTEXT | participant) + (1 + INFORMATION * CONTEXT | item).

That said, we should also note that there is a difference in the behavior of *ano* 'yes' and *ne* 'no'. While the effect of ACCORDANCE is very clear for *ne* (numerical difference of 63 % between accord and discord), it is much less pronounced for *ano* (numerical difference of 20 %). The ratings for *ano* 'yes' are closer to chance in both accord and discord, indicating a greater degree of uncertainty in the consistency ratings. This pattern would be expected under the conjunction of the following two premises: the negative polar question makes the negative proposition available as an antecedent (i.e., negation is not pleonastic) and *ano* 'yes' encodes [AGREE], i.e., its semantics is relative and agrees either with the positive antecedent ('yes, she did') or with a negative antecedent ('yes, she didn't'). The fact that agreement with the positive antecedent is judged as true significantly more often than agreement with the negative antecedent would then reflect on the relative availability of the two antecedent types. This explanation would further be consistent with the fact that the availability of the positive antecedent is modulated by the context (simple effect of CONTEXT), in line with Krifka (2013): if the context is positive, the positive antecedent is available more (71 %) than if the context is negative (56 %). What is unexpected is that there is no analogous simple effect of CONTEXT if *ano* 'yes' is in discord with the information provided, i.e., if the information provided is negative. In this latter case, we could expect the negative context to make the negative antecedent more accessible and hence increase the consistency judgment relative to the positive context condition. This expectation is not met: *ano* 'yes' is judged as consistent in 43 % of the cases irrespective of the value of the CONTEXT variable.

Turning to the interpretation of *ne* 'no', the overall results are consistent with the absolute semantics ([−]). What is unexpected under this view, however, is the simple effect of context in the accord condition, i.e., that *ne* 'no' is judged true in more cases if the context is negative (79 %) than if it is positive (75 %). While this effect is numerically smaller than for 'yes' responses, it is statistically stronger. A [REVERSE]-based semantics would get a handle on this effect, but would leave the very low consistency ratings in the discord condition unexplained.

3.3 Experiment 2: Syntactically declarative polar questions

3.3.1 Design and manipulated variables

This experiment focuses on responses to negative syntactically declarative polar questions, i.e., questions in which the verb is located after the subject (V2); see (13).

(13) Jitka neprodala ty   staré boty?
      Jitka NEG.sold   DEM old   shoes
      'Jitka didn't sell the old shoes?'

We only used two crossed factors in this experiment – INFORMATION and RE-
SPONSE. Context was always negative because negative declarative questions
only sound natural in contexts indicating negative evidential bias (Gunlogson
2002, Staňková 2023) and our primary interest was the interpretation (not so
much naturalness) of response particles. The overview of the individual condi-
tions is provided in Table 4. The materials were parallel to those in experiment
1; we do not include a token set here in the interest of space.

Table 4: Conditions in the factorial design of experiment 2

	INFORMATION	RESPONSE	ACCORDANCE
a	i_pos	yes	accord
b	i_pos	no	discord
c	i_neg	yes	discord
d	i_neg	no	accord

3.3.2 Results

Figure 3 shows the raw results of experiment 2, in particular the ratings of con-
sistency between the RESPONSE (yes vs. no) and the INFORMATION provided (i_pos
vs. i_neg). The values of the auxiliary ACCORDANCE variable are provided as labels,
for completeness. Figure 4 provides the corresponding 95% confidence intervals.
We fitted a generalized linear mixed model to estimate the effect of INFORMA-
TION, RESPONSE, and their interaction on the consistency rating. Both factors
were sum-coded. Random intercepts and slopes for both items and participants
were included.[11] The model confirms the visually clear main effect of INFORMA-
TION: responses are judged as more consistent with negative information (i_neg)
than with positive information (i_pos) ($z = -9.191$, $p < .001$). Furthermore, there
is an interaction between INFORMATION and RESPONSE: the effect of INFORMA-
TION is more pronounced with *ne* than with *ano* ($z = 4.091$, $p < .001$).

---

[11]The formula used was: CONSISTENCY RATING ~ INFORMATION * RESPONSE + (1 + INFORMATION
+ RESPONSE | participant) + (1 + INFORMATION + RESPONSE | item).
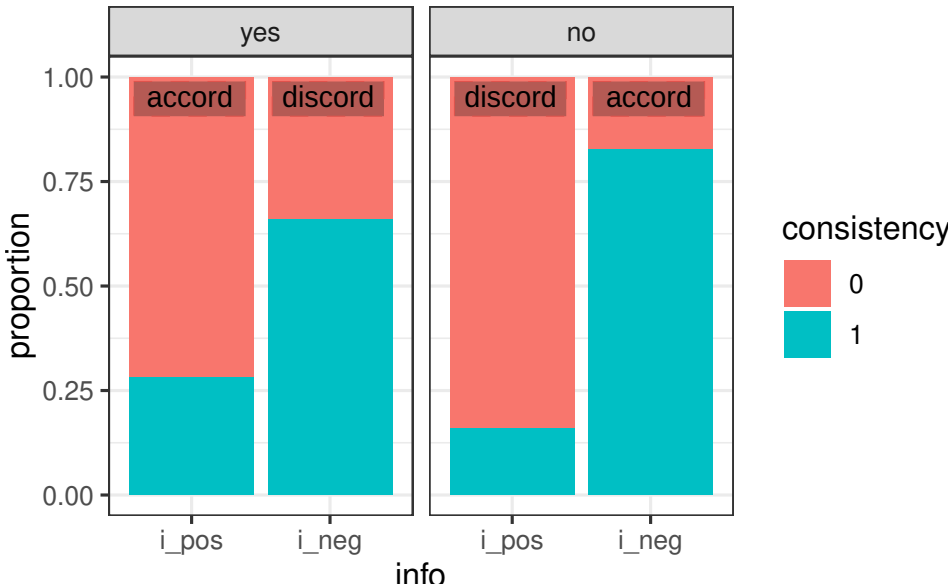
Figure 3: Experiment 2: Response–information consistency ratings
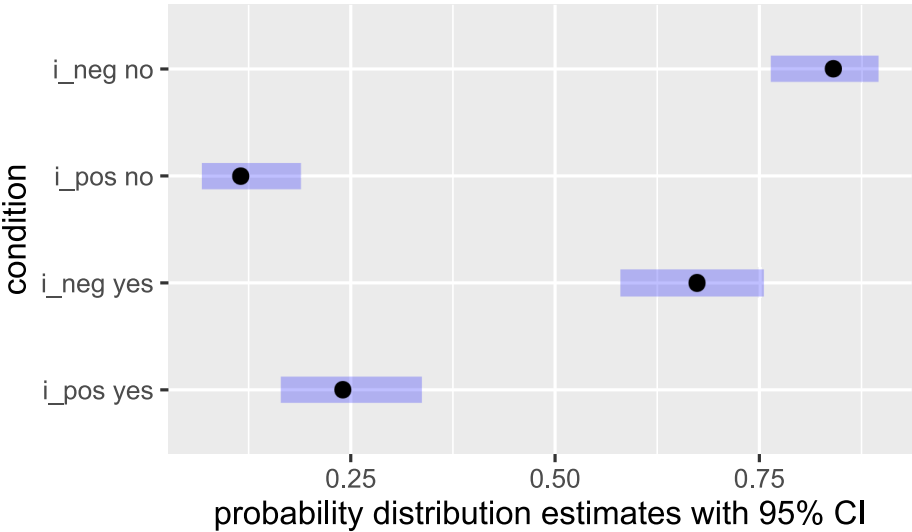
Figure 4: Experiment 2: 95% confidence intervals of consistency ratings

3.3.3 Discussion

The most significant result of experiment 2 is that both *ano* 'yes' and *ne* 'no' have the same truth conditions when responding to negative declarative polar questions: they are both judged as true if the information provided is negative (main effect of INFORMATION). This result follows from the premise that (i) *ano* 'yes' has relative semantics (encodes [AGREE]) and agrees with the negative antecedent ('yes, she didn't'), and, (ii) *ne* 'no' has absolute semantics (encodes [−]) and negative polarity ('(no,) she didn't').

The fact that the effect of INFORMATION is stronger for negative than for positive responses is consistent with this view. The relative semantics of *ano* 'yes' leaves some room for uncertainty as to which antecedent functions as the particle's prejacent. While the declarative form of the negative question makes the negative antecedent highly salient (making the 'yes, she didn't' interpretation true in 66 % of the cases in the i_neg/discord condition and 72 % in the i_pos/accord condition), the positive antecedent can also be accessed, at least when compared to the corresponding interpretations in the no condition (the 'yes, she did' interpretation is judged as true in 28 % of the cases in the i_pos/accord condition and 34 % in i_neg/discord condition). Compared to that, the hypothetical 'no, she did' interpretation is rather exceptional (only evident in about 17 % of the no responses overall).

4 General discussion

The results of our two experiments lend solid support to (i) relative, [AGREE]-based semantics of *ano* 'yes' and (ii) absolute, [−]-based semantics of *ne* 'no'. The proposed lexical encoding and the corresponding realization possibilities are represented in Table 5, an updated version of Table 1. The Czech particles *ano* 'yes' and *ne* 'no' are framed for clarity. The last column indicates where the evidence for the realization (im)possibilities stems from and whether the evidence is positive (judgment of truth) or negative (judgment of falsity).

In order to aid the discussion visually, we insert Figure 5, which includes data from both experiments: the top pane visualizes results of experiment 1 (aggregating over both levels of the CONTEXT variable), in which the question was interrogative (verb-first), and the bottom pane visualizes the results of experiment 2, in which the question was declarative (non-verb-first). For ease of reference, we label the individual stacked bars with capital letters.

Let us go through Table 5 step-by-step. If *ano* encoded [+] alone, we would expect the yes response in experiment 2 to be judged as consistent with the

Table 5: Feature bundles in the feature model (updated)

	Lexically encoded by	Realized by	(Positive/ Negative) evidence from
[+]	*yes*	*yes*, $\boxed{ano}$	exp 2 (neg)
[−]	*no*, $\boxed{ne}$	*no*, $\boxed{ne}$	exp 1, 2 (pos)
[AGREE]	*yes*, $\boxed{ano}$	*yes*, $\boxed{ano}$	exp 1, 2 (pos)
[REVERSE]	*no*	*no*, $\boxed{ne}$	exp 1, 2 (neg)
[AGREE, +]	n.a.	*yes*, $\boxed{ano}$	exp 1 (pos)
[AGREE, −]	n.a.	*yes* or *no*, $\boxed{ano}$ or $\boxed{ne}$	exp 1, 2 (pos)
[REVERSE, +]	*doch*	*yes* or *no*/*doch*, $\boxed{ano}$ / $\boxed{ne}$	exp 1, 2 (neg)
[REVERSE, −]	n.a.	*no*, $\boxed{ne}$	exp 1 (pos)
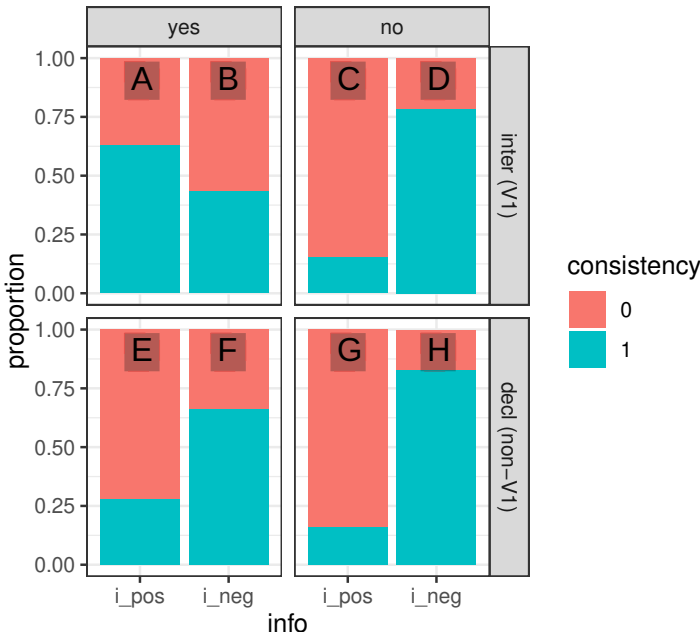
Figure 5: Both experiments: Response–information consistency ratings

positive information (bar E in Figure 5). The fact that it cannot "ignore" the negative antecedent (i.e., agrees with it; bar F), strongly supports its relative (rather than absolute) semantics.

The assumption that *ne* encodes the absolute feature [−] is supported by the stability of its consistency with the negative information, independently of the question type preceding it (bars D and H).

That *ano* encodes [AGREE] is witnessed primarily by the differential behavior of this particle in experiment 1 and experiment 2. In the former, *ano* is judged more consistent with positive information (bar A, vs. B), and in the latter, *ano* is judged more consistent with negative information (bar F, vs. E). This follows if *ano* agrees with its antecedent and if interrogative (V1) questions make the positive antecedent more salient (cf. pleonastic negation), while declarative (non-V1) questions make the negative antecedent more salient. The fact that the effect of the INFORMATION variable is less pronounced in the yes condition, as compared to the no condition, is – or so we hypothesize – also consistent with the relative vs. absolute semantics of *ano* vs. *ne*, respectively. While the absolute semantics of *ne* remains largely insensitive to the polarity of its antecedent (being sensitive merely to the polarity-free prejacent), the relative semantics of *ano* leaves room for pragmatic and contextual considerations as to which antecedent – whether positive or negative – is selected as the prejacent of *ano*, which in turn leads to a greater variance in the consistency judgments and their overall centering around chance. This is especially evident in the results of experiment 1 (see A vs. B), where we also observed the effect of the CONTEXT variable predicted by Krifka (2013): positive context (as compared to negative context) supports the selection of a positive prejacent (see Section 3.2.3).

The fact that *ne* does not encode [REVERSE] is supported by the results of both experiments, but especially of experiment 2: while *ano* switched its truth conditions between experiment 1 and experiment 2, the truth conditions of *ne* remain stable. This clearly indicates that the differential availability of the two polar antecedents in these experiments had no effect on the meaning of *ne*, militating against its relative semantics. What supports the relative semantics, and is unexpected under our analysis, is that the consistency of the *ne* response with the negative information (bar D) is modulated by context: the consistency is higher if the context is negative – a mirror image of what happens in A. This effect is numerically small (only about 4 %), but statistically significant.

Let us now turn to the realization of the four logically possible feature combinations. The [AGREE, +] bundle is realized by *ano*, which, by the subset principle employed in the feature model, spells out [AGREE] (leaving [+] unrealized). This

case is instantiated by bar A, where agreement is with a positive antecedent (supported by the tendentially pleonastic nature of the negation) and where the polarity of the response is, accordingly, positive ('yes, she did'). The [REVERSE, −] bundle is realized by *ne*, which spells out the subset [−]. This case is instantiated by bar D, where the polarity of the response is negative ('she didn't') and is reversed as compared to the primarily positive polarity of the antecedent. The [AGREE, −] bundle can in principle be realized in two ways – either by *ano,* which spells out [AGREE], or by *ne,* which spells out [−]. That precisely this is the case is witnessed by the identical truth conditions of the two particles in experiment 2 or, more specifically, by the analogous consistency ratings in bars F and H. In this case, the response agrees with the negative antecedent (≈ *ano*) and thus conveys a response of negative polarity (≈ *ne*). The most problematic case is represented by the bundle [REVERSE, +], which finds no suitable match in the lexical meanings of *ano* or *ne.* This scenario is represented by bars E and G and, as is evident from the consistency ratings, neither *ano* nor *ne* are capable of reliably conveying it. It follows that a response with positive polarity reacting to a clearly negative antecedent (contributed in experiment 2 by the negative declarative question) cannot be expressed by a standalone particle in Czech. Instead, a more complex structure is warranted, such as a fragment (elliptical) response containing a verb explicitly specified for polarity (Gruet-Skrabalova 2015, 2016) or the positive particle *ano* 'yes' preceded by *ale* 'but'; see (14). We hypothesize that the particle *ale* 'but' reverses the salience of the two polar alternatives, making the positive one, which is otherwise only latently present, more salient and hence available as an antecedent of the relative particle *ano* 'yes'.

(14) A: Jitka neprodala ty staré boty?
Jitka NEG.sold DEM old shoes
'Did Jitka not sell the old shoes?'

B$_1$: #Ano. / #Ne.
yes no
Intended: 'She sold the old shoes.'

B$_2$: Prodala.
sold
'She did.'

B$_3$: Ale ano.
but yes
'She did.'

This latter point brings us to a discussion of the predictions made by Gruet-Skrabalova (2016). Gruet-Skrabalova proposes that both of the Czech response particles can realize both the relative features and the absolute features. Our experimental results do not lend support to this claim. More particularly, we see only little evidence for [REVERSE] being realized by *ne* 'no' or for [+] being realized by *ano* 'yes'. The infelicity of the response (14B₁), which reflects our experimental results, is an example of this.

Our experimental results and the analysis we offer bear implications for the interpretation of Czech negative polar questions. Negative polar interrogatives (V1) are often considered to contain pleonastic negation, i.e., a negation which does not contribute propositional negation (see Staňková 2023 and the references therein). Such questions can thus be expected to only contribute positive propositions as antecedents available for anaphoric pick-up by the relative particle *ano* 'yes'. Counter to this expectation, we see that the negative proposition is not completely unavailable. In experiment 1, *ano* 'yes' is considered to be consistent with negative information in 43 % of the cases (bar B), a proportion which is hardly negligible (esp. when compared to the no+i_pos condition; see bar C). We take this to indicate that negation attached to a fronted verb in polar interrogatives is not necessarily pleonastic; it can either be marginally read as propositional negation or contributes an illocutionary negation (called FALSUM by Repp 2013), which can (marginally) participate in forming an antecedent – possibly a speech act – which can in turn function as the prejacent of *ano* 'yes'. While a more detailed investigation of the interactions between the semantics of negative polar interrogatives and the semantics of polar responses is still missing, the experimental results reported in Staňková (2023) are consistent with the view just suggested.

Negative declarative questions, on the other hand, primarily contribute a negative antecedent, witnessed by the high consistency of *ano* 'yes' with negative information in experiment 2 (bar F). While this is, *prima facie*, an expected result, we also know from Staňková's (2023) results that negative declarative questions readily contribute not only inner negation (licensing negative concord items) but also outer ("pleonastic") negation (compatible with positive polarity items). If this is the case, we would expect the positive proposition to be more readily available for anaphoric pickup by *ano* 'yes'. Yet this is only possible in 28 % of the cases (bar E). Admittedly, however, the salience of the positive proposition is reduced by two factors in our experimental design (of experiment 2): the absence of any polarity item indicating outer negation and the contextual negative evidence (bias). It is an open issue whether the manipulation of these factors would have an impact on the availability of the positive interpretation of the particle

ano (matched by an increased consistency in what would correspond to the E bar).

5 Conclusion

Our paper contributes the first experimental data pertaining to the semantics and interpretation of the two Czech polar response particles – *ano* 'yes' and *ne* 'no'. Building on the feature model of Roelofsen & Farkas (2015) and based on the results of our two experiments, we have argued that *ano* 'yes' lexically encodes the relative feature [AGREE] and *ne* 'no' encodes the absolute feature [−]. This stands in contrast to what has been proposed for Czech by Gruet-Skrabalova (2016) or for English by Roelofsen & Farkas (2015), namely that response particles are ambiguous between the relative and the absolute meaning. In addition, the results of experiment 1 reveal tentative evidence in favor of Krifka's (2013) proposal that context can affect the choice of the antecedent for relative response particles. More specifically, we saw that the relative particle *ano* 'yes' is resolved to a positive antecedent more often in cases in which it is preceded by a positive context, as compared to a negative context. What is puzzling is that an inverse effect is observed for the particle *ne* 'no', which otherwise exhibits a pattern consistent with absolute lexical semantics (which in turn should be insensitive to contextual manipulations). The effect is numerically much smaller, but statistically stronger.

Finally, we have drawn some implications for the semantics of polar questions. Counter to the common assumption that negation on the fronted verb in interrogative questions is pleonastic (e.g., Gruet-Skrabalova 2016), i.e. not interpreted, we have seen some tentative evidence for the availability of a negative structure being contributed by such questions. Whether it is a negative proposition or a negative speech act (as assumed e.g. by Staňková 2023) remains an open question. Likewise, it remains open how negative declarative questions (non-V1) in which negation is interpreted as outer negation (Staňková 2023) are responded to. The prediction is that the positive interpretation of *ano* 'yes' should be available to a greater extent in these cases.

Abbreviations

DEM	demonstrative	REFL	reflexive
NEG	negation		

Kateřina Hrdinková & Radek Šimík

Acknowledgments

References

AnderBois, Scott. 2019. Negation, alternatives, and negative polar questions in American English. In Klaus von Heusinger, Edgar Onea & Malte Zimmermann (eds.), *Questions in discourse – Volume 1: Semantics* (Current Research in the Semantics / Pragmatics Interface 35), 118–171. Leiden: Brill. DOI: 10.1163/9789004378308_004.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.

Claus, Berry, A. Marlijn Meijer, Sophie Repp & Manfred Krifka. 2017. Puzzling response particles: An experimental study on the German answering system. *Semantics and Pragmatics* 10(19). 1–51. DOI: 10.3765/sp.10.19.

Esipova, Masha. 2021. Polar responses in Russian across modalities and across interfaces. *Journal of Slavic Linguistics* 29(FASL 28 issue). 1–11. DOI: 10.1353/jsl.2021.a923057.

Espinal, M. Teresa & Susagna Tubau. 2019. Response systems: The syntax and semantics of fragment answers and response particles. *Annual Review of Linguistics* 5. 261–287. DOI: 10.1146/annurev-linguistics-011718-012613.

Geist, Ljudmila & Sophie Repp. 2023. Responding to negative biased questions in Russian. In Petr Biskup, Marcel Börner, Olav Mueller-Reichau & Iuliia Shcherbina (eds.), *Advances in formal Slavic linguistics 2021* (Open Slavic Linguistics 8), 145–177. Berlin: Language Science Press. DOI: 10.5281/zenodo.10123641.

Goodhue, Daniel. 2022. Isn't there more than one way to bias a polar question? *Natural Language Semantics* 30(4). 379–413. DOI: 10.1007/s11050-022-09198-2.

Goodhue, Daniel & Michael Wagner. 2018. Intonation, *yes* and *no. Glossa: a journal of general linguistics* 3(1). 5. 1–45. DOI: 10.5334/gjgl.210.

Gruet-Skrabalova, Hana. 2015. Verbs and particles in minimal answers to *yes-no* questions in Czech. In Gerhild Zybatow, Petr Biskup, Marcel Guhl, Claudia Hurtig, Olav Mueller-Reichau & Maria Yastrebova (eds.), *Slavic grammar from a formal perspective: The 10th anniversary FDSL conference*, 197–215. Frankfurt am Main: Peter Lang.

Gruet-Skrabalova, Hana. 2016. Yes or no, or how to answer a negative question. *Linguistica* 56(1). 127–142. DOI: 10.4312/linguistica.56.1.127-142.

Gunlogson, Christine. 2002. Declarative questions. In Brendan Jackson (ed.), *SALT 12: Proceedings from the 12th Conference on Semantics and Linguistic Theory*, 124–143. Ithaca: CLC Publications. DOI: 10.3765/salt.v12i0.2860.

Häussler, Jana & Tom Juzek. 2017. Hot topics surrounding acceptability judgment tasks. In Sam Featherston, Robin Hörnig, Reinhild Steinberg, Birgit Umbreit & Jennifer Wallis (eds.), *Proceedings of Linguistic Evidence 2016: Empirical, theoretical, and computational perspectives.* Tübingen: University of Tübingen. DOI: 10.15496/publikation-19039.

Kramer, Ruth & Kyle Rawlins. 2011. Polarity particles: An ellipsis account. In Suzi Lima, Kevin Mullin & Brian W. Smith (eds.), *NELS 39: Proceedings of the Thirty-Ninth Annual Meeting of the North East Linguistic Society*, vol. 2, 479–492. Amherst: GLSA Publications.

Kramer, Ruth & Kyle Rawlins. 2012. An ellipsis approach to answer particles in positive and negative contexts. Presented in the Workshop on the Syntax of Answers to Polar Questions (Newcastle University).

Krifka, Manfred. 2013. Response particles as propositional anaphors. In Todd Snider (ed.), *SALT 23: Proceedings from the 23rd conference on Semantics and Linguistic Theory*, 1–18. Ithaca: CLC Publications. DOI: 10.3765/salt.v23i0.2676.

Ladd, D. Robert. 1981. A first look at the semantics and pragmatics of negative questions and tag questions. In Roberta Hendrick, Carrie Masek & Mary Frances Miller (eds.), *CLS 17: Papers from the Seventeenth Regional Meeting of the Chicago Linguistics Society*, 164–171. Chicago: Chicago Linguistic Society.

Lenth, Russell V. 2024. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.10.2.090002. https://rvlenth.github.io/emmeans/.

Liberman, Mark & Ivan Sag. 1974. Prosodic form and discourse function. In *CLS 10: Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*, 416–427. Chicago: Chicago Linguistic Society.

Maldonado, Mora & Jennifer Culbertson. 2023. *You say yes, I say no*: Investigating the link between meaning and form in response particles. *Glossa: a journal of general linguistics* 8(1). 1–41. DOI: 10.16995/glossa.9185.

R Core Team. 2021. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.r-project.org.

Repp, Sophie. 2013. Common ground management: Modal particles, illocutionary negation and *verum*. In Daniel Gutzmann & Hans-Martin Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning*, 231–274. Brill. DOI: 10.1163/9789004183988_008.

Roelofsen, Floris & Donka F. Farkas. 2015. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language* 91(2). 359–414. https://www.jstor.org/stable/24672234.

Staňková, Anna. 2023. *The expression of speaker's bias in Czech polar questions*. Praha: Charles University. (MA thesis).

Staňková, Anna & Radek Šimík. Forthcoming. Negation in Czech polar questions. *Journal of Slavic Linguistics*.

Starschenko, Alexej & Marta Wierzba. 2023. L-Rex Linguistic rating experiments [software], version 1.0.2. GNU General Public License v3.0. https://github.com/2e2a/l-rex/.

Chapter 9

ABA in Russian adjectives, subextraction, and Nanosyntax

Daniar Kasenov

HSE University / Lomonosov Moscow State University

This paper's core focus is the ABA pattern exhibited by a number of Russian adjectives in their degree paradigms (positive, comparative, superlative). While the surface pattern seems to be a counterexample to the *ABA generalization about adjectival degree paradigms stated in Bobaljik (2012), a more involved exploration of Russian adjectival morphology shows that there are more classes of Russian adjectives that are problematic for the contemporary syntactic approaches to morphology (DM, Nanosyntax) given Bobaljik's containment hypothesis. This paper provides a description of these patterns in Russian adjectival morphology and provides an analysis for all the problematic classes in the framework of Nanosyntax, making use of two recent technical developments in the theory (Movement-Containing Trees of Blix 2022 and the subextraction Spell-Out algorithm of Caha & Taraldsen Medová 2022, 2023).

1 Introduction

This paper is concerned with an apparent counterexample to Bobaljik's ABA generalization in the domain of degree morphology (Bobaljik 2012). According to Bobaljik's cross-linguistic study, there is no language that has an adjective which has a suppletive stem α in positive and superlative and a suppletive stem β in the comparative. To give an example, a logically possible language English′, in which the adjective 'bad' has the forms *bad* 'bad.POS', *worse* 'bad.CMPR', *baddest* 'bad.SPRL', is impossible according to the generalization that Bobaljik draws from his typological study. The attested and unattested patterns of Bobaljik's three-cell paradigm are summarized in Table 1.

Table 1: (Un)attested suppletion patterns in adjectival paradigms

	POS	CMPR	SPRL	
AAA	*pretty*	*pretti-er*	*pretti-est*	English
ABB	*bad*	*worse*	*worst*	English
ABC	*bon-us*	*mel-ior*	*optim-us*	Latin
AAB	*bad*	*badd-er*	*worst*	Unattested!
ABA	*bad*	*worse*	*badd-est*	Unattested!

More generally, *ABA phenomena (such as the suppletion patterns in adjectives) are

> morphological patterns in which, given some arrangement of the relevant forms in a structured sequence, the first and third [forms] may share some property "A" only if the middle member shares that property as well. If the middle member is distinct from the first, then the third member of the sequence must also be distinct. (Bobaljik & Sauerland 2018: 1–2)

This work follows the tradition of examining *ABA phenomena (patterns of suppletion, syncretism, and other morphological properties) through the lens of the theories of morphology which assume the Single Engine Hypothesis (Marantz 2001), the idea that all complex expressions (including words) are built by syntax. Among others, these theories include the realizational approaches of Distributed Morphology (DM, Halle & Marantz 1994 and subsequent work) and Nanosyntax (Starke 2009 and subsequent work). In such approaches, *ABA phenomena are often understood structurally: cases of *ABA are due to the complex internal structure of examined wordforms, in which one form contains the other. Such analyses have been proposed for adjectival suppletion (Bobaljik 2012), case syncretism (Caha 2009), reflexive pronominal paradigms (Middleton 2021), numeral morphology (Sudo & Nevins 2022), and many other phenomena. Bobaljik himself has accounted for the ban on ABA via the containment structure provided in Figure 1.

Note, however, that containment structures do not rule out an ABA pattern by themselves. Such structures rule them out in conjunction with the widespread conception of (morphologically-conditioned) allomorphy phenomena in Distributed Morphology (the framework of choice in Bobaljik 2012) as contextual allomorphy (see Bonet & Harbour 2012 and Gouskova & Bobaljik 2020 for an

```
                        SPRLP
                       /     \
               CMPRP          SPRL
              /     \
          ADJ        CMPR
```
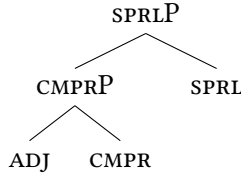
Figure 1: Containment structure for degree morphology

overview). The core logic is as follows: in DM, morphological forms are the results of Vocabulary Insertion rules which map syntactic objects onto morpho-phonological strings. Assuming that, allomorphy is understood as the same syntactic object being referenced by several mapping rules which differ by the contexts of their application. For example, the two allomorphs of the root of the English adjective *bad* are the results of two distinct insertion rules, which differ by their contexts (the *worse*-rule applied in the context of a CMPR node), as shown by the Vocabulary Insertion rules in (1).

(1)　Vocabulary Insertion rules for $\sqrt{\text{BAD}}$
　　a.　$\sqrt{\text{BAD}} \leftrightarrow$ *worse* /__] CMPR]
　　b.　$\sqrt{\text{BAD}} \leftrightarrow$ *bad*

In the absence of a specified rule for superlative forms, the containment structure ensures that the rule that applies in the context of the CMPR node also applies in the superlative form. Thus, the only allowed way for the ABA pattern to arise is accidental homophony: there should be two distinct insertion rules which just so happen to have the same morpho-phonological string as the result (like the Vocabulary Insertion rules in 2). Accidental homophonies are thought to be a rare occasion and definitely unlikely to hold across many "lexemes" (see Bobaljik & Sauerland 2018 for further discussion).

(2)　Vocabulary Insertion rules for an ABA pattern *bad-worse-baddest*
　　a.　$\sqrt{\text{BAD}} \leftrightarrow$ *bad* /__] CMPR] SPRL]
　　b.　$\sqrt{\text{BAD}} \leftrightarrow$ *worse* /__] CMPR]
　　c.　$\sqrt{\text{BAD}} \leftrightarrow$ *bad*

Given this background, the core data in this paper comes from a certain class of Russian adjectives, which exhibits an ABA pattern with respect to the presence of the augment affix *-(o)k*.[1] An example of such an adjective and its paradigm

---

[1] I follow Vanden Wyngaerd et al. (2020) in using the term AUGMENT for this sort of adjectival affix in Slavic languages.

is given in Table 2. Note that the alternation between *-(o)k* and *-(o)č* variants of the augment affix is due to morpho-phonological processes of palatalization. Without going into the depths of this phenomenon (see Blumenfeld 2003 and see Halle 1959 for a thorough treatment), it is relevant for our purposes that certain affixes turn segments /k g t s/ to /č ž č š/, respectively.

Table 2: *ABA-violating paradigm of Russian adjective *vysokij* 'high'

POS	CMPR	SPRL
vys-ok-ij	*vyš-e*	*vys-oč-aj-š-ij*
high-AUG-AGR	high-AGR	high-AUG-CMPR-SPRL-AGR
'high'	'higher'	'highest'

The problem is clear: Bobaljik's *ABA generalization rules out such patterns and yet they are found in these adjectives. While one could consider the observed surface pattern to be a reason to abandon Bobaljik's structure, this paper aims to provide an analysis of Russian *ABA-violating adjectives that does not abandon the containment structure for degree morphology.

This paper's goals are twofold. The first goal is to provide a thorough examination of Russian adjectival morphology and to pinpoint the problems it poses for contemporary generative approaches to morphology. The second goal is to resolve said problems following existing Nanosyntax work on degree morphology (Caha et al. 2019, Vanden Wyngaerd et al. 2020, Caha & Taraldsen Medová 2023). The technical solution will be based upon two novel ideas in the Nanosyntax literature: the movement-containing trees (MCTs) of Blix (2022) and the subextraction spell-out algorithm of Caha & Taraldsen Medová (2022, 2023).

The paper is structured as follows. Section 2 examines the data and argues that there are three distinct classes of Russian adjectives whose degree paradigms are problematic for a Bobaljik-style approach. Section 3 introduces the theoretical framework of the analysis to come, namely, Nanosyntax. Section 4 presents my own solution to the puzzles posed by Russian adjectival morphology while introducing unfamiliar technical elements (Movement-Containing Trees and the subextraction spell-out algorithm) and showing which parts of the data require them. Section 5 concludes the paper.

2 The landscape of Russian adjectival morphology

This section presents the main patterns found in Russian adjectival morphology and discusses the existing allomorphs of the comparative and the superlative affixes, the periphrastic forms and three classes of adjectives that are problematic from the viewpoint of a simple containment structure for the degree morphology (Bobaljik 2012).

2.1 Basic adjectival morphology of Russian

Let us consider a basic Russian adjective with a basic degree paradigm. The adjective *glupyj* 'dumb' is an exemplar. As shown in Table 3, Russian degree morphology shows a straightforward containment of the comparative form *glup-ej-* in the superlative *glup-ej-š-*, once we consider the affix *-e* of the comparative form to be a φ-deficient agreement affix (or something else but crucially something irrelevant to the degree morphology).

Table 3: The basic degree paradigm of the adjective *glupyj* 'dumb'

POS	CMPR	SPRL
glup-yj	*glup-ej-e*	*glup-ej-š-ij*
dumb-AGR	dumb-CMPR-AGR	dumb-CMPR-SPRL-AGR
'dumb'	'dumber'	'dumbest'

The paradigm in Table 3 shows that the comparative affix is *-ej-* while the superlative affix is *-š-*. From here, I will refer to them as such, even if they arise in a form other than comparative or superlative, respectively. Most Russian adjectives have a similar paradigm, some of which are given in Table 4.

However, the presented synthetic paradigm is not the only way of forming Russian comparatives and superlatives and, for the sake of completeness of the overview of Russian adjectival morphology, I should introduce the other morphological strategies as well. The first thing to mention are the analytic forms *bolee*+ADJ and *samyj*+ADJ, as shown in Table 5 for the adjective *glupyj* 'dumb'. Since these are outside the scope of this paper, I refer the reader to Matushansky (2002) for discussion of analytic comparatives and Goncharov (2015) for discussion of analytic superlatives. However, I will make an important observation that some speakers of Russian outright reject synthetic superlative forms and tend to prefer the analytic form across the board. The interaction between analytic and

Table 4: More regular adjectives

POS	CMPR	SPRL
Paradigm for *umnyj* 'smart'		
umn-yj	*umn-ej-e*	*umn-ej-š-ij*
smart-AGR	smart-CMPR-AGR	smart-CMPR-SPRL-AGR
'smart'	'smarter'	'smartest'
Paradigm for *važnyj* 'important'		
važn-yj	*važn-ej-e*	*važn-ej-š-ij*
important-AGR	important-CMPR-AGR	important-CMPR-SPRL-AGR
'important'	'more important'	'most important'
Paradigm for *krasivyj* 'pretty'		
krasiv-yj	*krasiv-ej-e*	*krasiv-ej-š-ij*
pretty-AGR	pretty-CMPR-AGR	pretty-CMPR-SPRL-AGR
'pretty'	'prettier'	'prettiest'

synthetic comparatives is more intricate and appears to be conditioned by many factors, including syllabic length of the adjective (see Kosheleva 2016 for discussion). Given this preference, some speakers may find the forms presented later in the text to be dubious. I put this difference in idiolects aside and leave them for a further sociolinguistic exploration.

Table 5: The analytic paradigm of the adjective *glupyj* 'dumb'

POS	CMPR	SPRL
glup-yj	*bolee glup-yj*	*samyj glup-yj*
dumb-AGR	more dumb-AGR	most dumb-AGR
'dumb'	'dumber'	'dumbest'

Another thing to note are the *nai*-superlatives, which consist of the prefix *nai*- and the synthetic superlative form, exemplified for the adjective *glupyj* 'dumb' in Table 6. These superlatives seem to be in free variation with regular synthetic superlatives, but some speakers consider them a "more marked" form conveying a focus on the degree. As far as I am aware, *nai*-superlatives never present suppletion/allomorphy patterns distinct from synthetic comparatives, hence they will not be discussed in this paper in detail.

Table 6: Russian *nai*-superlatives

POS	CMPR	SPRL
glup-yj	*glup-ej-š-ij*	*nai-glup-ej-š-ij*
dumb-AGR	dumb-CMPR-SPRL-AGR	*nai*-dumb-CMPR-SPRL-AGR
'dumb'	'dumber'	'dumbest'

To round up this short section, for our purposes it is important that, in the regular case, the Russian comparative affix is *-ej-*, the Russian superlative affix is *-š-*, and the superlative form contains the comparative affix – while the comparative affix is *-ej*, the superlative is ADJ-*ej*-*š*-AGR (with *-ij* 'AGR.M.SG' being the concord affix used throughout the paper). These observations are perfectly in line with the theory of adjectival degree morphology laid out in Bobaljik (2012), according to which the superlative form is built on top of the comparative form (the containment hypothesis), as discussed in the introduction. In light of the accordance of the data of basic Russian adjectives with Bobaljik's theory, the next subsection is devoted to showing the adjectives which deviate from the basic pattern of comparatives being formed with *-ej-* and superlatives being formed with *-š-* on top of the comparative form.

2.2 Three problematic classes of adjectives

Exemplars of the three puzzling classes are the adjectives *strog-ij* 'strict', *rez-k-ij* 'harsh', and *vys-ok-ij* 'high'. Let us go through these adjectives one by one. The paradigm of the adjective *strog-ij* 'strict' (shown in Table 7) presents the following puzzle: despite it forming a zero-comparative with no overt comparative affix, the comparative affix *-aj-* appears in addition to the superlative affix *-š-* in the superlative form. Recall that alternations like *strog-*/*strož-* are due to the palatalization phenomena (Blumenfeld 2003) and are not relevant for the present study's focus on the morphologically-conditioned allomorphy in Russian adjectives.

One could argue that *-ajš-* should not be decomposed and rather be treated as an allomorph of the superlative affix for adjectives which form a zero-comparative. However, the adjective *krut-oj* 'cool' and similar ones (the paradigms of which are shown in Table 8) provide circumstantial evidence against such a hypothesis: *krut-oj* forms a zero-comparative *kruč-e* 'cooler' and a superlative form *krut-ej-š-ij* 'coolest', which suggests that the *-aj-* found in *strož-aj-š-ij* 'strictest' is

Table 7: The degree paradigm of the adjective *strog-ij* 'strict'

POS	CMPR	SPRL
strog-ij	*strož-e*	*strož-aj-š-ij*
strict-AGR	strict-AGR	strict-CMPR-SPRL-AGR
'strict'	'stricter'	'strictest'

the allomorph of *-ej-*, the comparative affix in Russian. And again, the *krut-*/*kruč-*, *čist-*/*čišč-* and *bogat-*/*bogač-* alternations are morpho-phonological in nature and are thus irrelevant to the morphosyntactically-conditioned allomorphy patterns discussed in this paper.

Table 8: Zero comparatives with *-ej-*

POS	CMPR	SPRL
The degree paradigm of the adjective *krut-oj* 'cool'		
krut-oj	*kruč-e*	*krut-ej-š-ij*
cool-AGR	cool-AGR	cool-CMPR-SPRL-AGR
'cool'	'cooler'	'coolest'
The degree paradigm of the adjective *čistyj* 'clean'		
čist-yj	*čišč-e*	*čist-ej-š-ij*
clean-AGR	clean-AGR	clean-CMPR-SPRL-AGR
'clean'	'cleaner'	'cleanest'
The degree paradigm of the adjective *bogatyj* 'rich'		
bogat-yj	*bogač-e*	*bogat-ej-š-ij*
rich-AGR	rich-AGR	rich-CMPR-SPRL-AGR
'rich'	'richer'	'richest'

Taking both *strog-ij* 'strict' and *krut-oj* 'cool' into account, a puzzling picture emerges: while these adjectives form zero-comparatives (without the comparative affix *-ej-*/*-aj-*), the comparative affix emerges in the superlative form. While one could argue that we are dealing with an affix *-ejš-*/*-ajš-*, such an analysis misses a clear parallel to the regular adjectives like *glup-yj* 'dumb' in the superlative form. In what follows, I assume that the *-ej-*/*-aj-* in the superlative form is the same morphological entity (= result of the same insertion rule) as the *-ej-*

found in the comparative forms of regular adjectives. Furthermore, I make the assumption that the alternation between *-ej-* and *-aj-* is morpho-phonological in nature, which is supported by the observation that *-aj-* allomorph is only found after /k/-, /g/-, and /x/-final adjectival stems (which are transformed into /č ž š/, respectively). Of course, this argument predicts that /k/-, /g/-, and /x/-final adjectives form their comparative forms with *-aj-* but such adjectives always form zero-comparatives, so the prediction cannot be tested.

Assuming that the overtness of the comparative affix is the default option (*-ej-*/*-aj-* is the default allomorph), the pattern of zero-comparatives presents a non-trivial problem for a theory like Bobaljik's. In order to account for the covertness of CMPR in the comparative form *strože* 'stricter', one has to posit a VI rule like (3b) which expones CMPR as a zero in the context of adjectives like *krutoj* 'cool' and *strogij* 'strict', but then posit a more specified rule like (3a) which expones CMPR as the default allomorph *-ej-*/*-aj-* since without such a rule there would be no way for the CMPR to be exponed in the superlative form.

(3) Zero-comparatives require accidental homophony of CMPR

    a. CMPR ↔ *-ej-*/*-aj-* /X]__] SPRL] where $X \in \{\sqrt{\text{STRICT}}, \sqrt{\text{COOL}}, ...\}$

    b. CMPR ↔ ∅ /X]__ where $X \in \{\sqrt{\text{STRICT}}, \sqrt{\text{COOL}}, ...\}$

    c. CMPR ↔ *-ej-*/*-aj-*

Given that having a zero-comparative is a property of multiple lexical items, the accidental homophony solution appears dubious and hence I consider the pattern to be problematic for a straightforward DM approach to the presented data. In addition to that, there are two other problematic classes of adjectives left to be presented in this section, the first of which is exemplified by the adjective *rez-k-ij* 'harsh'. I dub this class AUGMENT ADJECTIVES, borrowing the term for the *-(o)k-* affix from Vanden Wyngaerd et al. (2020). Example paradigms of augment adjectives are provided in Table 9. The alternation between *-(o)k-* and *-(o)č-* is due to palatalization phenomena and the presence of *o* is conditioned by stress (cf. *vy-ˈsokij* and *ˈredkij*) and hence both alternations are ignored for present purposes.

On the surface, the augment adjectives present the very same pattern as zero-comparative adjectives: in the context of some syntactic nodes (be it $\sqrt{\text{STRICT}}$ or the augment *-(o)k-*) the CMPR node is zero but is exponed as its default form once the SPRL node enters the structure. The problem posed by augment adjectives is thus the same as posed by zero-comparatives, which raises the question of whether it is even sensible to draw a distinction between the two classes. However, foreshadowing my analysis, I will pursue the analytic strategy of deriving the patterns as portmanteaux – hence, the distinction between a root- and augment-triggered zero-comparative will prove useful in the later sections.

Table 9: Augment adjectives

POS	CMPR	SPRL
The degree paradigm of the adjective *rezkij* 'harsh'		
rez-k-ij	*rez-č-e*	*rez-č-aj-š-ij*
harsh-AUG-AGR	harsh-AUG-AGR	harsh-AUG-CMPR-SPRL-AGR
'harsh'	'harsher'	'harshest'
The degree paradigm of the adjective *žutkij* 'eerie'		
žut-k-ij	*žut-č-e*	*žut-č-aj-š-ij*
eerie-AUG-AGR	eerie-AUG-AGR	eerie-AUG-CMPR-SPRL-AGR
'eerie'	'eerier'	'eeriest'
The degree paradigm of the adjective *žarkij* 'hot'		
žar-k-ij	*žar-č-e*	*žar-č-aj-š-ij*
hot-AUG-AGR	hot-AUG-AGR	hot-AUG-CMPR-SPRL-AGR
'hot'	'hotter'	'hottest'
The degree paradigm of the adjective *gromkij* 'loud'		
grom-k-ij	*grom-č-e*	*grom-č-aj-š-ij*
loud-AUG-AGR	loud-AUG-AGR	loud-AUG-CMPR-SPRL-AGR
'loud'	'louder'	'loudest'

Now, consider the final class of adjectives: the *ABA-violating adjectives like *vys-ok-ij* (already mentioned in the introduction) in Table 10. Descriptively, the pattern is that the augment is not present in the comparative form but is present in the positive and superlative forms, which fits the ABA pattern as formulated by Bobaljik & Sauerland (2018), and is, thus, highly problematic for a theory that adheres to the containment hypothesis of Bobaljik (2012), which was put forward in order to exclude ABA patterns in degree morphology of adjectives.

Given the observations about the zero-comparative adjectives and the augment adjectives, however, we can decompose the *ABA-violating pattern into the combination of the observations about zero-comparative adjectives and augment adjectives in the following way. The ABA pattern consists of (i) the CMPR node being zero-exponed in comparative form only in the context of the augment (augment-adjectives pattern); (ii) the node adjacent to the adjectival root being zero-exponed in the comparative form only (zero-comparatives pattern). I believe that decomposing the ABA pattern into two distinct and attested patterns in Russian allows for a more grounded analysis (even though the two phenomena are still problematic).

Table 10: *ABA-violating adjectives

POS	CMPR	SPRL
Degree paradigm of the adjective *vys-ok-ij* 'high'		
vys-ok-ij	*vyš-e*	*vys-oč-aj-š-ij*
high-AUG-AGR	high-AGR	high-AUG-CMPR-SPRL-AGR
'high'	'higher'	'highest'
Degree paradigm of the adjective *red-k-ij* 'rare'		
red-k-ij	*rež-e*	*red-č-aj-š-ij*
rare-AUG-AGR	rare-AGR	rare-AUG-CMPR-SPRL-AGR
'rare'	'rarer'	'rarest'
Degree paradigm of the adjective *šyr-ok-ij* 'wide'		
šyr-ok-ij	*šyr-e*	*šyr-oč-aj-š-ij*
wide-AUG-AGR	wide-AGR	wide-AUG-CMPR-SPRL-AGR
'wide'	'wider'	'widest'
Degree paradigm of the adjective *gad-k-ij* 'disgusting'		
gad-k-ij	*gaž-e*	*gad-č-aj-š-ij*
disgusting-AUG-AGR	disgusting-AGR	disgusting-AUG-CMPR-SPRL-AGR
'disgusting'	'more disgusting'	'most disgusting'

To sum up, we have discussed the three classes of Russian adjectives that pose a problem for the containment hypothesis of Bobaljik (2012). The descriptive contribution of this paper ends here. The next section is devoted to introducing Nanosyntax (but I presuppose basic knowledge of the main tenets of Distributed Morphology). The section after that presents my Nanosyntactic analysis while introducing recent technical developments of the theory along the way.

3 Nanosyntax: the basics

This section presents the basics of Nanosyntax: its theoretical commitments and the inner workings of Nanosyntactic analyses, using the main building blocks of Russian adjectival morphology as the example (comparative *-ej-/-aj-* and superlative *-š-*). The first subsection presents the basic ideas behind Nanosyntax. The second subsection presents an analysis of regular adjectives in Russian.

3.1 The basics of Nanosyntax

Nanosyntax (Starke 2009, Baunaz & Lander 2018), like the mainstream Distributed Morphology approach to the syntax-morphology interface (Halle & Marantz 1994), is committed to the Single Engine Hypothesis (Marantz 2001): all complex expressions in languages are built by the same computational system (or module) – syntax. Unlike Distributed Morphology, however, Nanosyntax does not assume that individual syntactic terminals are morphemes / bundles of features (*pace* Embick 2015). Instead, Nanosyntax assumes a version of the One Feature – One Head thesis (Kayne 2005): all features are individual heads (and are, thus, privative). Thus, where DM would have a single "bundle" of, for example, φ-features on AGR nodes on adjectives (as in existing DM work on nominal concord, see Norris 2014 and the tree in Figure 2a), Nanosyntactic work on nominal concord assumes a hierarchy of feature heads (as in Caha 2023 and in the tree in Figure 2b).
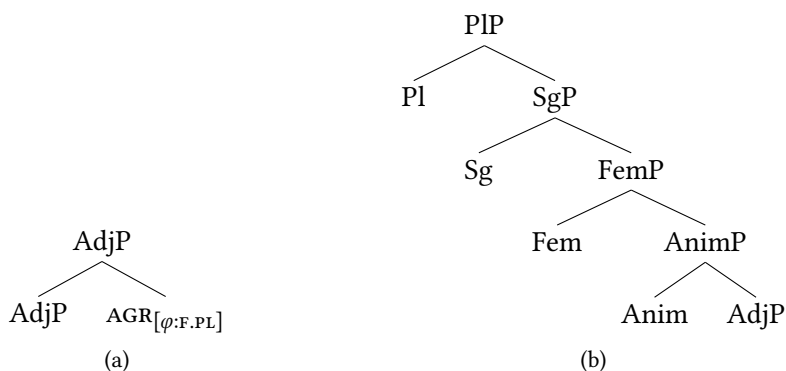
Figure 2: Structure for nominal concord in DM and Nanosyntax

The question is, how are the individual features grouped together to be matched to morphemes? This question requires a two-step answer. The first step is to introduce the notion of phrasal spell-out. While DM assumes that Vocabulary Insertion maps syntactic terminals onto morpho-phonological representations, Nanosyntactic work assumes that Vocabulary Insertion targets constituents. The idea is, then, that the bundles of features form syntactic constituents in order to be lexicalized together. Nanosyntax forms such constituents of features/feature-heads via syntactic movement according to the Spell-Out algorithm provided in (4). The core idea behind this algorithm is that after a new feature-head is merged, the resulting structure must be transformed into a structure that can be spelled-out, or to put it another way, whose subconstituents can

be matched to existing lexical entries. The core property of the algorithm is that cumulative exponence is preferred (the separate exponence of the newly merged F is only possible via the step in 4c).

(4)   Spell-Out algorithm

    a.   Merge F to XP and spell out

    b.   If (a) fails, move Spec,XP to Spec,FP and spell out

    c.   If (b) fails, move XP to Spec,FP and spell out

    d.   If (c) fails, move to the next option in the previous cycle (backtracking)

Now the question lies in the precise nature of lexical entries in Nanosyntax and matching the feature structures to these entries. In Nanosyntax, lexical entries (or L-trees, to use proprietary terminology) are pairs of morpho-phonological representations and syntactic trees; an example is given in Figure 3.[2]

ZP ⇔ /blick/

Z   YP

Y   XP
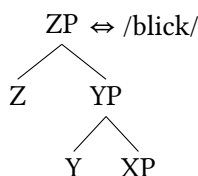
Figure 3: An example of an L-tree

The matching of constituents to L-trees is regulated by the Superset Principle, which states that an L-tree can be matched to any subconstituent of the structure in the L-tree. So, given the structure XP, the two lexical entries in Figure 4 match it (since XP is a subconstituent of both), which requires a principled way of choosing between the two matching L-trees.

The choice between L-trees that match to the structure is regulated by the following principle: the L-tree with the least amount of structure not found in the syntactic constituent undergoing spell-out is chosen. So, between the two matching L-trees in Figure 4, the second one is to be preferred since it contains less "excess" structure.

Finally, to end this quick introduction, I want to emphasize that phrasal spell-out and spell-out-driven movement are the only operations available in the Nano-syntax machinery. There is no contextual allomorphy (or readjustment rules, or

---

[2]I do not touch on the topic of the syntax-semantics interface in Nanosyntax (or any Late Insertion theory) due to the complexity of the issue and its lack of direct relevance to the paper.
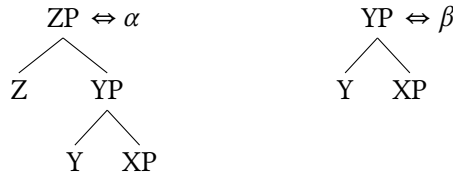
Figure 4: Two matching L-trees for XP

impoverishment rules, or any other familiar DM operation) in Nanosyntax, only portmanteaux, and, thus, a difference in form implies the presence of additional structure or a phonological analysis.

3.2 A case study: basic adjectival morphology of Russian

To recap the previous subsection, Nanosyntax assumes phrasal spell-out of syntactic constituents consisting of individual feature-heads which are formed via movement. The sequence of features (or f-seq) for degree morphology, according to Nanosyntactic work (see Caha et al. 2019 for the argumentation in favour of the split structure for degree morphology) is provided in (5).

(5)   Nanosyntactic f-seq for degree morphology
      AdjP – Q – C1 – C2 – S1 – S2

In this subsection, I will provide a Nanosyntactic analysis of the basic paradigm of regular adjectives in Russian, repeated in Table 11. The main goal is to provide a lexical entry for the comparative affix *-ej-/-aj-* and for the superlative affix *-š-*.

Table 11: The paradigm of a regular adjective

POS	CMPR	SPRL
glup-yj	*glup-ej-e*	*glup-ej-š-ij*
dumb-AGR	dumb-CMPR-AGR	dumb-CMPR-SPRL-AGR
'dumb'	'dumber'	'dumbest'

The split comparative and split superlative structures proposed by Caha et al. (2019) increase the number of analytical choices we are facing. While a simple Bobaljik-style structure would require CMPR being realized as *-ej-/-aj-* and SPRL as *-š-*, the f-seq in (5) allows for various lexicalizations. Since the data of regular adjectives underdetermines the analysis, I will provide the lexical entries, which

allow for the analyses of the three problematic classes that will be presented in the next section. One thing to note is that the constituents in the L-trees for *-ej-* and *-š-* are remnant constituents (constituents, out of which something has moved, as shown by the presence of unary branching at the foot of the tree), see Figure 5. Such remnant constituents are exclusively associated with suffixes in the Nanosyntax literature (see Starke 2018).
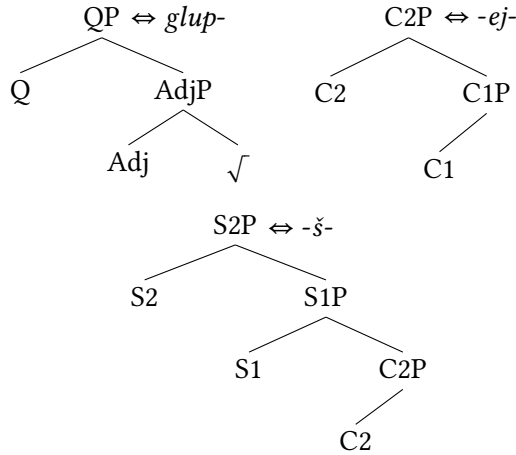
Figure 5: Lexical entries for regular adjective paradigms

Here, we shall go through the derivations step by step to show that the proposed lexical entries result in the observed paradigm. The thing to keep in mind is the standard Nanosyntax Spell-Out algorithm (Baunaz & Lander 2018) repeated in (6).

(6) Standard Nanosyntax Spell-Out algorithm
   a. Merge F to XP and spell out
   b. If (a) fails, move Spec,XP to Spec,FP and spell out
   c. If (b) fails, move XP to Spec,FP and spell out
   d. If (c) fails, move to the next option in the previous cycle (backtracking)

The derivation of the positive form is trivial: AdjP can be realized by the adjectival stem due to the Superset Principle and QP is the exact match of the lexical entry for *glup-*, as shown in Figure 6b. I want to note here that, in presenting the Nanosyntactic derivations, I will match subconstituents to the affixes in their

underlying form. While I understand that it hurts the readability of the lexical-izations themselves, the clarity of the paper overall benefits from this decision, in my opinion.

(a) Start with AdjP   (b) Merge Q to AdjP and spell out
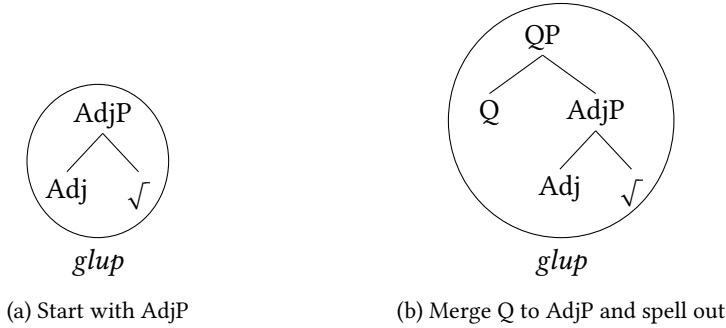
Figure 6: Deriving the positive form *glup-*

The derivation of the comparative is also rather straightforward, as shown in Figure 7. Since the [C1 [Q AdjP]] structure does not match any lexical entry, as indicated by the double exclamation marks in Figure 7a, the next step is to move the specifier of QP to Spec,C1. However, there is no specifier of QP and, thus, the next step is to move QP to Spec,C1, which results in the proper lexicalization of C1P in Figure 7b. Then, the C2 head is merged and the resulting structure does not match any lexical entry in Figure 7c, which results in movement of QP to Spec,C2P in Figure 7d. This structure results in the observed form *glup-ej-*, given our lexical entries.

The derivation of the superlative form is more complex and requires backtrack-ing, the final step in the spell-out algorithm provided in (6). After S1 is merged, there is no licit lexicalization even with movement of QP to Spec,S1P in Figure 8b and movement of the whole comparative structure to Spec,S1P in Figure 8c. The reason for this is that the lexical entry for the superlative affix *-š-* requires there to be a subconstituent with [C2P [C2]] at its foot, which isn't present at this point in the derivation. Hence, backtracking happens and the procedure goes back to the "next option in the previous cycle" step, namely, movement of C1P to Spec,C2P in Figure 8d. After that, merging S1 in Figure 8e and moving C1P to Spec,S1P results in a lexicalizable structure in Figure 8f. After merging S2 in Figure 8g and moving C1P to Spec,S2P in Figure 8h, we end up with a structure that is realized as the observed form for the superlative *glup-ej-š-*.

Even though it is not necessary for an analysis of the paradigm of regular adjectives on its own, the core property of the proposed analysis is that lexical entries for *-ej-* and *-š-* overlap in their inclusion of the C2 head: superlatives thus

(a) Merge C1 to QP

(b) Move QP to Spec,C1P

(c) Merge C2 to C1P

(d) Move QP to Spec,C2P
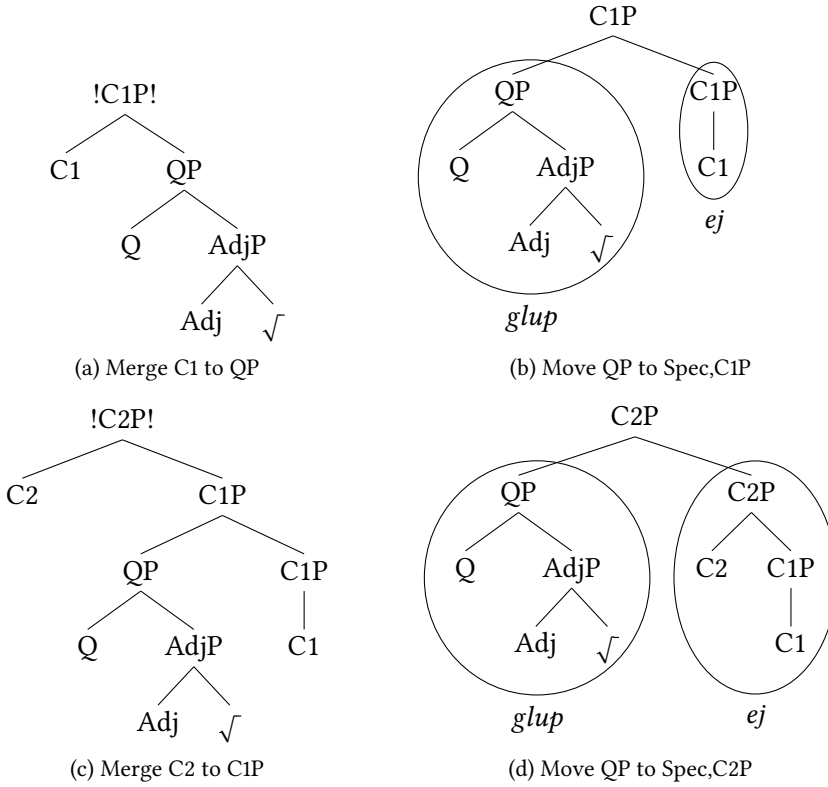
Figure 7: Deriving the comparative form *glup-ej-*

require backtracking (informally, splitting of C2 from C1 in lexicalization) and this property of the analysis presented will become relevant in the analysis of zero-comparatives, which is presented in the following section, along with the analyses for augment adjectives and *ABA-violating adjectives.
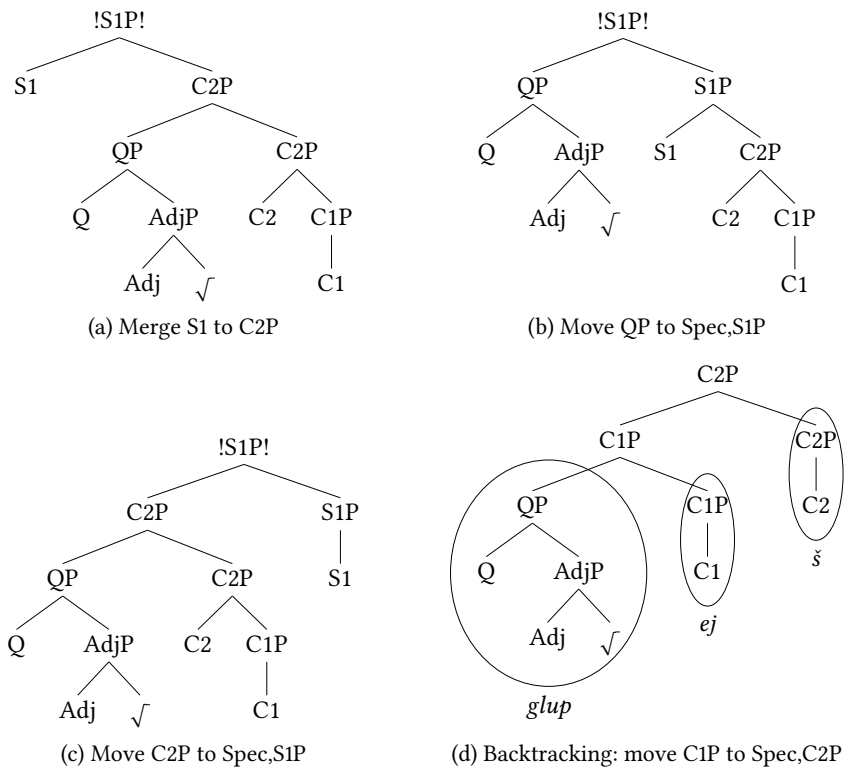
(a) Merge S1 to C2P

(b) Move QP to Spec,S1P

(c) Move C2P to Spec,S1P

(d) Backtracking: move C1P to Spec,C2P

Figure 8: Deriving the superlative form *glup-ej-š-*

(e) Merge S1 to C2P

(f) Move C1P to Spec,S1P

(g) Merge S2 to S1P

(h) Move C1P to S2P
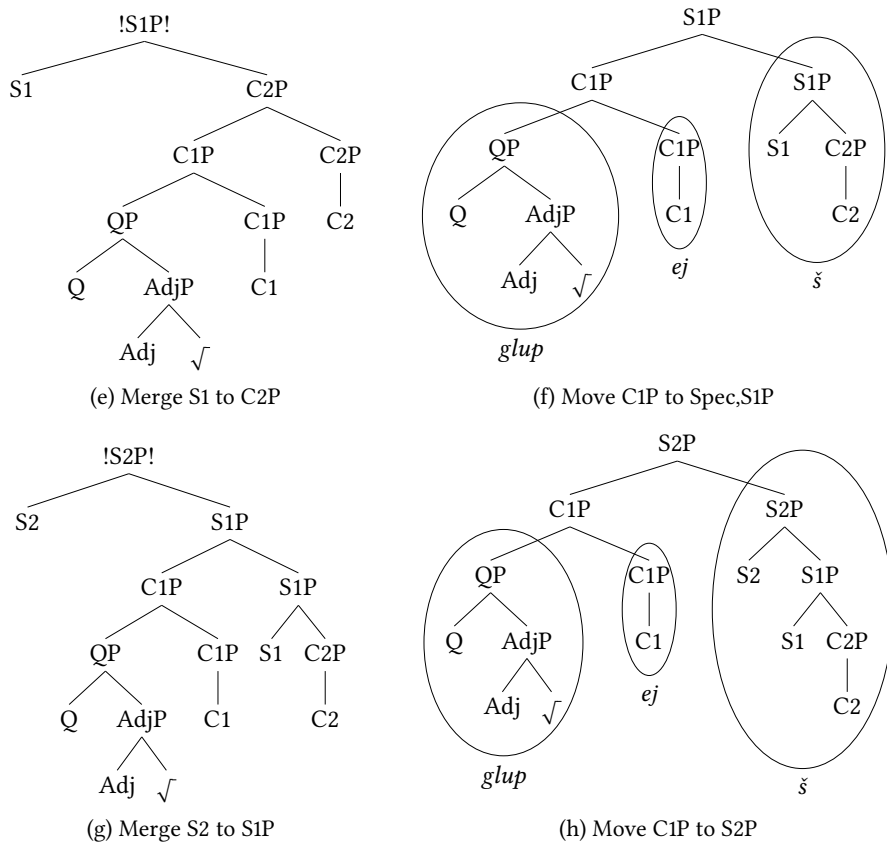
Figure 8: Deriving the superlative form *glup-ej-š-* (continued)

4 Analysis of three problematic classes

This section presents the Nanosyntactic analysis of the three problematic adjective classes: *strogij*-type adjectives (zero-comparatives), *rezkij*-type adjectives (augment adjectives), and *vysokij*-type adjectives (*ABA-violating adjectives). In Section 4.1, I present an analysis of zero-comparatives and introduce the notion of Movement-Containing Trees (Blix 2022) along the way. In Section 4.2, I present an analysis of augment-adjectives and introduce the novel spell-out algorithm of Caha & Taraldsen Medová (2022, 2023). Finally, Section 4.3 puts the analyses in Section 4.1 and Section 4.2 together to derive the ABA pattern established in the introduction. The core idea behind the analyses lies in the backtracking step forced by the L-tree of the superlative affix, as discussed in the previous section: in all three analyses, the backtracking step will trigger re-bundling of the features resulting in the exponence of the comparative and the augment.

4.1 Zero-comparatives: the need for movement-containing trees

Let me repeat the pattern and the problem for a DM-style approach posed by zero-comparative here. The basic pattern is as follows: the comparative affix *-ej-*/*-aj-* is absent from the comparative form itself, but arises in the decomposition of the superlative form, as shown in the paradigm in Table 12 for the adjective *strogij* 'strict'.

Table 12: The degree paradigm of the adjective *strog-ij* 'strict'

POS	CMPR	SPRL
strog-ij	*strož-e*	*strož-aj-š-ij*
strict-AGR	strict-AGR	strict-CMPR-SPRL-AGR
'strict'	'stricter'	'strictest'

The problem for a DM-style analysis was that one appears to need a zero-insertion rule for the CMPR node, which is sensitive to the adjacent adjective. However, this rule needs to be overridden in the superlative form, which results in an accidental homophony for the default VI rule for CMPR and the rule which is sensitive to both the adjective and the presence of SPRL.

A basic Nanosyntax model (like the one introduced in Baunaz & Lander 2018) cannot accommodate these findings either. In Nanosyntax, having a zero-comparative entails that the adjectival root (like *strog-*) has the comparative structure (C1 and C2 heads) in its lexical entry, as shown in Figure 9.

```
              C2P ⇔ strog-
             ╱         ╲
          C2            C1P
                       ╱    ╲
                    C1       QP
                            ╱  ╲
                          Q     AdjP
                               ╱   ╲
                             Adj    √
```
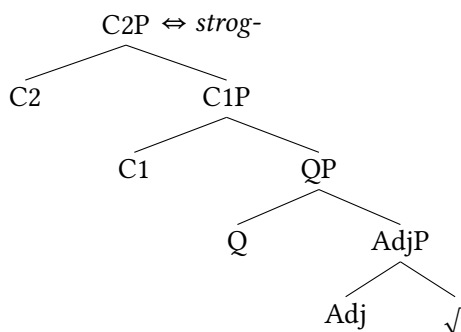
Figure 9: A putative lexical entry (L-tree) for *strog-*

The problem then is that when superlative heads are introduced into the derivation (S1 and S2), there is no way to trigger the overt comparative affix no matter what the lexical entry for the superlative affix – all comparative structures will be realized either by the adjectival stem or by the superlative affix. Given backtracking, all structures will be divided into the adjectival stem and the superlative affix, one way or another (see Figure 10).

The solution for the problem of the comparative affix suddenly being overt in the superlative form comes from the work of Hagen Blix arguing that phrasal spell-out entails the possibility of spelling out constituents that "include" movement. To be more substantive, Blix suggests that L-trees like Figure 11 are available in the lexicon, given the possibility of spelling out whole constituents (see Blix 2022 for an exploration of this idea based on Kipsigis number morphology). In accordance with an anonymous reviewer's comments, I emphasize that the idea of Movement-Containing lexical entries is not a theoretical addition to the Nanosyntactic project but rather an under-explored representational possibility.

For our purposes, the main consequence of the proposed lexical entry is that there is no subconstituent of the L-tree in Figure 11 that contains both C1 and the adjective to the exclusion of C2. Hence, if we force C2 to be spelled-out together with superlative structure (via the backtracking step, see the previous subsection), the comparative affix will arise, as shown in Figure 12. Note that the derivational steps are the same as with the regular adjectives – the only difference comes from the fact that adjectival stems like *strog-* are able to realize the whole comparative structure.

To put it informally, Blix's (2022) proposal allows for a formalization of the intuition that the comparative affix is zero-exponed in the comparative only: it is "inside" a portmanteau form, which is possible in the comparative form only

(a)

(b)

(c)
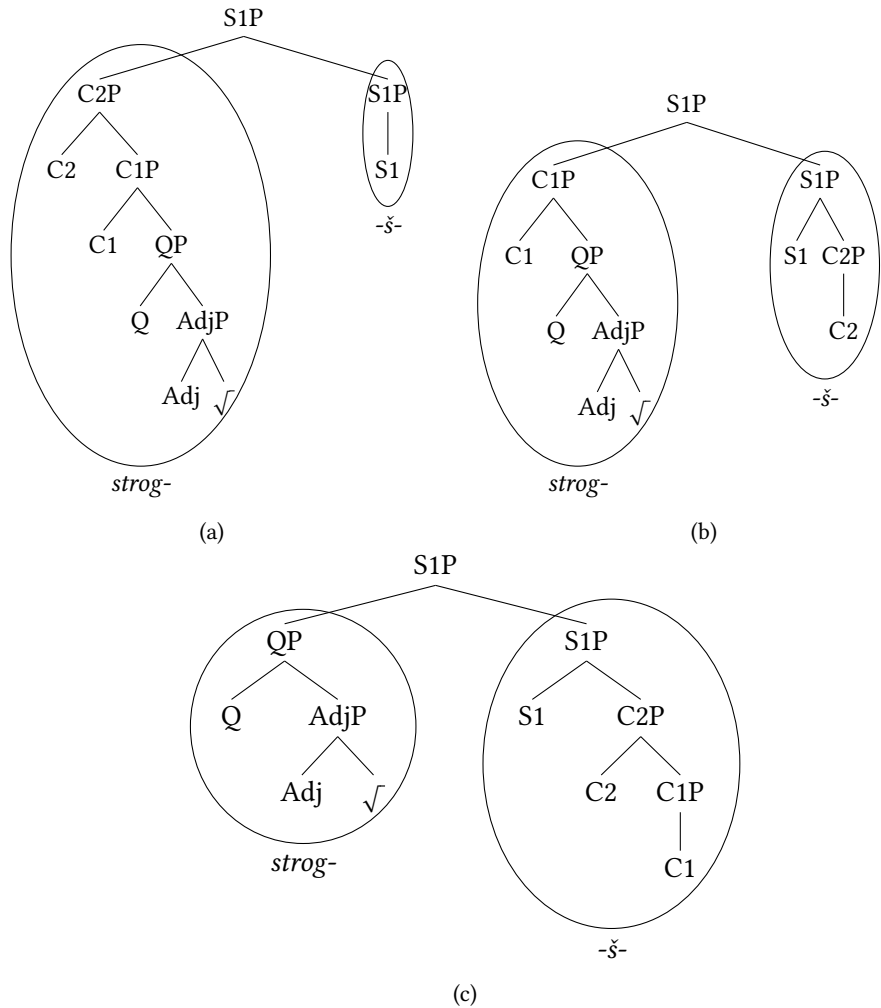
Figure 10: Possible lexicalizations of S1P given the L-tree for *strog-*
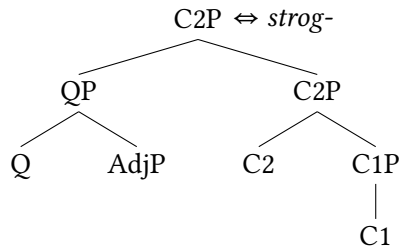
Figure 11: Movement-Containing Tree for the lexical entry of *strog-*

(a) Lexicalization of the comparative *strož-* (cf. *glup-ej-*)

(b) Lexicalization of the superlative *strož-aj-š-* (cf. *glup-ej-š-*)
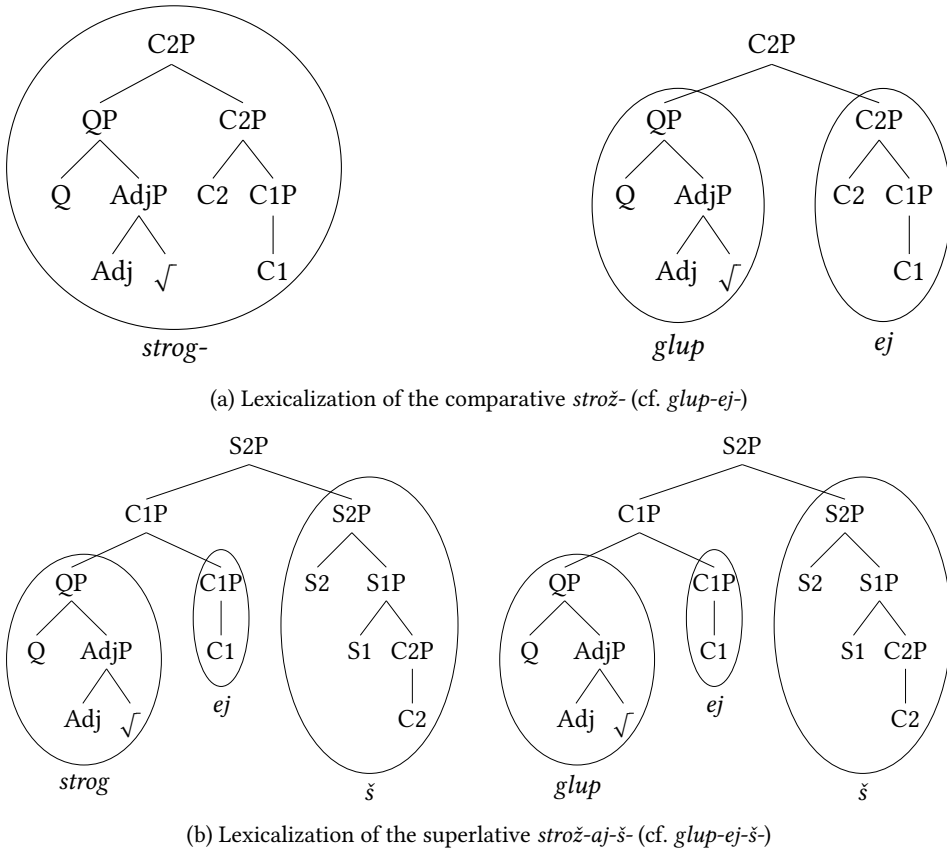
Figure 12: Lexicalizations of comparative and superlative forms of *strogij* 'strict'

due to the internal structure of the lexical entry. Once S1 merges, the lexicalization requires bundling C2 together with S1, which results in QP being the only available subconstituent of the L-tree in Figure 11, forcing the exponence of *-ej-* in the superlative form.

4.2 Augment adjectives: the need for subextraction

Although the movement-containing trees (together with backtracking) have allowed us to capture the zero-comparative class of adjectives, the augment adjectives present an additional puzzle: we need the augment affix itself to realize the comparative structure. The desired lexicalization is as follows: there is some right branch that spells out Q, C1 and C2 together in the comparative form, and

the superlative form must look like every other superlative form does: Adj, Q, C1 and C2-S1-S2 are lexicalized by distinct affixes.

The question is, how does one come to these lexicalizations given the regular Nanosyntax Spell-Out algorithm provided in Baunaz & Lander (2018) and Starke (2018). My answer is: it is impossible. Let us see why. Given the regular Nanosyntax algorithm, the fact that the comparatives only have a single affix implies the lexicalization in Figure 13.
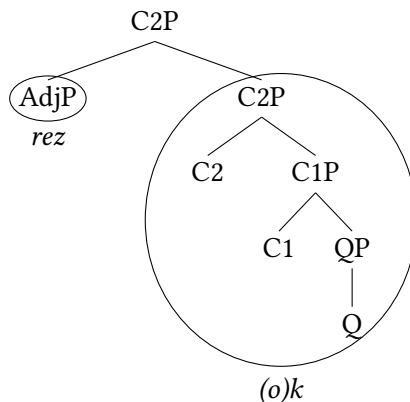
Figure 13: Lexicalization with a single affix according to regular Nanosyntax algorithm

The following problem then arises: any L-tree which matches with the right branch of the tree in Figure 13 will match to the subconstituent without C2 (given the Superset Principle). Hence, the predicted lexicalization for the superlative form does not include the comparative affix, contrary to the data, see Figure 14.

This problem motivates a theoretical addition to the Nanosyntax model. To account for augment adjectives, I employ the novel subextraction Nanosyntax algorithm, which is given in (7). Here, I take "non-remnant" to mean "not containing a unary branch". See Caha & Taraldsen Medová (2023) for a similar algorithm.

(7) Subextraction spell-out algorithm (cf. Caha & Taraldsen Medová 2023)
   a. Merge F and spell-out
   b. If (a) fails, move the closest non-remnant constituent to Spec,FP
   c. If (b) fails, move the dominating node to Spec,FP (recursive step)
   d. If (c) fails, try the next option in the previous cycle

Compared to the regular Nanosyntax algorithm, the steps are the same in a single-affix structure as the steps in the standard algorithm: the first step is to move the
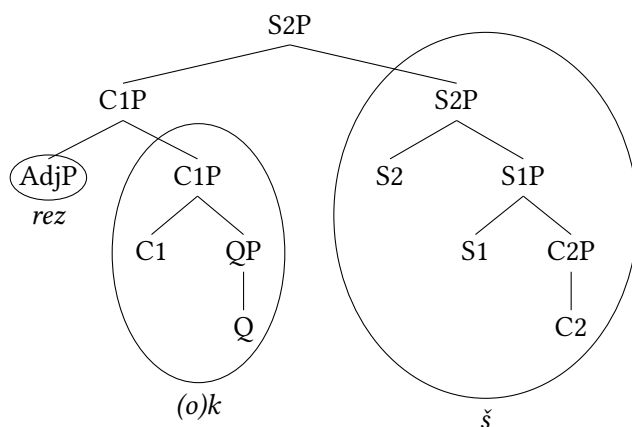
Figure 14: Lexicalization of superlative given the regular Nanosyntax algorithm

specifier (the closest non-remnant constituent) and the second step is to move the whole structure (the dominating constituent). The difference comes with multiple affix structures: given the structure in Figure 15a, the first step would be to move the XP to Spec,HP (as in Figure 15d) and not YP (as was the case with the old algorithm, see the step in Figure 15c) – the novel algorithm makes heavy use of subextraction (moving a subconstituent from a specifier).

The right branch of the structure in Figure 15d is a peculiar one because if there is backtracking in the derivation, the next step in the derivation is the structure in Figure 15c – hence, if there is an L-tree for the right branch in Figure 15d, backtracking may force the exponence of the affix which realizes [YP [Y]]. The core idea is shown graphically in Figure 16.

This property is important given the lexical entry I propose for the augment affix, given in Figure 17. The main idea is that the backtracking step forced by the L-tree for the superlative affix -š- will trigger exponence of the affix which realizes [C1P [C1]] (namely, -ej-). Before I show the step-by-step derivation for the superlative form, let us go through the necessary steps for such a constituent to arise in the first place. The first step is to provide an L-tree for the adjectival stem, which does not include Q since it it realized by the augment.

Now, let us go through the whole derivation in order to show that the lexical entry in Figure 17, coupled with the subextraction spell-out algorithm, results in the observed morphological pattern of augment adjectives. After Q is merged (Figure 18b), the movement of AdjP to Spec,QP is necessary to lexicalize the structure (Figure 18c).

(a) A multi-affix structure

(b) Merge H

(c) Old: move YP to Spec,HP

(d) New: move the closest non-remnant constituent (XP) to Spec,HP
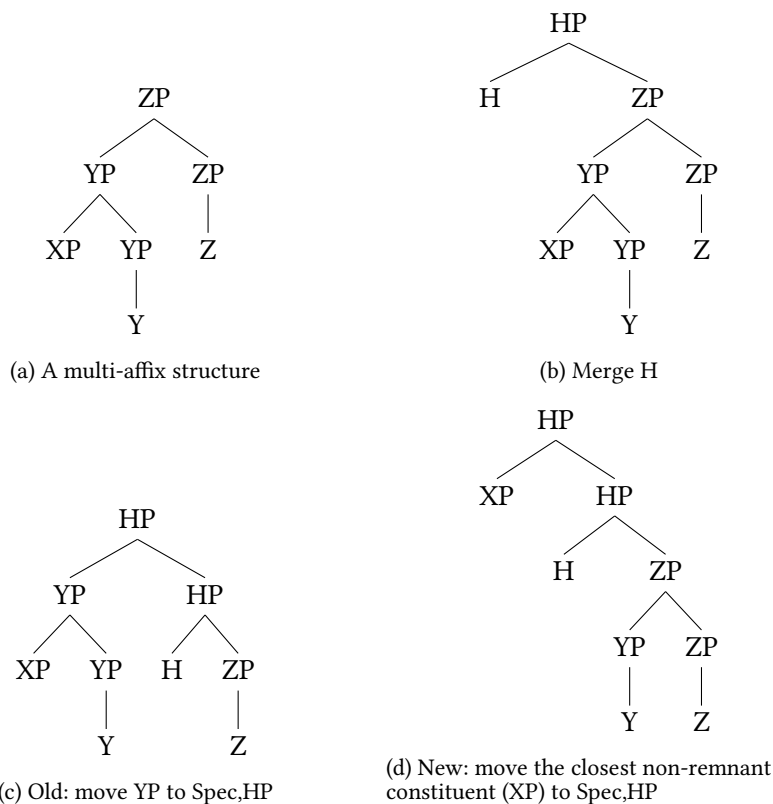
Figure 15: Complex specifier and old/new spell-out algorithm

Then, after C1 is merged (Figure 19a), the movement of AdjP to Spec,C1P does not allow proper lexicalization (Figure 19b) and the next step is done (movement of QP to Spec,C1P, see Figure 19c), which results in a licit lexicalization. Note that this lexicalization does not correspond to any existing form – that is not an issue since C1P does not occur in the absence of C2P.

After C2 is merged (Figure 20a), the first step of the subextraction algorithm is to move AdjP (and not QP, since it contains a unary branch [QP [Q]]) to Spec,C2P (Figure 20b), which results in a right branch that matches the L-tree for the augment, deriving the fact that the comparative affix is not present in the comparative form of augment adjectives.

Now, let us see what happens when the superlative structure is introduced into the derivation. After S1 is merged (Figure 21a), no operation in the cycle (movement of AdjP in Figure 21b, C2P in Figure 21c) results in a proper lexicalization. Thus, backtracking is necessary.

HP

XP
α

HP

H ZP

YP ZP

Y Z

β

(a) Lexicalization of the right branch

HP

YP HP

XP YP H ZP
α

Y Z

γ δ

(b) The next step (after backtracking)
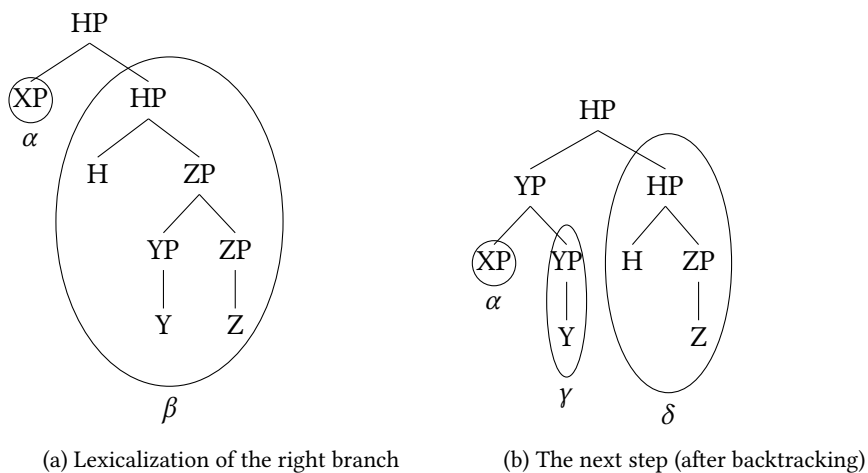
Figure 16: Affix emergence with backtracking and subextraction algorithm

C2P ⇔ *(o)k-*

C2 C1P

QP C1P

Q C1

AdjP ⇔ *rez-*

Adj √
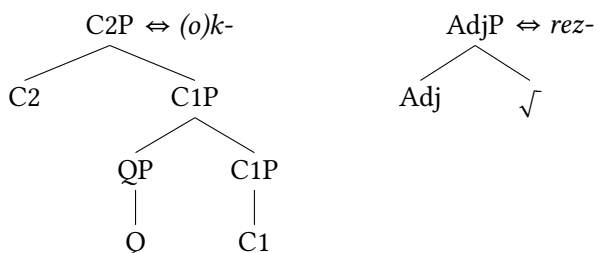
Figure 17: L-trees for the augment and for the stem *rez-*

AdjP

Adj √

rez

(a) Start with AdjP

!QP!

Q AdjP

Adj √

(b) Merge Q

QP

AdjP QP

Adj √ Q

rez *(o)k*

(c) Move AdjP to Spec,QP
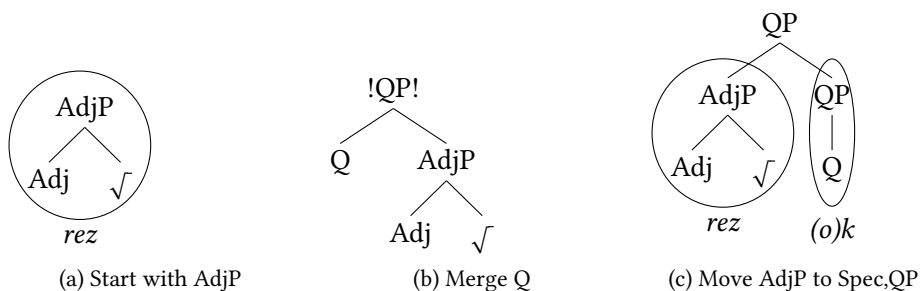
Figure 18: Deriving the positive form *rez-k-*

(a) Merge C1

(b) Move the closest non-remnant constituent (AdjP) to Spec,C1P

(c) Move the dominating constituent (QP) to Spec,C1P
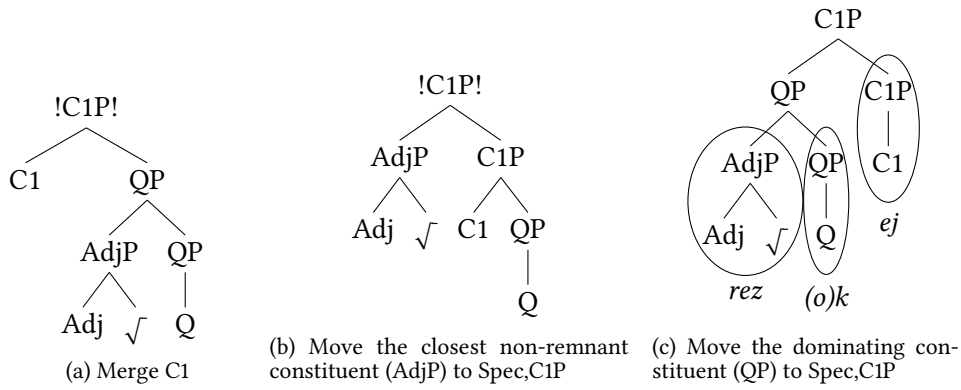
Figure 19: Lexicalizing C1P

(a) Merge C2

(b) Move the closest non-remnant constituent (AdjP) to Spec,C2P
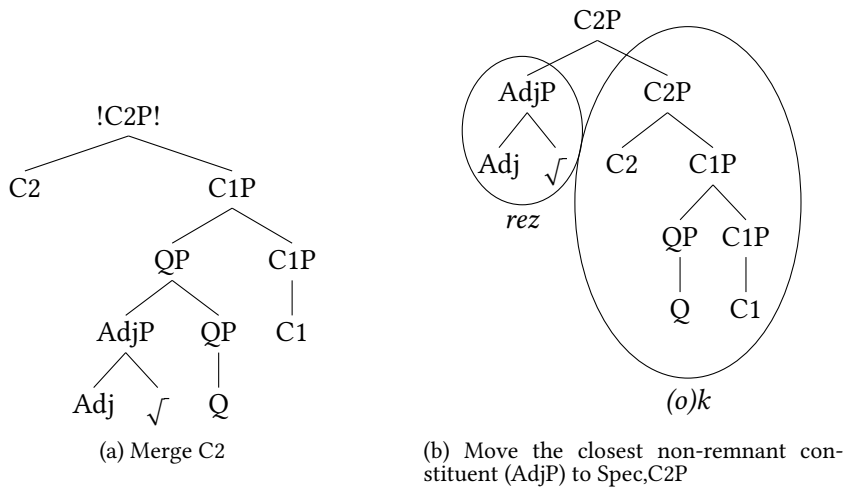
Figure 20: Lexicalizing C2P

(a) Merge S1

(b) Move the closest non-remnant constituent (AdjP) to Spec,S1P

(c) Move the dominating constituent (C2P) to Spec,S1P
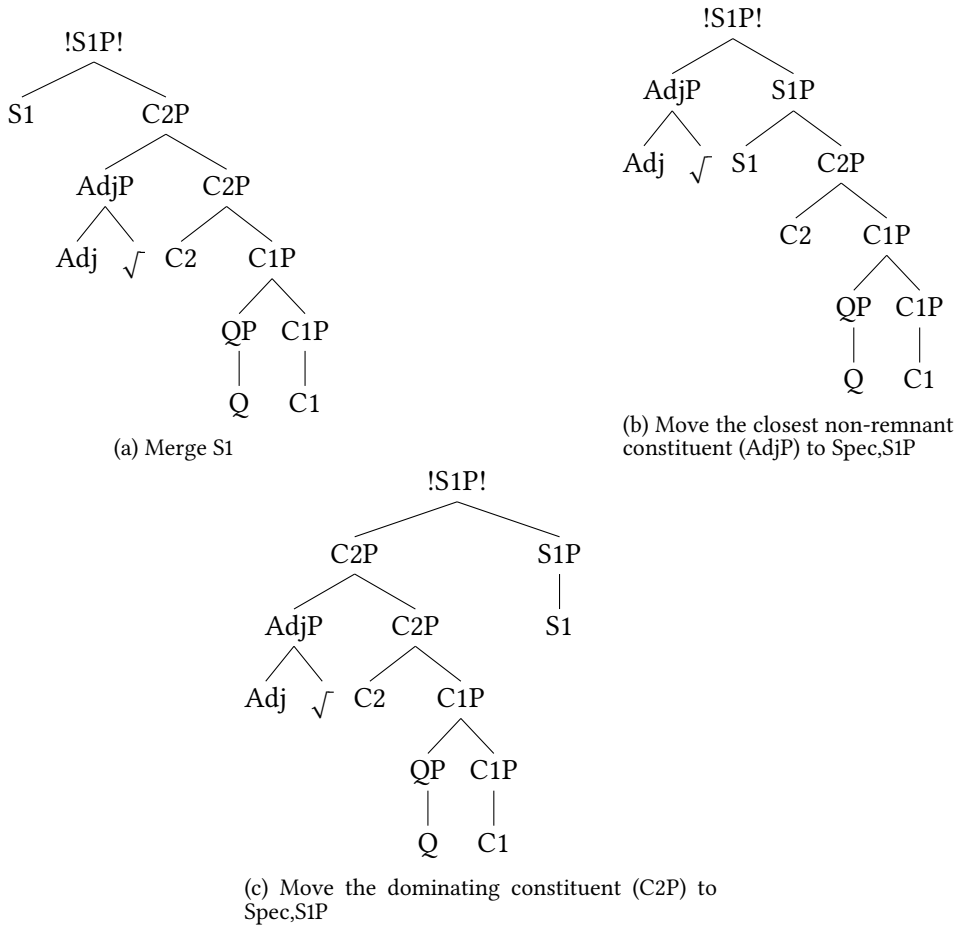
Figure 21: First cycle in lexicalizing S1P

However, the first step of backtracking will be movement of QP to Spec,C2P, resulting in a proper lexicalization more reminiscent of the regular adjectives, as shown in Figure 22.
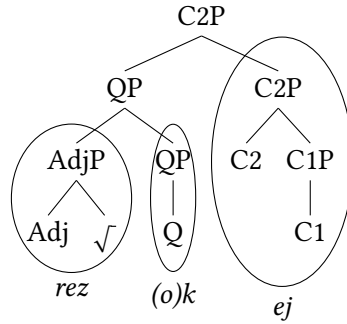
Figure 22: Backtracking: Move the dominating constituent (QP) to Spec,C2P

As was the case with regular adjectives, merging S1 still does not result in a proper lexicalization (as shown in Figure 23), no matter what the operation (movement of AdjP, QP, C2P) and thus backtracking is necessary once again.

The next backtracking step is to move the C1P to Spec,C2P, resulting in a structure that allows for future lexicalization of C2 together with S1 and S2, as shown in Figure 24. Note that this instance of backtracking mirrors the derivational steps necessary to lexicalize the superlative form of the regular adjectives.

After merging S1 (Figure 25a), movement of AdjP (Figure 25b) and QP (Figure 25c) does not result in a proper lexicalization – but movement of C1P does, as shown by Figure 25d.

The same thing happens after merging S2 (Figure 26a): movement of AdjP (Figure 26b) and QP (Figure 26c) does not result in a proper lexicalization, but movement of C1P does (Figure 26d). In the end, we derive the superlative form *rez-č-aj-š-ij* with an overt comparative affix.

This subsection has presented an analysis for the problematic degree morphology pattern of Russian augment adjectives. The core analytical move was the L-tree for the augment: its shape and the novel subextraction spell-out algorithm guarantee that backtracking (which happens due to C2 being in the L-tree for the superlative affix) results in a structure where [QP [Q]] is the only subconstituent matching to the L-tree of the augment, allowing for the independent realization of the [C1P [C1]] subconstituent. In the next subsection, I will synthesize ideas from the proposed analyses of zero-comparatives and augment adjectives in order to account for the *ABA-violating adjectives.

(a) Merge S1

(b) Move the closest non-remnant constituent (AdjP) to Spec,S1P

(c) Move the dominating constituent (QP) to Spec,S1P

(d) Move the dominating constituent (C2P) to Spec,S1P
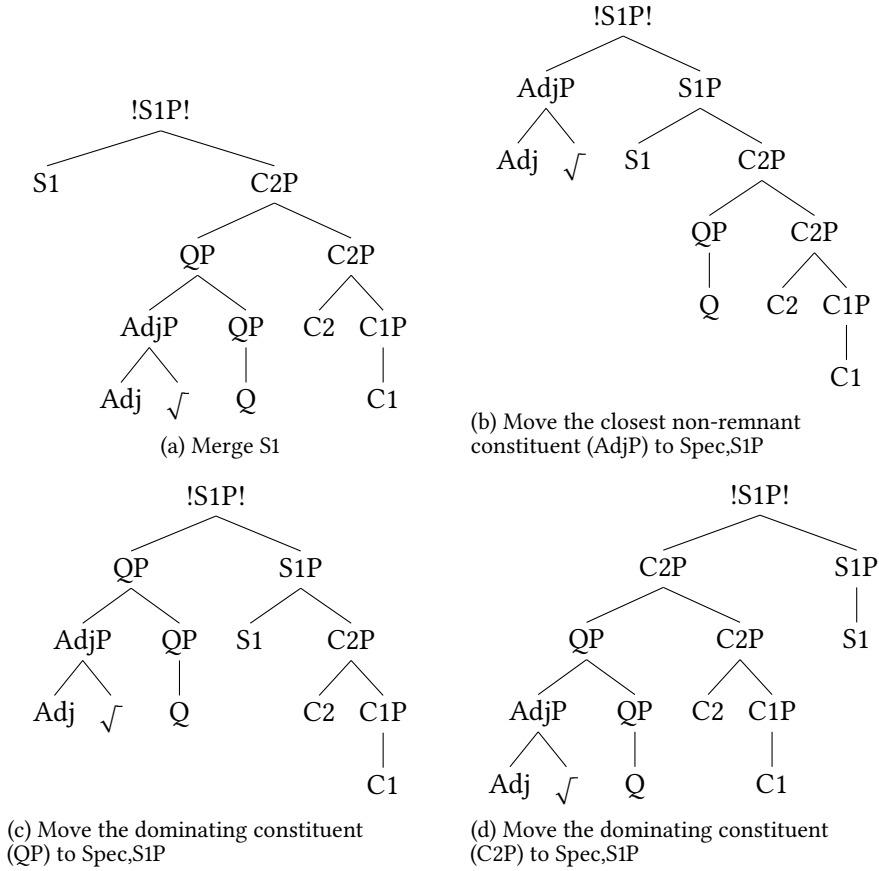
Figure 23: Lexicalizing S1P after first backtracking
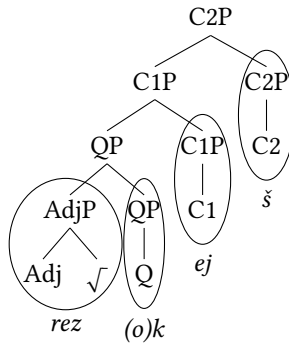
Figure 24: Backtracking: Move the dominating constituent (C1P) to Spec,C2P

(a) Merge S1

(b) Move the closest non-remnant constituent (AdjP) to Spec,S1P

(c) Move the dominating constituent (QP) to Spec,S1P

(d) Move the dominating constituent (C1P) to Spec,S1P
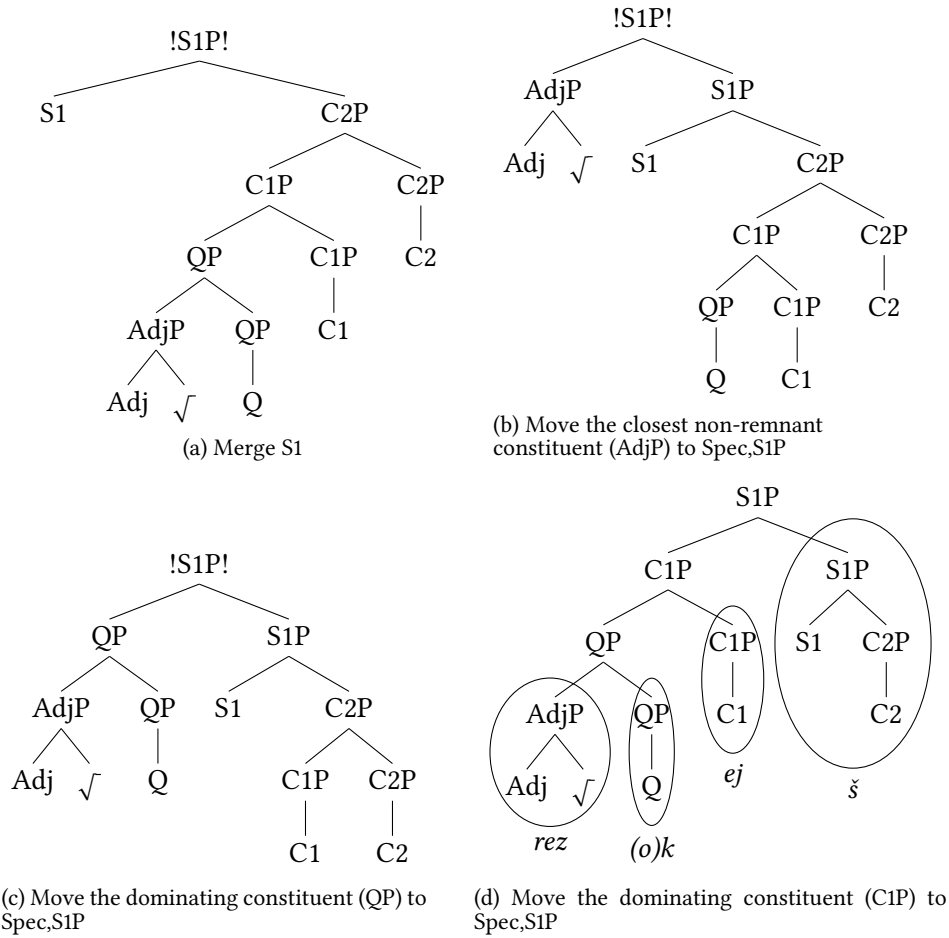
Figure 25: Lexicalizing S1P

(a) Merge S2

(b) Move the closest non-remnant constituent (AdjP) to Spec,S2P

(c) Move the dominating constituent (QP) to Spec,S2P

(d) Move the dominating constituent (C1P) to Spec,S2P
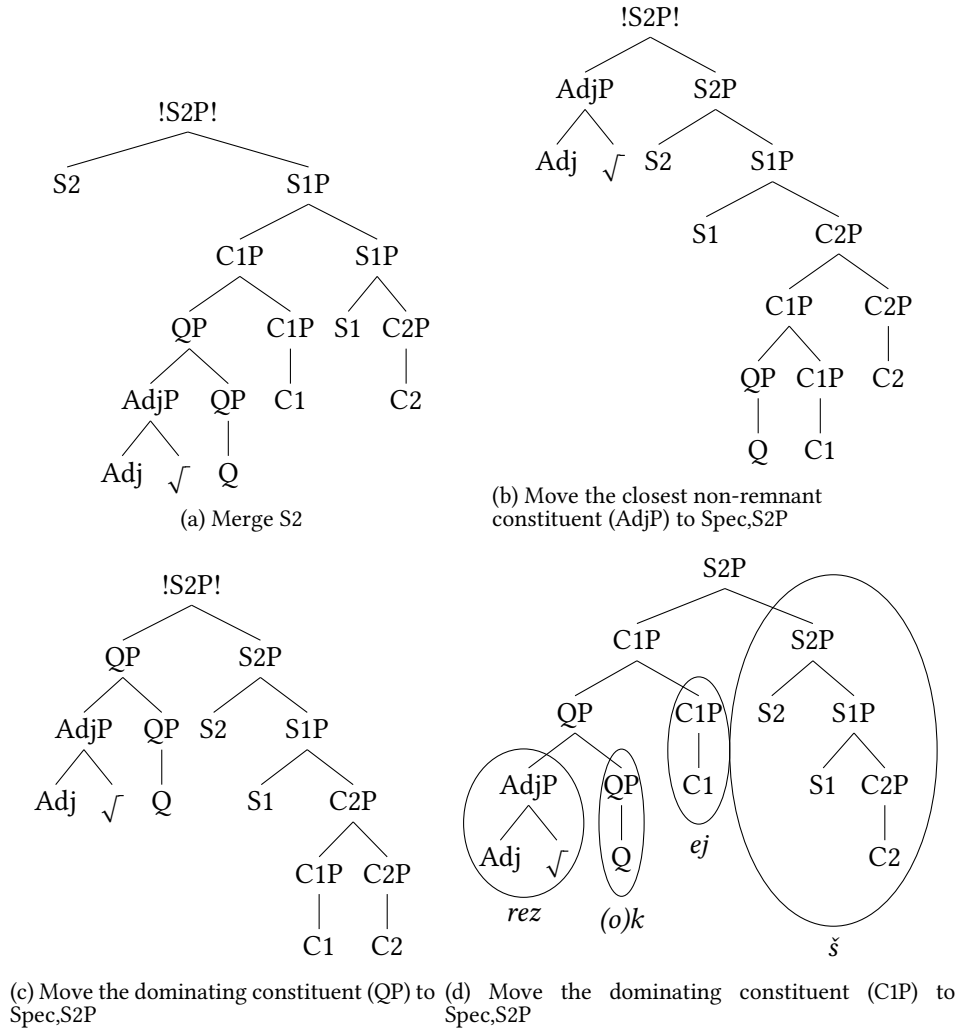
Figure 26: Lexicalizing S2P

4.3 *ABA-violating adjectives: putting the pieces together

At this moment, the solution to the *ABA-violating class of adjectives should be rather clear. The final lexicalized structure for comparatives of augment adjectives should be the L-tree for the adjectival stems of *ABA-violating adjectives, as given in Figure 27 for *vys-*. The solution to the ABA distribution of zero-exponence of the augment in the *ABA-violating class is thus the same as the solution to the distribution of zero-exponence of the comparative affix in the zero-comparative class: the lexical entry for the adjectival stem is such that the whole structure for the comparative form is a portmanteau.
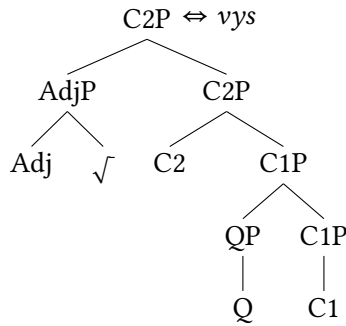
Figure 27: Lexical entry for *vys*

The core property of this proposal is that this lexical entry does not provide any subconstituent with AdjP that is not an AdjP itself or the whole tree, which means that the adjectival root will not be available to spell-out anything but AdjP in the positive form (resulting in *vys-ok-ij*, the lexicalization of which is given in Figure 28a) and the superlative form (resulting in *vys-oč-aj-š-ij*, the lexicalization of which is given in Figure 28b).

Note that, from the derivational point of view, there is no difference in the spell-out steps for augment adjectives and *ABA-violating adjectives – the only difference is the lexical property of *ABA-violating adjectives that they happen to have the correct right branch in their lexical entry, which creates an appearance of an ABA pattern with respect to the overtness of the augment. The core analytical contribution here is that the puzzling ABA pattern results from a combination of two independent phenomena (with theory-laden description) found in the domain of Russian adjectival morphology: the first phenomenon is the pattern of zero-comparatives (the adjectival root triggers zero-exponence of morphosyntactic material in the comparative form only), which is captured by

(a) Lexicalized structure for *vys-ok-*

(b) Lexicalized structure for *vys-oč-aj-š-*
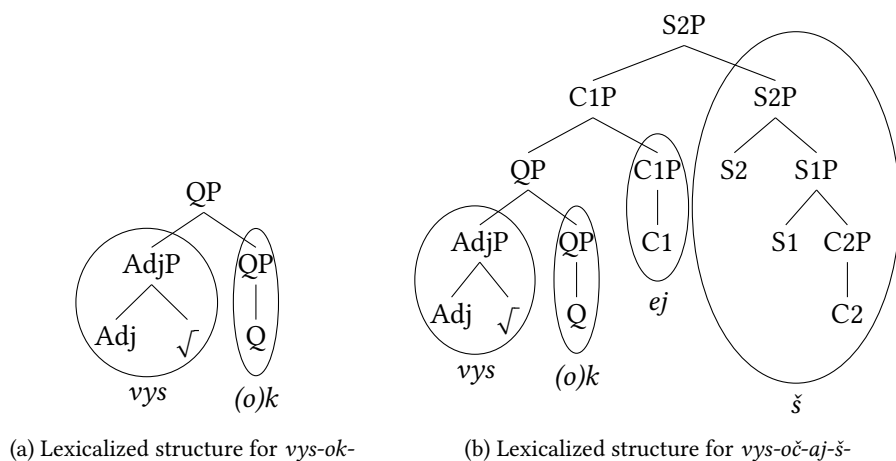
Figure 28

positing a movement-containing L-tree for the adjectival stem. The second phenomenon is the pattern of augment adjectives (the augment affix zero-exponence of morphosyntactic material in the comparative form only). When one combines the derivational steps necessary for the analysis of augment adjectives with a movement-containing L-tree for adjectival stems such as *vys-*, an ABA pattern emerges.

Some theorists may take the fact that the morphological theory used in this work generates ABA patterns as a matter of concern since many works (this paper included) have taken the impossibility of such patterns as the starting point of the investigation. However, recent research on similar (pseudo-)ABA patterns (Middleton 2021, Davis 2021) has come to the conclusion that the middle cell (the B of ABA) needs to be a portmanteau – this paper can be seen as adding to the body of evidence in favor of this idea.

5 Conclusion

In this paper, I have taken a look at the adjectival morphology of Russian through the lens of the comparative–superlative containment hypothesis put forth by Bobaljik (2012). I have provided evidence for there being a number of adjectives whose morphological behavior in the comparative and the superlative forms is problematic for contemporary proposals that follow Bobaljik's general idea.

Although the reported surface patterns may be taken as counter-evidence to Bobaljik's claims, I have argued that his core ideas need not be abandoned and

have proposed a Nanosyntactic analysis of the pattern building on the idea of Movement-Containing Trees (which are implied by the notion of phrasal spell-out but have been only recently argued for and used by Blix 2022) and the novel spell-out algorithm which allows subextraction from specifiers (Caha & Taraldsen Medová 2022, 2023).

Given that the morphological patterns discussed in this work are problematic for both Distributed Morphology and the standard version of Nanosyntax found in Baunaz & Lander (2018) and preceding work, it is possible to take the proposed analysis as an argument for accepting the generative power of the version of Nanosyntax with the subextraction algorithm presented in this work.

Abbreviations

AUG	augment	PL	plural
AGR	agreement	POS	positive
CMPR	comparative	SG	singular
F	feminine	SPRL	superlative
M	masculine		

Acknowledgments

References

Baunaz, Lena & Eric Lander. 2018. Nanosyntax: The basics. In Lena Baunaz, Liliane Haegeman, Karen De Clercq & Eric Lander (eds.), *Exploring Nanosyntax*, 3–56. Oxford: Oxford University Press.

Blix, Hagen. 2022. Interface legibility and nominal classification: A nanosyntactic account of Kipsigis singulatives. *Glossa: a journal of general linguistics* 7(1). 1–46. DOI: 10.16995/glossa.5825.

Blumenfeld, Lev. 2003. Russian palatalization and Stratal OT: Morphology and [back]. In Wayles Brown, Ji-Yung Kim, Barbara Partee & Robert Rothstein (eds.), *Annual Workshop on Formal Approaches to Slavic Linguistics: The Amherst meeting, 2002* (Formal approaches to Slavic linguistics 11), 141–158. Ann Arbor: Michigan Slavic Publications.

Bobaljik, Jonathan David. 2012. *Universals in comparative morphology: Suppletion, superlatives, and the structure of words.* Cambridge: MIT Press.

Bobaljik, Jonathan David & Uli Sauerland. 2018. ABA and the combinatorics of morphological features. *Glossa: a journal of general linguistics* 3(1). 15. 1–34. DOI: 10.5334/gjgl.345.

Bonet, Eulàlia & Daniel Harbour. 2012. Contextual allomorphy. In Jochen Trommer (ed.), *The morphology and phonology of exponence* (Oxford Studies in Theoretical Linguistics), 195–235. Oxford: Oxford University Press.

Caha, Pavel. 2009. *The nanosyntax of case.* Tromsø: University of Tromsø. (Doctoral dissertation).

Caha, Pavel. 2023. The layered structure of concord. In Pavel Caha, Pavel Rudnev & Svetlana Toldova (eds.), *The many facets of agreement* (LINCOM Studies in Theoretical Linguistics (LSTL) 68), 3–36. München: LINCOM GmbH.

Caha, Pavel, Karen De Clercq & Guido Vanden Wyngaerd. 2019. The fine structure of the comparative. *Studia Linguistica* 73(3). 470–521. DOI: 10.1111/stul.12107.

Caha, Pavel & Lucie Taraldsen Medová. 2022. *On the feminine singular declension of the Russian demonstrative.* Conference paper. 15th European Conference on Formal Description of Slavic Languages (FDSL 15). https://is.muni.cz/publication/2222757/fdsl15-dem.pdf.

Caha, Pavel & Lucie Taraldsen Medová. 2023. *Syncretism and markedness paradoxes in the Russian demonstrative declension.* Ms. Masaryk University. https://ling.auf.net/lingbuzz/007511.

Davis, Colin P. 2021. Case-sensitive plural suppletion in Barguzin Buryat: On case containment, suppletion typology, and competition in morphology. *Glossa: a journal of general linguistics* 6(1). 116. 1–26. DOI: 10.16995/glossa.5885.

Embick, David. 2015. *The morpheme: A theoretical introduction.* Berlin, München: De Gruyter Mouton.

Goncharov, Julie. 2015. *In search of reference: The case of the Russian adjectival intensifier 'samyj'.* Toronto: University of Toronto. (Doctoral dissertation).

Gouskova, Maria & Jonathan David Bobaljik. 2020. *Allomorphy and vocabulary insertion.* Ms. NYU and Harvard. https://ling.auf.net/lingbuzz/005388.

Halle, Morris. 1959. *The sound pattern of Russian: A linguistic and acoustical investigation.* The Hague: Mouton.

Halle, Morris & Alec Marantz. 1994. Some key features of Distributed Morphology. In Andrew Carnie, Heidi Harley & Toby Bures (eds.), *Papers on phonology and morphology* (MIT Working Papers in Linguistics 21), 275–288. Cambridge: MITWPL.

Kayne, Richard S. 2005. *Movement and silence* (Oxford Studies in Comparative Syntax). Oxford: Oxford University Press.

Kosheleva, Daria. 2016. Rival forms of comparatives in Russian. *Poljarnyj vestnik: Norwegian Journal of Slavic Studies* 19. 42–50. DOI: doi.org/10.7557/6.4059.

Marantz, Alec. 2001. Words. Handout, MIT. https://babel.ucsc.edu/~hank/mrg.readings/Marantz_words.pdf.

Matushansky, Ora. 2002. More of a good thing: Russian synthetic and analytic comparatives. In Jindřich Toman (ed.), *Formal Approaches to Slavic Linguistics 10: The second Ann Arbor Meeting 2001*, 143–161. Ann Arbor: Michigan Slavic Publications.

Middleton, Jane. 2021. Pseudo-ABA patterns in pronominal morphology. *Morphology* 31(4). 329–354. DOI: 10.1007/s11525-021-09377-7.

Norris, Mark. 2014. *A theory of nominal concord.* Santa Cruz: University of California. (Doctoral dissertation).

Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. *Nordlyd* 36(1: Special issue on Nanosyntax). 1–6. DOI: 10.7557/12.213.

Starke, Michal. 2018. Complex left branches, spellout, and prefixes. In Lena Baunaz, Liliane Haegeman, Karen De Clercq & Eric Lander (eds.), *Exploring Nanosyntax*, 239–249. Oxford: Oxford University Press.

Sudo, Yasutada & Andrew Nevins. 2022. No ABA patterns with fractionals. In Linnaea Stockall, Luisa Martí, David Adger, Isabelle Roy & Sarah Ouwayda (eds.), *For Hagit: A celebration* (QMUL Occasional Papers in Linguistics 47), 1–10. London: Department of Linguistics Queen Mary University of London. https://www.qmul.ac.uk/sllf/media/sllf-new/department-of-linguistics/hagit-borer-celebration/Sudo_and_Nevins.pdf.

Vanden Wyngaerd, Guido, Michal Starke, Karen De Clercq & Pavel Caha. 2020. How to be positive. *Glossa: a journal of general linguistics* 5(1). 23. 1–34. DOI: 10.5334/gjgl.1114.

Chapter 10

Two types of secondary imperfectives: Evidence from Polish and Bulgarian

Dorota Klimek-Jankowska[a] & Vesela Simeonova[b]

[a]University of Wrocław [b]University of Graz

Secondary imperfective (SI) morphology differs in its productivity in Polish (PL) and Bulgarian (BG): in PL, the SI morphology combines with some but not all prefixes. By contrast, almost every BG perfective verb has a SI variant. To our knowledge, there is no research that has attempted to get a closer understanding of the source of this discrepancy. To fill in this niche, we conducted a comparative study of the interaction of SI morphology with different classes of aspectual prefixes in PL and BG and the meaning effects they give rise to. We present novel observations and account for them by proposing that there are two distinct layers at which SI morphemes are generated in BG and only one such layer in PL.

1 Introduction

There is an ongoing debate in the literature on Slavic aspect concerning the status of aspectual morphemes. Little agreement has been reached as to the status of secondary imperfective (SI) /(y)v/ in Slavic languages. The views vary as to where /(y)v/ is generated in the structure (see Schoorlemmer 1995, Babko-Malaya 1999, Istratkova 2004, Milićević 2004, Svenonius 2004a,b, Romanova 2004, Filip 2005, Di Sciullo & Slabakova 2005, Arsenijević 2006, Romanova 2007, Ramchand 2008a,b, Łazorczyk 2010, Markova 2011, Tatevosov 2011, 2015, Biskup 2012, 2019, Wiland 2012, Žaucer 2012, Rothstein 2020, Klimek-Jankowska & Błaszczak 2022, 2023, Kwapiszewski 2022). In most of these studies, generalizations about the status of secondary imperfective morphology are made based on the data from a

single language. However, Slavic languages differ considerably in the productivity of forming SI verbs. For example, the Bulgarian (BG) secondary imperfective is considerably more productive as compared to Polish (PL): almost every BG perfective verb has a SI variant (see Markova 2011, Rivero & Slavkov 2014; Nicolova 2017: 5.3.25), while in PL, SI morphology combines with some but not all prefixed verbs (see Łazorczyk 2010, Łaziński 2011, 2020, Wiemer et al. 2020, Kwapiszewski 2022, Klimek-Jankowska & Błaszczak 2022, 2023). While it is a well-known observation, there are no works that attempt to explain the locus of variation. Our goal in this paper is to fill that gap. We provide two major novel empirical observations:

1. Within Bulgarian: while all previous works on SI in general and in Bulgarian specifically treat it as a single class, we identify two distinct SI classes within Bulgarian with systematically different sets of formal derivational and semantic properties.

2. Between languages: we identify the locus of cross-linguistic variation in the productivity of SI between Polish and Bulgarian: the two languages share one of the two classes of SI and Polish lacks the second class of SI that Bulgarian has.

We conclude that the difference in the productivity of SI in Polish and Bulgarian is not random, but is systematically determined based on the range of derivational possibilities in the two languages, with respective semantic consequences.

To arrive at these conclusions, we tested the interaction of SI morphology in PL and BG with two classes of prefixes – lexical and purely perfectivizing prefixes – and the meaning effects these different combinations of the tested aspectual prefixes with SI morphemes give rise to. In this paper, the empirical scope of the environments tested is limited to past tense contexts. We discuss possible extensions in §5.

We show that in BG there are two types of SI morphemes that bear different meanings; PL has only one of these SI morphemes. We propose that the two types of SI in BG are realized in two different syntactic layers, following Cinque's (1999) model.

2 Background on the secondary imperfective in Polish and Bulgarian

2.1 Similarities between Polish and Bulgarian

In both PL and BG, aspectual distinctions are encoded on almost all verbs.[1] The least morphologically complex aspectual forms are primary imperfectives (bare, i.e. "unprefixed'' verbs) and they can be perfectivized by means of a prefix, cf. (1). Some prefixes do not change the verb in any way other than its aspectual value; they are called PURELY PERFECTIVIZING or EMPTY prefixes (Bogusławski 1963, Svenonius 2004a,b, Młynarczyk 2004, Willim 2006, Ramchand 2008a).[2]

(1)   a. pisać   – napisać   PL
         write.IPFV   write.PFV

      b. piša   – napiša   BG
         write.IPFV   write.PFV

Another class of prefixes are the so-called LEXICAL PREFIXES (Babko-Malaya 1999, Svenonius 2004a,b, Romanova 2004, 2007, Ramchand 2008a,b, Biskup 2012, 2019, a.m.o.): they have an idiosyncratic meaning where the prefix changes the lexical interpretation of the verb, but not in a predictable way, for example the prefix *prze-* in (2) and the prefix *pod-* in (3) have very different meaning in various verbs they participate in. Lexical prefixes cause idiosyncratic changes in the meaning of a verbal predicate that is not derivable from either the verb or the prefix, cf. (2) and (3).[3]

---

[1]The encoding of aspectual distinctions can be blocked, e.g. for phonological reasons, or in certain loanwords, e.g. *printiram* 'print' only has one form in BG.

[2]Janda & Nesset (2010) emphasize that Russian has at least 16 prefixes forming natural perfectives (those perfectives which are not semantically distinct from the unprefixed base verb), which may suggest that they encode hidden distinctions. They propose that in the case of natural perfectives there is a semantic overlap between the meaning of the prefix and the meaning of the base verb and the diversity of prefixes used in natural perfectives follows from the fact that the base verbs from which they are derived fall into semantically diverse classes. Building on that, Janda & Lyashevskaya (2013) propose that the verbal prefixes act as classifiers in that they select verbs according to broad semantic traits, categorizing them the way numeral classifiers in some languages categorize nouns. We think that irrespective of the terminology used, there is a general consensus that the prefixes in natural perfectives do not modify the meaning of the base verbs but they may only impose selectional restrictions on the base verbs they combine with. Therefore, we will maintain the terminology 'purely perfectivizing'.

[3]Because of these properties, Romanova (2004), Svenonius (2004a,b), Ramchand (2004, 2008a,b), Łazorczyk (2010) argue that lexical prefixes are merged vP-internally.

(2) a. kupić   – *prze*-kupić   PL
         buy.ᴘꜰᴠ   bribe.ᴘꜰᴠ
         'to buy' – 'to bribe'

     b. grać   – *prze*-grać   PL
         play.ɪᴘꜰᴠ lose.ᴘꜰᴠ
         'play' – 'lose'

     c. łączyć   – *prze*-łączyć   PL
         connect.ɪᴘꜰᴠ switch.ᴘꜰᴠ
         'connect' – 'switch'

(3) a. seštam *(se) – *pod*-seštam   BG
         recall   ʀᴇꜰʟ   remind.ᴘꜰᴠ
         'recall' – 'remind'

     b. budja   – *pod*-budja   BG
         wake.up.ɪᴘꜰᴠ incite.ᴘꜰᴠ
         'wake up' – 'incite, instigate'

     c. igraja   – *pod*-igraja   BG
         play.ɪᴘꜰᴠ mock.ᴘꜰᴠ
         'play' – 'mock'

Furthermore, lexical prefixes (can) alter the argument structure/selectional restrictions of a verb, cf. (4)–(5).

(4) a. znać   {kogoś   / *uszkodzenia}   PL
         know.ɪᴘꜰᴠ someone   damage
         'to know someone'

     b. do-znać   {*kogoś   / uszkodzenia}
         suffer.ᴘꜰᴠ someone   damage
         'to suffer damage'

(5) mislja (*se) – za-mislja *(se)   BG
     think   ʀᴇꜰʟ consider   ʀᴇꜰʟ
     'think' – 'consider'

Lexically prefixed perfective verbs are imperfectivized by means of an *-yw-* or *-a-* suffix in Polish and by a *-va-* suffix or vowel alternations in Bulgarian, cf. (6). These imperfective forms derived from perfective verbs are called ꜱᴇᴄᴏɴᴅᴀʀʏ ɪᴍᴘᴇʀꜰᴇᴄᴛɪᴠᴇ (SI). Table 1 shows more verbs from this morphological pattern.

(6) a. podpisać – podpisywać   PL
         sign.PFV   sign.SI
         'sign' – 'sign'

     b. podpiša – podpisvam   BG
         sign.PFV   sign.SI
         'sign' – 'sign'

Table 1: Lexical prefixes and SI in Polish and Bulgarian

Polish		Bulgarian		English
PFV	SI	PFV	SI	
podpisać	podpisywać	podpiša	podpisvam	'sign'
odpowiedzieć	odpowiadać	otgovorja	otgovarjam	'reply'
naprawić	naprawiać	popravja	popravjam	'repair'
wyjaśnić	wyjaśniać	objasnja	objasnjavam	'explain'
sprzedać	sprzedawać	prodam	prodavam	'sell'
opisać	opisywać	opiša	opisvam	'describe'

2.2 A major difference: SI productivity

In the previous section, we showed that SI is possible with lexically prefixed verbs both in PL and in BG. However, there is a major difference between PL and BG in that almost every BG verb can form SI (Dickey 2000: 11; Nicolova 2017: 5.3.25). Most verbs with empty prefixes also have SI forms, as illustrated in (7b). In PL, empty prefixed forms cannot form SI, see the ungrammatical form in (7a). The pattern described in example (7) is systematic in the two languages, as demonstrated in Tables 2 and 3.[4]

(7) a. pisać   – napisać   – *napisywać   PL
         write.IPFV   write.PFV   write.SI

---

[4]The fact that one may find some rare instances of these verbs on the internet suggests that someone either used them creatively or mistakenly. Such rare uses may suggest that the two projections that we will argue for in §4 high SI and low SI are universally there in the hierarchy of projections but in some languages such as Polish, for example, the high SI morpheme generally does not merge in this position (it is blocked), but it may exceptionally be unblocked when used creatively or in speech production errors.

b. piša   – napiša   – napisvam   BG
    write.IPFV   write.PFV   write.SI

Table 2: Purely perfectivizing prefixes and SI in Bulgarian

Bulgarian			English
IPFV	PFV	SI	
stroja	postroja	postrojavam	'build'
piša	napiša	napisvam	'write'
půrža	izpărža	izpăržvam	'fry'
gladja	izgladja	izglaždam	'iron'
broja	prebroja	prebrojavam	'count'
molja	pomolja	pomolvam	'ask'
četa	pročeta	pročitam	'read'
gubja	izgubja	izgubvam	'lose'
merja	izmerja	izmervam	'measure'
zvănja	pozvănja	pozvănjavam	'call'
čupja	sčupja	sčupvam	'break'

Table 3: Purely perfectivizing prefixes and SI in Polish

Polish			English
IPFV	PFV	SI	
budować	zbudować	*zbudowywać	'build'
pisać	napisać	*napisywać	'write'
smażyć	usmażyć	*usmażywać	'fry'
prasować	wyprasować	*wyprasowywać	'iron'
liczyć	policzyć	*policzać	'count'
prosić	poprosić	*popraszać	'ask'
czytać	przeczytać	*przeczytywać	'read'
gubić	zgubić	*zgubiać	'lose'
mierzyć	zmierzyć	*zmierzać	'measure'
dzwonić	zadzwonić	*zadzwaniać	'call'
łamać	złamać	*złamywać	'break'

In other words, in BG, there is a morphological triplet for verbs with purely perfectivizing prefixes and a pair for lexically prefixed forms, as in Tables 2 and 4. And in PL, verbs form morphological pairs: either the bare imperfective and a verb with a purely perfectivizing prefix or the lexically prefixed imperfective and the derived SI, as in Tables 3 and 5. The crucial difference between the two languages in the two tables is marked with shading.[5]

Table 4: Purely perfectivizing vs. lexical prefixes and SI in Bulgarian

primary imperfective	perfective	secondary imperfective
N/A	podpiša 'sign'	podpisvam
N/A	poleja 'water'	polivam
stroja 'build'	postroja	postrojavam
piša 'write'	napiša	napisvam

Table 5: Purely perfectivizing vs. lexical prefixes and SI in Polish

primary imperfective	perfective	secondary imperfective
N/A	podpisać 'sign'	podpisywać
N/A	podlać 'water'	podlewać
budować 'build'	zbudować	*zbudowywać
pisać 'write'	napisać	*napisywać

While the morphological determinant of the restrictive SI in PL is well-known – the availability of SI counterparts of perfective verbs depends on the prefix type, as we described above – and it is also well-known that BG SI is fully productive, these observations raise many questions that remain unanswered to date. The questions we address in this paper are whether there are semantic differences between the SI forms in BG that do not have equivalents in PL and the ones that do have equivalents in PL and why the BG SI forms are not possible in PL.

---

[5]We acknowledge that the type of classification of triplets and pairs that we are using to make this claim is not the only one that exists in the literature. In a very recent study on aspectual triplets in Russian, Czech, Polish, Wiemer et al. (2020) identify triplets based on a different set of criteria. They assume that lexically prefixed verbs also form triplets. We explain how we understand triplets and pairs Section, §2.1 that we use in this paper are based on the works cited in §2.

3 Novel findings: Two types of SI

Our first finding is that the pair SI in PL is equivalent to the pair SI in BG and it is ambiguous between the single ongoing and habitual reading. The second finding is that within BG, the triplet SI is qualitatively different from the pair SI. The triplet SI is habitual only, while the pair SI is ambiguous between single ongoing and habitual reading. This means that the properties of SI are not uniform across languages and even within the same language they are not homogeneous.

3.1 Pair SI in Bulgarian and Polish

Both in Bulgarian and in Polish, SI forms derived from lexically prefixed verbs are ambiguous between an ongoing reading, as in (8), and a habitual reading, as in (9). This is not idiosyncratic of a specific verb, but holds across the morphological paradigm represented in Table 1.

(8) Ongoing context

    a. Kogato vljazoh   v ofisa na Ivan, toj (točno)
         when enter.PFV.AOR.1SG in office of Ivan he just
         {podpis-va-še   dokumenti / poprav-ja-še   koleleta /
          sign-SI-IMPF.3SG documents   repair-SI-IMPF.3SG bikes
         otgovar-ja-še   na imejli}.   BG
         reply-SI-IMPF.3SG to emails

    b. Kiedy weszłam   do gabinetu Jana, (właśnie)
         when entered.PFV.PST.1SG to office   John just
         {podpis-yw-ał dokumenty / naprawi-a-ł   rower /
          sign-SI-PST.3SG documents   repair-SI-PST.3SG bike
         odpowiad-a-ł   na maila}.   PL
         reply-SI-PST.3SG to email
         'When I entered John's office, he was (in the middle of) {signing documents / repairing bikes / replying to e-mails}.'

(9) Habitual context

    a. Predi obiknoveno {podpis-va-še   dokumentite /
         before usually   sign-SI-IMPF.3SG documents
         poprav-ja-še   koleletata / otgovar-ja-še   na imejli} po-bărzo
         repair-SI-IMPF.3SG bikes.DEF   reply-SI-IMPF.3SG to emails faster
         ot   men, no veče   ne.   BG
         than me,   but already not

b. Kiedyś zwykle {podpis-yw-ał   dokumenty / na-prawi-a-ł
before usually sign-SI-PST.3SG documents   repair-SI-PST.3SG
rowery / odpowiad-a-ł   na maile} szybciej niż ja, ale teraz
bikes   respond-SI-PST.3SG to emails faster   than I   but now
już   nie.   PL
already not
'In the past, usually he (used to) {sign (the) documents / repair bikes /
respond to e-mails} faster than me but not anymore.'

All the verbs in Table 1 behave in a way analogous to the pattern shown in
examples (8) and (9), allowing both habitual and ongoing readings. We were
unable to find any counterexamples.

3.2 Triplet SI in Bulgarian

The triplet SI in Bulgarian cannot be used with ongoing actions, cf. (10) and only
has habitual readings, cf. (11). The examples also show that the bare imperfective
is grammatical in both environments.[6]

(10)   ONGOING CONTEXT

a. Kogato telefonăt zvănna,   točno {păržeh   /
when   phone.DEF rang.AOR.3SG just   fry.IMPF.1SG.IPFV
*izpăržvah}   kjufteta.   BG
 fry.IMPF.1SG.SI meatballs
'When the phone rang, I was (right in the middle of) frying meatballs.'

b. Kogato telefonăt zvănna,   točno {gladeh   /
when   phone   rang.AOR.3SG just   iron.IMPF.1SG.IPFV
*izglaždah}   drehi.   BG
 iron.IMPF.1SG.SI clothes
'When the phone rang, I was (right in the middle of) ironing clothes.'

c. Kogato telefonăt zvănna,   točno si   {pišeh   /
when   phone   rang.AOR.3SG just   REFL.GEN write.IMPF.1SG.IPFV
*napisvah}   domašnoto.   BG
 write.IMPF.1SG.SI homework.DEF
'When the phone rang, I was (right in the middle of) writing my
homework.'

---

[6]While we assume that it is always available in ongoing contexts, we do not claim that it is
always possible in every habitual context.

(11)   Habitual context

  a. Kogato praveh   zakuska, obiknoveno {păržeh   /
      when   make.IMPF.IPFV.1SG breakfast usually   fry.IMPF.IPFV.1SG
      izpăržvah}   po   3 kjufteta   na čovek.   BG
      fry.IMPF.SI.1SG DISTR 3 meatballs per person
      'When I made breakfast, I used to fry 3 meatballs per person.'

  b. Predi vinagi {gladeh   / izglaždah}   drehite
      before always iron.IMPF.IPFV.1SG iron.IMPF.SI.1SG clothes.DEF
      vednaga   sled prane.   BG
      immediately after washing
      'Before I always ironed the clothes immediately after washing.'

  c. Predi   obiknoveno {pišeh   / napisvah}   po
      before usually   wrote.IMPF.IPFV.1SG wrote.IMPF.SI.1SG DISTR
      njakolko knigi na godina, no sega samo po   edna.   BG
      several   books per year,   but now only   DISTR one
      'In the past, I used to write several books per year, but now only one.'

This semantic pattern observed with triplet SI above is valid across the paradigm of triplet SI, a sample of which was presented in Tables 2 and 3. Since there is no triplet SI in Polish, this part of the data is not directly comparable between the two languages. In both single ongoing and habitual scenarios presented for Bulgarian in (10) and (11), Polish uses primary imperfective verbs only.

To summarize, while PL uses primary IPFV verbs to render both habitual and ongoing readings in purely perfectivized verbs and the SI forms of such verbs are blocked, BG productively uses the SI forms of those verbs to exclusively "mark" the special kind of habitual reading (consisting of a series of temporally non-overlapping bounded events happening on separate occasions).

Our novel observation is that the properties of SI morphology are not semantically uniform across the two languages. In the next section, we propose that the two types of SI morphemes merge at different syntactic positions in BG but not in Polish.

4 The syntax of SI

In order to formally capture the observations presented in the previous section, we propose that in BG, the two types of SI morphemes merge at two syntactic layers – one higher and one lower – while in PL, the low SI morpheme merges only in the lower one.

(12)   BG: SI_{high} >> SI_{low}
       PL: SI_{low}

This proposal allows us to syntactically distinguish between the properties of lexical prefixes and purely perfectivizing ones (see §2). We argue that SI is not a uniform category within Bulgarian because it merges in two different syntactic positions with different properties each. Moreover, pair SI is equivalent across Bulgarian and Polish because it is merged in the same projection SI_{low} with the same properties. PL does not have aspectual triplets because the high SI morpheme is blocked and it cannot merge in the SI_{high} layer while Bulgarian developed a specialized habitual meaning of the SI_{high} morpheme. Because verbs with SI_{high} have obligatory habitual readings in BG, which are missing in PL, we propose that the SI_{high} morpheme merges in a projection corresponding to Cinque's (1999) Asp_{HAB}.

We assume Baker's (1985) Mirror Principle, according to which syntax reflects morphology and vice versa and the linearization of functional morphemes is syntactically motivated. Additionally, we follow Cinque (1999), who argues that there is a fixed hierarchy of functional projections which regulates the way adverbs and functional morphemes are merged in syntax. Based on a large survey of languages, Cinque shows that among temporal/aspectual affixes, e.g. repetitive, frequentative, terminative, continuative, retrospective, durative, progressive, completive, those that are specifically dedicated to expressing habituality scope the highest (Cinque 1999: p. 56; 70). Crucially, this means that the dedicated habitual functional head is syntactically higher than the progressive aspectual head. The complete functional hierarchy is provided below in (13), in which the two functional heads are highlighted with boxes.

(13) $Mood_{speech\ act}$ > $Mood_{evaluative}$ > $Mood_{evidential}$ > $Mood_{epistemic}$ > $T_{(past)}$
       > $T_{(Future)}$ > $Mood_{irrealis}$ > $\boxed{Asp_{habitual}}$ > $T_{(Anterior)}$ > $Asp_{perfect}$ >
       $Asp_{retrospective}$ > $Asp_{durative}$ > $\boxed{Asp_{progressive}}$ > $Asp_{prospective}/Mod_{root}$ >
       $Voice$ > $Asp_{celerative}$ > $Asp_{completive}$ > $Asp_{(semel)repetitive}$ > $Asp_{iterative}$

Cinque (1999: 76)

Our examples in the preceding sections showed that this is the case also in BG and PL. In addition, we showed that the adverbs are optional in the case of SI_{high}, that is, it encodes the habitual reading itself rather than being merely compatible with it. SI_{low}, on the other hand, is compatible with both frequentative and ongoing adverbs just like a null IPFV operator, for example the one proposed

in Ferreira (2016) selecting for VPs referring to singular or plural events respectively: IPFV [VPsg / VPpl].[7]

In this way the proposal offers a formalization in syntactic terms which captures the differences in the morphological productivity of SI in the two languages, the within-language split (in BG), as well as the fact that the lower SI is equivalent in the two languages.

5 Discussion and conclusion

This study presented a formal description of a systematic difference in the productivity of SI in PL and BG: the Polish perfectives that do not allow subsequent secondary imperfectivization are precisely those cases where Bulgarian SI forms only have a habitual reading and the single ongoing reading is unavailable. These are the perfective forms which in Polish contain purely perfectivizing prefixes. By contrast, Polish perfectives that allow subsequent secondary imperfectivization are those cases where Bulgarian SI forms are ambiguous between a single ongoing and a habitual reading. These are the perfective forms which both in Polish and Bulgarian contain lexical prefixes.

Based on these novel observations we proposed that there are two distinct types of SI in BG: SI_{high} » SI_{low} and only one in Polish, SI_{low}. Crucially, we showed that SI_{low} is uniform in the two languages – it is ambiguous between a single ongoing and habitual reading. SI_{high} is merged in a projection corresponding to Cinque's (1999) Asp_{HAB} and Bulgarian syntax generates this position, while Polish does not.

Previous works on SI cannot capture the novel observations we present here. For example, Rivero & Slavkov (2014) assume that all SIs in BG have both habitual and ongoing readings. This is reflected in their formal account, attributing this duality to context. As we have shown, this is accurate for SI with lexical prefixes in BG and in Polish, but it overgenerates for SI with purely perfectivizing prefixes.

Conversely, Markova (2011) assumes (but provides no evidence) that the SI morphology in BG is in Cinque's (1999) Asp habitual projection. This has the opposite problem: it undergenerates the available ongoing interpretations of SI with lexical prefixes.

One limitation of this study is that the syntactic proposal put forth here still does not explain why it may be that one language is able to generate the higher

---

[7]We assume Tatevosov's (2011, 2015) proposal that aspectual morphology can be lower than the actual aspectual interpretation.

SI layer, while the other language is not. We leave such a comprehensive explanatory account for future work.

Additionally, one may reasonably ask why SI_{high} is not blocked by primary imperfective forms which can also express a habitual reading. In order to address this issue, we show below that in temporal *after*-clauses imposing sequential ordering between two events only SI_{high} is possible and simple imperfective is not, as shown in (14). By contrast in temporal *while*-clauses with two events temporally overlapping only simple imperfective is possible and SI_{high} is not, as shown in (15).

(14) Vseki păt sled kato {*?stroeše   / postrojavaše} kăšta, tja se
      every time after when   built.IPFV.3SG built.SI.3SG   house it.F REFL
      srutvaše.   BG
      collapsed.SI.3SG
      'Every time 'after' he built (=finished building) a house, it collapsed.'

(15) Vseki păt dokato { stroeše   / *postrojavaše} kăšta, imaše
      every time while   built.IPFV   built.SI.3SG   house have.IPFV.3SG
      incidenti.   BG
      incidents
      'Every time while he was building a house, there were incidents.'

One possible answer could be that there are two homophonous SI morphemes in Bulgarian, the one applying higher in syntax being specialized in expressing habituality consisting of a series of temporally non-overlapping bounded events (instead of serving the more general task of "undoing the perfectivizing contribution of the prefix"). Panini's Principle (also referred to as Elsewhere Principle), according to which the application of a specific rule or operation overrides the application of a more general rule, would then link the function of expressing habituality to the specialized SI_{high} form.

What remains to be studied in more detail is the interaction of SI morphology with different classes of superlexical prefixes. It is also necessary to extend the empirical scope to other Slavic languages in order to identify which languages pattern with Bulgarian and which ones pattern parametrically with Polish, and whether there are other possibilities. It also remains to be tested whether the observations we report in this study for past tense contexts can be extended to non-past tense as well.[8]

---

[8]We note that both present tense and imperfective aspect can have a multitude of meanings, e.g. non-actual readings, see Rivero & Slavkov (2014), Nicolova (2017): 364; due to this, a study of the interaction of SI and present tense deserves a longer paper that we leave to future work.

Finally, Bulgarian and Macedonian have two properties that other Slavic languages lack: they have definite articles and they have preserved the Imperfectum and Aorist tenses. To that end, it would be relevant to test whether the SI morphology interacts with Imperfectum and Aorist and behaves differently in the present and in the past, as well as whether number and referentiality of nominal complements impact the interpretation of either types of SI.

Abbreviations

1	first person	IPFV	imperfective
3	third person	PFV	perfective
AOR	aorist	PP	past participle
DEF	definite	PRS	present tense
DISTR	distributive	PST	past
IPFV	imperfective	REFL	reflexive
GEN	genitive	SG	singular
F	feminine gender	SI	secondary imperfective

Acknowledgments

References

Arsenijević, Boban. 2006. *Inner aspect and telicity: The decompositional and the quantificational nature of eventualities at the syntax-semantics interface* (LOT Dissertation Series 142). Utrecht: LOT.

Babko-Malaya, Olga. 1999. *Zero morphology: A study of aspect, argument structure and case.* Rutgers: The State University of New Jersey-New Brunswick.

Baker, Mark C. 1985. The mirror principle and morphosyntactic explanation. *Linguistic inquiry* 16(3). 373–415.

Biskup, Petr. 2012. Slavic prefixes and adjectival participles. In Markéta Ziková & Mojmír Dočekal (eds.), *Slavic languages in formal grammar: Proceedings of FDSL 8.5, Brno 2010*, 271–289. Frankfurt am Main: Peter Lang.

Biskup, Petr. 2019. *Prepositions, case and verbal prefixes: The case of Slavic.* Amsterdam: John Benjamins. DOI: 10.1075/la.255.

Bogusławski, Andrzej. 1963. Czasowniki dokonane i niedokonane i ich stosunki wzajemne. In Witold Doroszewski (ed.), *Z polskich studiów slawistycznych*, 153–159. Warszawa: Polska Akademia Nauk.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective.* Oxford: Oxford University Press. DOI: 10.1093/oso/9780195115260.001.0001.

Di Sciullo, Anna Maria & Roumyana Slabakova. 2005. Quantification and aspect. In Angeliek van Hout, Henriëtte de Swart & Henk Verkuyl (eds.), *Perspectives on aspect*, 61–80. Dordrecht: Springer. DOI: 10.1007/1-4020-3232-3_4.

Dickey, Stephen M. 2000. *Parameters of Slavic aspect: A cognitive approach.* Stanford: CSLI Publications.

Ferreira, Marcelo. 2016. The semantic ingredients of imperfectivity in progressives, habituals, and counterfactuals. *Natural Language Semantics* 24. 353–397. DOI: 10.1007/s11050-016-9127-2.

Filip, Hana. 2005. On accumulating and having it all: Perfectivity, prefixes and bare arguments. In Henk Verkuyl, Henriëtte de Swart & Angeliek van Hout (eds.), *Perspectives on aspect*, 125–148. Dordrecht: Springer. DOI: 10.1007/1-4020-3232-3_7.

Istratkova, Vyara. 2004. On multiple prefixation in Bulgarian. *Nordlyd, Tromso Working Papers on Language and Linguistics: Special Issue on Slavic Prefixes* 32(2). 301–321. DOI: 10.7557/12.71.

Janda, Laura A. & Olga Lyashevskaya. 2013. Semantic profiles of five Russian prefixes: *po-, s-, za-, na-, pro-. Journal of Slavic Linguistics* 21(2). 211–258. DOI: 10.1353/jsl.2013.0012.

Janda, Laura A. & Tore Nesset. 2010. Taking apart Russian *raz-. The Slavic and East European Journal* 54(3). 476–501. http://www.jstor.org/stable/23345086.

Klimek-Jankowska, Dorota & Joanna Błaszczak. 2022. The status of secondary imperfectivization in Polish: Evidence from VP idioms. *Journal of Slavic Linguistics* 30(FASL 29 issue). 1–19. DOI: 10.1353/jsl.2022.a923068.

Klimek-Jankowska, Dorota & Joanna Błaszczak. 2023. The interaction of idioms and aspect in Polish. *Natural Language & Linguistic Theory* 42. 153–207. DOI: 10.1007/s11049-023-09577-2.

Kwapiszewski, Arkadiusz. 2022. *Aspect and event structure: The morphosyntax of Polish verbs from a cross-linguistic perspective.* Oxford: University of Oxford. (Doctoral dissertation).

Łaziński, Marek. 2011. Polish aspectual prefixes, their order and functions: A study based on the National Corpus of Polish. *Word Structure* 4(2). 231–243. DOI: 10.3366/word.2011.0012.

Łaziński, Marek. 2020. *Wykłady o aspekcie polskiego czasownika.* Warszawa: Wydawnictwa Uniwersytetu Warszawskiego. DOI: 10 . 31338 / uw . 9788323542346.

Łazorczyk, Agnieszka A. 2010. *Decomposing Slavic aspect: The role of aspectual morphology in Polish and other Slavic languages.* Los Angeles: University of Southern California. (Doctoral dissertation).

Markova, Angelina. 2011. On the nature of Bulgarian prefixes: Ordering and modification in multiple prefixation. *Word Structure* 4(2). 244–271. DOI: 10.3366/word.2011.0013.

Milićević, Nataša. 2004. The lexical and superlexical verbal prefix *iz-* and its role in the stacking of prefixes. *Nordlyd* 32(2). 279–300.

Młynarczyk, Anna Katarzyna. 2004. *Aspectual pairing in Polish* (LOT Dissertation Series 87). Utrecht: LOT.

Nicolova, Ruselina. 2017. *Bulgarian grammar.* Berlin: Frank & Timme.

Ramchand, Gillian Catriona. 2004. Time and the event: The semantics of Russian prefixes. *Nordlyd* 32(2). 323–361. DOI: 10.7557/12.72.

Ramchand, Gillian Catriona. 2008a. Perfectivity as aspectual definiteness: Time and the event in Russian. *Lingua* 118(11). 1690–1715. DOI: 10.1016/j.lingua.2007.03.008.

Ramchand, Gillian Catriona. 2008b. *Verb meaning and the lexicon: A first-phase syntax*, vol. 116. Cambridge: University Press Cambridge. DOI: 10 . 1017 / CBO9780511486319.

Rivero, Maria Luisa & Nikolay Slavkov. 2014. Imperfect(ive) variation: The case of Bulgarian. *Lingua* 21. 232–277. DOI: 10.1016/j.lingua.2014.07.019.

Romanova, Eugenia. 2004. Superlexical versus lexical prefixes. *Nordlyd* 32(2). 255–278. DOI: 10.7557/12.69.

Romanova, Eugenia. 2007. *Constructing perfectivity in Russian.* Tromsø: University of Tromsø. (Doctoral dissertation).

Rothstein, Susan. 2020. Perfective is a V operator in Slavic, though not in English. *Current development in Slavic linguistics. Twenty years after (based on selected papers from FDSL 11).* 155–172.

Schoorlemmer, Maaike. 1995. *Participial passive and aspect in Russian.* Utrecht: OTS.

Svenonius, Peter. 2004a. Slavic prefixes and morphology: An introduction to the Nordlyd volume. *Nordlyd* 32(2). 177–204. DOI: 10.7557/12.67.

Svenonius, Peter. 2004b. Slavic prefixes inside and outside VP. *Nordlyd* 32(2). 205–253. DOI: 10.7557/12.68.

Tatevosov, Sergei. 2011. Severing perfectivity from the verb. *Scando-Slavica* 57(2). 216–244. DOI: 10.1080/00806765.2011.631782.

Tatevosov, Sergei. 2015. Severing imperfectivity from the verb. In Gerhild Zybatow, Petr Biskup, Marcel Guhl, Claudia Hurtig, Olav Mueller-Reichau & Maria Yastrebova (eds.), *Slavic grammar from a formal perspective: The 10th anniversary FDSL conference, Leipzig 2013*, 465–494. Frankfurt/M.: Peter Lang.

Wiemer, Björn, Joanna Wrzesień-Kwiatkowska & Marek Łaziński. 2020. Badania aspektu w językach polskim, czeskim i rosyjskim za pomocą korpusów i baz danych (pierwsze podsumowanie tematu). *Forum Lingwistyczne* 7. 45–58. DOI: 10.31261/FL.2020.07.04.

Wiland, Bartosz. 2012. Prefix stacking, syncretism and the syntactic hierarchy. In Markéta Ziková & Mojmír Dočekal (eds.), *Slavic languages in formal grammar: Proceedings of FDSL 8.5, Brno 2010*, 307–324. Frankfurt am Main: Peter Lang.

Willim, Ewa. 2006. *Event, individuation and countability: A study with special reference to English and Polish*. Kraków: Jagiellonian University Press.

Žaucer, Rok. 2012. The syntax of perdurative-prefixed verbs (and the VP-internal/VP-external prefix distinction). In Markéta Ziková & Mojmír Dočekal (eds.), *Slavic languages in formal grammar: Proceedings of FDSL 8.5, Brno 2010*, 339–355. Frankfurt am Main: Peter Lang.

Chapter 11

The (un)expectedly stacked prefixes in Slovenian

Franc Lanko Marušič, Petra Mišmaš & Rok Žaucer
University of Nova Gorica

When a Slavic verb occurs with multiple prefixes their order is often claimed to follow certain restrictions of a fairly formal character. Firstly, lexical prefixes, which can modify the argument structure of the verb and contribute idiosyncratic interpretations, are always found adjacent to the verbal root, while superlexical prefixes, which do not alter the argument structure and whose interpretative contribution is adverbial, can be stacked over the lexicals. And secondly, when multiple superlexicals stack on a verbal stem, they follow a fixed order. We set out to test these two generalizations with a corpus study. We find that there exist a number of verbs which seem to have more than one lexical prefix, in direct contradiction of the standard assumptions about prefixation.

1 Introduction

In Slovenian and in Slavic languages more generally, simplex verbs consist of a root, a theme vowel [TV] and a tense and agreement ending [T/AGR], and are typically imperfective (though this is not a rule, cf. e.g. the Slovenian perfective simplex verb *kupiti* 'to buy'). Verbs can also carry one or more prefixes, with the prefixed form generally being interpreted perfectively (unless imperfectivized through, for example, suffixation in the process called secondary imperfectivization [SI]). We demonstrate this for the verb *znati* 'to know' and some of its derivatives in Table 1.[1]

---

[1]Unless indicated otherwise, all examples in this paper are Slovenian.

Table 1: The various parts of the Slavic verb

prefix	prefix	root	SI	TV	T/AGR	Gloss
		zn		a	ti	'to know.IPFV'
	po	zn		a	ti	'to know.PFV'
	po	zn	av	a	ti	'to know.IPFV'
pre	po	zn		a	ti	'to recognize.PFV'
pre	po	zn	av	a	ti	'to recognize.IPFV'

Turning to verbal prefixes, these are, in general, all formally related to prepositions (e.g., *ob* 'by/next to', *pri* 'at', etc., cf. Matushansky 2002, Gehrke 2008, Caha & Ziková 2022, a.o.), but are often assumed to differ among themselves in terms of their position within the verbal domain. Typically, a distinction is made between so-called lexical and superlexical prefixes. The former are often seen as affixal prepositions functioning as VP-internal resultative secondary predicates, similarly to resultative particles in Germanic, the latter as affixal prepositions functioning as VP-external, INFL-level material, e.g., Ramchand (2004), Romanova (2004), Svenonius (2004), and each type is said to behave uniformly with respect to a number of properties. The tree in Figure 1 sketches the relevant positions. A more detailed overview is given in Section 2.
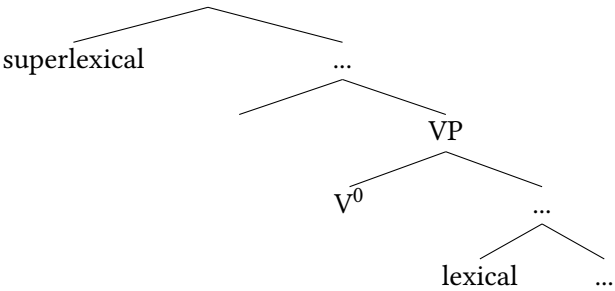
Figure 1: A sketch of the two positions of the two types of prefixes

One important distinction between the two types of prefixes that the literature often seems to convey (even if sometimes unintentionally) is that a verb will – generally – only have one lexical prefix, while superlexical prefixes can stack. The strong tendency that there will only be a single lexical prefix stems from the fact that there is a single position for lexical prefixes, as in Svenonius (2004) or Romanova (2004), or that the semantics of lexical prefixes preclude there being

more than one with a given verb, as in Babko-Malaya (2003). When both types of prefixes appear in a verb, the superlexical prefix(es) linearly precede the lexical prefix, and if a verb has multiple superlexical prefixes, these appear in a certain order (e.g., Milićević 2004, Istratkova 2006, Wiland 2012).

The main goal of this paper is to see if we can find a reflection of these generalizations in Slovenian corpus data, if we can use Slovenian corpus data to corroborate these generalizations about the lexical–superlexical division, in particular the view that stacked prefixes will generally not be lexical prefixes and that superlexical prefixes are governed by strict ordering constraints. If we find the generalizations reflected in corpus data, this can be seen as support for the theoretical claims; but note that if we do not, the claims can still be correct, as theoretical possibilities for the existence of specific structures per se do not necessarily imply anything about these structures' frequency in use.

Whereas we find that our corpus data are of limited use for testing fine-grained proposals for orderings of superlexicals, we do also find that they offer corpus support for some aspects of the ordering claims. At the same time, our corpus data also reveal some cases that may appear to be at odds with the expected division. Specifically, while isolated examples of verbs that seem to have two lexical prefixes have been pointed out in the past, e.g., *iz-pod-riniti* (lit.: from-under-drive) 'to push out' and *s-pod-makniti* (lit.: from-under-move) 'to jerk away' have been considered in Svenonius (2004: 242), and see also Markova (2011: 260) for Bulgarian and Biskup (2023: 20) for Russian and Czech, our corpus leads us to an expanded set of verbs that display this unexpected combination. Using this set of verbs we then consider how to analyze verbs in which two prefixes both exhibit properties typical of lexical prefixes.

The paper is organized as follows. In Section 2 we review some widely assumed properties ascribed to the two classes of prefixes. Section 3 presents a corpus study that focuses on stacked verbal prefixes. Section 4 discusses the data with potentially unexpectedly stacked prefixes, Section 5 presents the conclusions.

2 What we know: Lexical and superlexical prefixes in Slavic verbs

A fairly standard division of prefixes that is also characteristic of the more traditional literature (e.g. Vidovič Muha 1993, Toporišič 2000), and is typically assumed to hold for all Slavic languages, establishes two main uses of prefixes. Lexically used prefixes tend to have spatial or idiosyncratic meanings, where "idiosyncratic" is meant to capture situations in which the prefix's addition to the

verb does not lead to a systematically predictable interpretation of the prefix-verb stem complex, as shown in (1). With superlexically used prefixes, on the other hand, the addition of the same prefix predictably adds the same (adverbial) interpretation, and the interpretation of the verb stays transparent and constant across the prefixed verb class, (2).[2]

(1) ob-delati | ob-soditi | ob-noviti | ob-leteti
     at-work   at-judge   at-new   at-fly
     'to process' | 'to sentence' | 'to renew' | 'to fly around'

(2) po-sedeti | po-bingljati | po-plesati | po-igrati se
     over-sit   over-dangle   over-dance   over-play REFL
     'to sit for | 'to dangle | 'to dance | 'to play for a
      a while' | for a while' | for a while' | while'

The two classes are said to differ in a number of other properties. Lexical prefixes are said to appear directly on the verb root while superlexicals can be separated from the root by another prefix, and consequently, lexical prefixes can never be stacked, while there should be no such restriction, across the board, for superlexicals. Also, only lexical prefixes are said to be able to affect argument structure. And only lexical prefixes, but not superlexicals, can form secondary imperfectives, cf., e.g., Svenonius (2004: 229) for the diagnostics for superlexical prefixes, though note also that even for Svenonius some subclasses of superlexical prefixes can violate this last constraint (Svenonius 2004: 230). These properties are summarized in Table 2.

Table 2: Lexical and superlexical prefixes

Lexical prefixes	Superlexical prefixes
adjacent to the root	outside of lexical prefixes
idiosyncratic/PP meanings	adverbial meanings
affect argument structure	don't affect argument structure
form secondary impf.	don't form secondary impf.
generally don't stack	can stack

---

[2]For expository reasons, we ignore Slovenian orthography and separate prefixes from the rest of the verb with a hyphen. Prefixes are glossed on the basis of the basic meanings of their prepositional counterparts.

Many aspects of these generalizations, however, have also been questioned. Žaucer (2009), for example, shows that the cumulative prefix *na-* introduces an unselected object – generally considered a hallmark of lexicality – but can, at the same time, also stack over another prefix. A number of authors argued that the split should be in more than two groups: for example, Tatevosov (2008) argues for an independent, third class of *intermediate* prefixes; Babko-Malaya (2003) splits lexical prefixes in two groups; Markova (2011) proposes a four-part division into *outer*, *higher inner*, *lower inner*, and *lexical* prefixes (where the "traditional" lexical prefixes are split into *lower inner* and *lexical* prefixes).

2.1 Identity of prefixes

What is phonologically one and the same prefix can often be used as either a lexical or a superlexical prefix, as shown in (3)–(4). So if prefixes are defined with their phonological shape one should really only talk of their lexical or superlexical uses, rather than of lexical and superlexical prefixes.

(3)   a.   po-liti
          over-pour
          'to spill'

      b.   po-sedeti
          over-sit
          'to sit for a while'

(4)   a.   do-staviti
          to-put
          'to deliver'

      b.   do-od-pirati
          to-off-push
          'to finish opening'

Po- will standardly be analyzed as a lexical prefix resulting in a spatio-idiosyncratic interpretation on the verbal stem in (3a) and as a superlexical prefix with adverbial interpretation in (3b), and *do-* as a lexical prefix added to the verbal stem *staviti* (which never occurs on its own without a prefix in most varieties of Slovenian) and as a superlexical prefix added to an already prefixed stem in (4b).

Moreover, a prefix can have more than one superlexical use, as shown by the Polish example (5), where *po-* serves once as a delimitative and once as a distributive prefix (cf. also Žaucer 2009).

(5) Kucharze po-po-roz-kładali   przez chwilę naczynia i   zajęli
     cooks   po.DELIM-po.DIST-roz-put.SI over all   tables   and began
     się czymś   innym.
     REFL something else
     'The cooks put the dishes on the table for a while and they turned their
     attention to something else.'

<div align="right">(Polish; Klimek-Jankowska & Błaszczak 2022)</div>

2.2 Stacking

As mentioned above, it has been observed that when Slavic verbal prefixes stack, their ordering is not random, but rather reveals certain restrictions of a fairly formal character. For one, lexical prefixes attach to the verb before superlexical prefixes, and as a consequence, in any form with multiple prefixes, if the form includes a lexical prefix, the lexical prefix will appear closest to the verb, as sketched in (6). The other observation, also sketched in (6), is that superlexical prefixes (and only superlexical prefixes) can stack even over other superlexical prefixes so that a single verb can have more than one superlexical prefix but, normally, just one lexical prefix (cf. Romanova 2004, Svenonius 2004, Gehrke 2008) (though some authors, e.g. Tatevosov 2008, argue that Russian actually does not allow stacking of "genuine" superlexical prefixes (i.e., inceptive *za-*, delimitative *po-*, cumulative *na-* and distributive *pere-*) but only of "intermediate" prefixes, cf. above).

(6) superlexical prefix > superlexical prefix > lexical prefix > verb

The restriction to no more than one lexical prefix is taken to reflect the widely assumed general restriction to one independent resultative secondary predicate per verb (a.o. Rappaport Hovav & Levin 2001, Ramchand 2008), and suggests a further difference between lexical and superlexical prefixes.[3] Slavic lexical prefixes are parallel to resultative secondary predicates in languages like English, while superlexicals appear to be something different (cf. also Spencer & Zaretskaya 1998).

The superlexical prefixes are also said to follow a fixed order when stacked to the same verbal stem (Istratkova 2006, Wiland 2012, Endo & Wiland 2014, Klimek-Jankowska & Błaszczak 2022). For example, as claimed by Wiland (2012), who

---

[3]This restriction is sometimes also suggested to have a non-structural, conceptual explanation, e.g. the Single Delimiting Constraint in Tenny (1994) and Filip (2003). In this paper, we focus on the structural approach.

develops an even more fined-grained, cartography-inspired differentiation of su-
perlexical prefixes, the cumulative prefix *na-* needs to precede the completive
prefix *do-*, as shown in (7). Istratkova (2006) proposes the order in (8) for Bul-
garian, Wiland (2012) proposes the sequence in (9) for Polish, which was later
modified by Klimek-Jankowska & Błaszczak (2022) to (10).

(7) a. na-do-kładaj   sobie jeszcze
         CUML-COMPL-put self   more
         'get yourself some more (e.g. food)'

     b. *do-na-kładaj sobie jeszcze   (Polish; Wiland 2012)

(8) ATT > INCP > COMPL > DIST > CUML > EXC > REP
     po   *za*   *iz*   *po*   *na*   *raz*   *pre*
                                              (Bulgarian; Istratkova 2006)

(9) DIST > ATT > DELIM > SAT > CUML > EXC > REP > PERD > COMPL > TERM
     po   *pod* *po*   *na*   *na*   *na*   *prze* *prze*   *do*   *od*
                                                  (Polish; Wiland 2012)

(10) DELIM > DIST > SAT > CUML > {PERD, EXC, REP, ATT, TERM, PURELY PFV}
     po   *po*   *na*   *na*   *prze prze prze pod od*   *s/na*
                       (Polish; Klimek-Jankowska & Błaszczak 2022)

3 Corpus-study results

In order to get better empirical insight into multiply prefixed verbs in Slovenian,
we considered two sets of data. First, we looked at the 3000 most common verbs
in Slovenian using the *WeSoSlav* database (see Arsenijević et al. 2024), to explore
the behavior of common verbs with more than one prefix in general (assuming
that such a 3000-verb sample is representative of the language). In the second
step we created a list of multiply prefixed verbs from the list of all verbs occurring
in the *Gigafida 2.0* reference corpus of written standard Slovenian (Čibej et al.
2019).

Starting with *WeSoSlav*, while we were able to confirm that multiple prefixa-
tion exists, we found that only 6 out of 3000 verbs had 3 prefixes (no verbs have
more), 178 verbs had 2 prefixes, while 2,076 had a single prefix.[4] Table 3 gives the

---

[4]The 6 verbs with three prefixes include two aspectual pairs (i.e. *s-po-raz-umeti* 'to agree/com-
municate.PFV', *s-po-raz-umevati* 'to agree/communicate.IPFV' and *s-po-pri-jeti* 'to cope/deal
with.PFV' *s-po-pri-jemati* 'cope/deal with.IPFV') so that there are really only 4 different verbs

relevant results.[5] Note that each verb was counted only once (that is, verbs with three prefixes were not counted also as verbs with one prefix and as verbs with two prefixes).

Table 3: Prefixation in WeSoSlav (Arsenijević et al. 2024)

number of prefixes	number of verbs	percent
zero	740	24.67%
(exactly) one	2,076	69.2%
(exactly) two	178	5.93%
(exactly) three	6	0.2%
Total	3,000	

This data leads us to certain conclusions. On the one hand, prefixed verbs are more common than verbs without prefixes (the latter are not necessarily simplex, since some have a suffix, e.g. *kup-ova-ti* 'to buy.ɪᴘꜰᴠ'). But more importantly, while verbs with a single prefix are extremely common, multiple prefixation is not. Given the relatively low number of multiply prefixed verbs, no proper quantificational analysis of the relative order of prefixes can be conducted. In order to create a better empirical base for investigating multiple prefixation, we created a larger list of multiply prefixed verbs.

3.1 Corpus-study results, additional data

The new set of data was created from the list of all 90,000+ verbs found in the *Gigafida 2.0* corpus (Čibej et al. 2019). We only looked at verbs that had more than 5 occurrences in the corpus as the number of typos, misspelled words and incorrectly classified words only increases with less frequent strings of characters. Prefixed verbs were automatically extracted from the list using a simple formula that looked at each individual verb and checked whether it begins with one of the prefixes. The prefix was subtracted from the verb and the verb was

---

with three prefixes. Applying this same aspectual-pair exclusion criterium also to verbs with two and with one prefix, there are only around 125 different verbs with two prefixes and around 1500 different verbs with a single prefix.

[5]Verbs that have a non-Slavic prefix like *re-* in *re-organizirati* 'to reorganize' or *dis-* in *dis-kvalificirati* 'to disqualify' were counted as unprefixed. Similarly we also disregarded the negative prefix *ne-*, as in *o-ne-sposobiti* 'to disable'.

checked again if the remaining part of the verb starts with one of the listed prefixes. This procedure was repeated five times. The automatically extracted multiply prefixed verbs were then also checked manually, since in some cases the automatic procedure counted some beginnings of stems/roots as prefixes, as in the case of verbs like *stati* (incorrectly analyzed as *s-tati*) 'to stand' or *vleči* 'to pull' (incorrectly analyzed as *v-leči*), and some combinations of prefixes could be misparsed as combinations of different prefixes, e.g. *pod-o-...* 'under-about-...', which is string-homophonous with *po-do-...* 'over-to-...', etc.

Table 4: Prefixation in the expanded database

number of prefixes	number of verbs	percent
zero	4,186	29.45%
one	9,181	64.58%
two	833	5.86%
three	16	0.11%
Total	14,216	

With this procedure we were able to retrieve a list of 849 multiply-prefixed verbs that exhibit at least 5 occurrences in the corpus. As above, the list contains some aspectual pairs, see footnote 4, but we did not exclude aspectual pairs for the figures we made. Verbs with three prefixes are extremely rare in Slovenian (see Section 3.2), and among the verbs with at least 5 occurrences in the corpus, there were no verbs with more than three prefixes.

In Figure 2 the prefixes are ordered on the basis of their likelihood, increasing from left to right, to appear as the prefix closest to the verb. The first thing to note is that no prefix is restricted to the root-adjacent position: in the presented set of verbs they all appear in the first position of a pair of prefixes at least once.

This last observation is very clearly visible also from Figure 3. Even the prefixes *pod-* 'under-' and *vz-* 'up-', which can be, based on Šekli (2016), taken as essentially exclusively lexical prefixes in Slovenian, appear stacked over another prefix in up to 20% of the cases. Actually, even the prefixes which seem to be most common in the root-adjacent position (*vz-* 'up-', *v-* 'in-', *ob-* 'around-', *pod-* 'under-' according to Figure 2 and Figure 3) also appear stacked over another prefix in at least 10% of the cases. Thus, all prefixes that are possible in the root-adjacent position can also be used as stacked prefixes (cf. Łaziński 2011 for a similar dictionary-based result from Polish) and thus – according to the descrip-
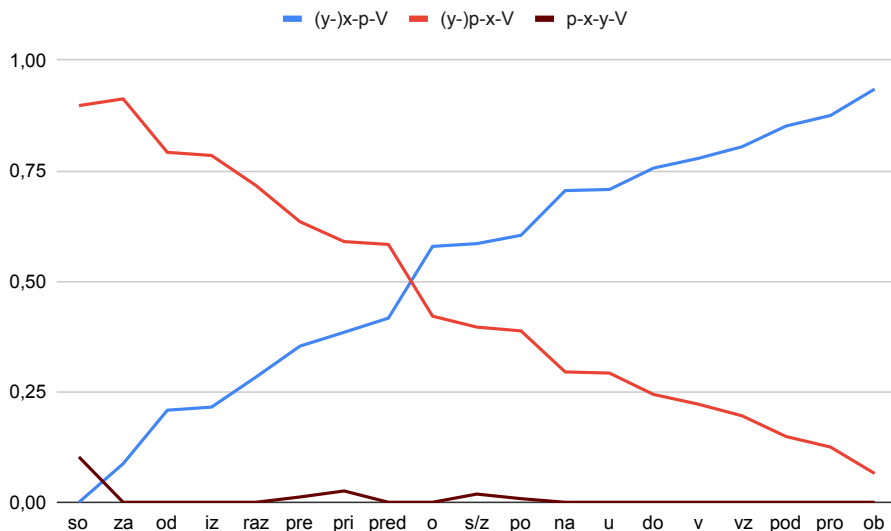
Figure 2: The frequency of prefixes relative to their position in a multiply prefixed verb (counting tokens of combinations)

tion so far – as superlexical prefixes. The implication does not go both ways, as *so-* 'co-' is never used as verb-adjacent in multiply prefixed verbs.

Table 5 confirms a tendency for a hierarchy, but does not confirm a true hierarchy. Most pairs of two prefixes only exist in one order, as is evident from the fact that the lower left half of Figure 5 has fewer cells filled in than the upper right half of the table. Table 5 shows that many pairs of prefixes exist with both orders of prefixes, so for example, there are 10 different verbs with the sequence *za-pre-* 'for-over', and 4 different verbs with this sequence reversed, i.e. *pre-za-*. Given that certain prefixes have more than one use, that is, that they can be either used as lexical or superlexical prefixes, one would need to determine case by case whether the second prefix of a sequence of two prefixes is indeed an instance of a superlexical prefix or a lexical prefix (which means coding your data on the basis of previous qualitative data analysis, which we wanted to avoid here as much as possible). Further, some prefixes have even more than one superlexical use (cf. Wiland 2012, Klimek-Jankowska & Błaszczak 2022), so that they can appear in more than one position within the proposed hierarchy of superlexical prefixes. These two facts presumably explain why we find so many different combinations where both orders of the two prefixes are possible, and we can only conclude that automatic extraction of prefixes cannot produce a clear sequence of superlexical
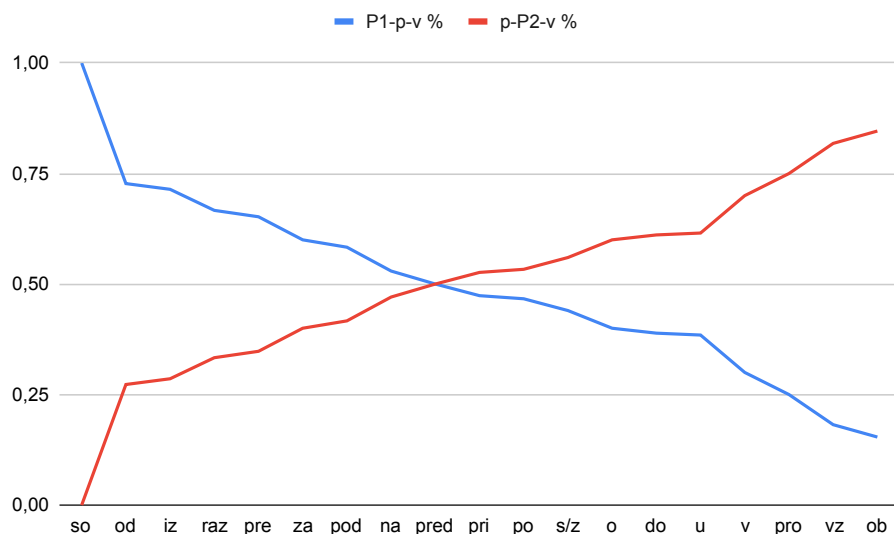
Figure 3: Relative amount of prefixes that a prefix can appear with either when it comes first or second in a pair of prefixes (counting types of combinations).

Table 5: The cross-table of prefix combinations. The first prefix of a pair is listed vertically, the second horizontally. The order of prefixes is a slightly modified sequence from Figure 2.

	so	od	pred	raz	iz	za	pre	o	pri	s/z	po	na	u	v	do	vz	pod	pro	ob
so		3	2			1	3	1	1	4	12	3			4		2	1	2
od						1		1	2		9	1			1	2	2		
pred							1	1			2	3							
raz							2	3	5		15			2			1	7	3
iz							6	4	4		11	4	1		1	2	14		4
za		1					10	9	5	127	20		7	3	8	4		3	12
pre	1		6	5	4			8	5	6	10	27	14	3	1	4		2	12
o						1	4		2	4	11		2						
pri							5	1			14	11	1		5	7		1	3
s/z		1		2		36		5	1		35	2			1		41	8	2
po	1		9	5	5		2	6		18		14	10	8	2	13	2		2
na				1	1				1		5		1	7	4	2			1
u								2	5		9						2		1
v									2		5				1				
do				2	1	1	2	1			2								1
vz											4				5				
pod			1	1		2					1	2	2			2			
pro				2	2										1	1			
ob				2							1								

prefixes, and therefore none of the proposed orders can be either confirmed or rejected.

If we assume that, generally, only the prefix closest to the root will potentially be a lexical prefix (see Section 1), we would need to look at verbs with at least three prefixes to be able to get a sequence of superlexical prefixes, but we only have 16 verbs with three prefixes to work with.

3.2 Verbs with three prefixes

Given that prefixes should be able to stack, and that quite some claims have been made on the basis of the possible and impossible ordering patterns in stacking, we expected that we will find substantial numbers of verbs with three or more prefixes. However, this prediction was not confirmed since out of 849 multiply-prefixed verbs no verb included more than three prefixes and only 16 included three prefixes. Specifically, a closer review of the 16 verbs showed that this number is actually even smaller, as "deduplication" of aspectual pairs reduces the number to a mere 10 verbs, listed in (11)–(20).[6]

Moreover, even some of the 10 verbs in (11)–(20) are odd-looking and unknown to us, such as *priopoteči* in (17), but as these verbs' few occurrences in the corpus seem to exhibit similar uses, we did not exclude them manually.[7,8]

(11) pre-raz-po-rediti[PFV] – pre-raz-po-rejati[IPFV]
      over-from-over-order   over-from-over-order
      'to rearrange'

(12) s-po-pri-jeti[PFV] – s-po-pri-jemati[IPFV]
      with-over-at-hold   with-over-at-hold
      'to tackle'

---

[6]Why are verbs with three or more prefixes so rare in actual language is a question we leave for future work. In discussing the rarity of some predicted orders of superlexical stacking, Markova (2011: 269) suggests that this might have to do with processing constraints.

[7]The verb *prisprehoditi* has 5 occurrences in *Gigafida 2.0* and *priopoteči* has 6 occurrences, and these 5/6 occurrences even include more than one example by the same author, so these are possibly forms that have been used/coined by two or three speakers. *Posprehoditi* has 27 occurences and *porazporediti* 30 occurrences in *Gigafida 2.0*. With the exception of *posprehoditi*, none of these are listed in any of the dictionaries available to us; the translations we provide for these verbs are thus our context- and form-based inferences.

[8]One could perhaps also exclude verbs with the prefix *so-* (similar to the English *co-*), such as (14) and (15). This prefix behaves differently from other verbal prefixes in several respects, can also appear in non-verbal contexts, e.g. *so-avtor* 'co-author', and is consequently often not even included in works on verbal prefixation, e.g. Vidovič Muha (1993).

(13) s-pre-ob-rniti^PFV   – s-pre-ob-račati^IPFV
     with-over-around-turn   with-over-around-turn
     'to convert'

(14) so-u-po-rabiti^PFV – so-u-po-rabljati^IPFV
     co-in-over-use   co-in-over-use
     'to co-use'

(15) so-po-vz-ročiti^PFV – so-po-vz-ročati^IPFV
     co-over-up-hand   co-over-up-hand
     'to co-cause'

(16) s-po-raz-umeti^PFV   – s-po-raz-umevati^IPFV
     with-over-from-understand   with-over-from-understand
     'to agree/communicate'

(17) pri-o-po-teči
     at-around-over-run
     'to get somewhere staggering'

(18) po-raz-po-rediti
     over-from-over-order
     'to distribute'

(19) po-s-pre-hoditi
     over-with-over-walk
     'to take a brief walk'

(20) pri-s-pre-hoditi
     at-with-over-walk
     'to get somewhere taking a walk'

Ignoring the interpretation of individual prefixes, we can extract several partial orders of prefixes from the above examples. Partial orders are given in (21). Interestingly *s/z-* 'with-' and *po-* 'over' appear in both orders, which is not surprising if both *po-* and *s/z-* have more than one superlexical use and thus more than one position in the hierarchy of superlexical prefixes.

(21) so > u, po
     s/z > po, pre > raz
     pri > o, s/z
     po > raz, s/z

But what seems to be going on is probably something else. A closer look at the verbs in (11)–(20) reveals that actually none of them seems to have a sequence of two obvious superlexical prefixes, and that for some of them no prefix seems very much like a standard, VP-external-looking superlexical prefix. In verbs like (12) and (13) all three prefixes have some of the properties of lexical prefixes – they affect the argument structure or have spatial PP meanings.

Automatic extraction of prefixes out of a list of verbs has limitations, and even though we were able to show that there is a tendency for a hierarchy, we did not arrive at a single order of superlexical prefixes; we were just able to show that there are certain prefixes that prefer to stay closer to the root and others that prefer to be further away, and that this preference is different for different prefixes, but different methods of establishing this preference gave different sequences of prefixes.

We will devote the remainder of this paper to the observed unexpected sequences of prefixes. As mentioned above, even the prefixes that have been suggested as being exclusively lexical appear in up to 20% of cases as the first prefix in a sequence of prefixes. Consider the verb *vz-po-staviti* 'to set up'. The prefix *vz-* generally has the meaning 'up' and is rarely associated with an adverbial meaning (e.g. *vz-ljubiti* 'to start to love') that we generally expect with the outermost prefix of a verb with two prefixes – certainly such a meaning is absent in *vz-po-staviti*. Similarly, the inner prefix of *vz-po-staviti*, as expected, has a meaning that can only be associated with a lexical prefix ('over'). This type of verbs – which we will call *vz-po-staviti*-type verbs – is what we turn to in Section 4.

4 Examples with two seemingly lexical prefixes

Considering the mainstream view in the literature on prefixation (Section 1), one expectation is that if a verb has two (or more) prefixes, at most one will tend to be a VP-internal, lexical prefix, while the rest will tend to be superlexical (or intermediate). However, our corpus study presented in Section 3 turned up a sizeable number of multiply prefixed verbs in which the outermost prefix also contributes a typically lexical meaning (i.e., *vz-po-staviti*-type verbs). Examples (22) to (28) give a sample of such verbs. These examples are presented here in triplets: The first form is the unprefixed version, the last is the relevant example with two prefixes, and the middle example is the form (which is always an attested form) with a single prefix. Example (28) stands out somewhat, as it has three seemingly lexical prefixes, and the version of the verb with just two is not attested in modern Slovenian (though it is attested in older versions of Slovenian). We

use a #hashtag to mark unprefixed forms that are unattested in modern standard Slovenian and also in many dialects, such as #*staviti*, though they are attested in some present-day dialects of Slovenian, in closely related BCMS, or are historically attested. Note also that in *vz-po-staviti*-type verbs, the verb with a single prefix always seems to exist, which makes these different from Žaucer's (2002) examples like (30), discussed in Svenonius (2004), in which the version with a single prefix is not attested.[9]

(22) klicati | po-klicati | v-po-klicati
      call   over-call   in-over-call
      'to call' | 'to call up' | 'to enlist'

(23) #staviti | po-staviti | vz-po-staviti
       set   on-set   up-on-set
                | 'to set' | 'to set up/establish'

(24) #jeti | pri-jeti | o-pri-jeti
       grab   at-grab   around-at-grab
              | 'to grab' | 'to hold on to'

(25) #peti | vz-peti | po-vz-peti
       pull   up-pull   on-up-pull
             | 'to climb' | 'to climb'

(26) #deti | o-deti | raz-o-deti
       put   around-put   from-around-put
             | 'to wrap' | 'to reveal'

(27) nesti | za-nesti | pri-za-nesti
      carry   behind-carry   at-behind-carry
      'to carry' | 'to carry in' | 'to spare'

(28) #umeti | raz-umeti |#po-raz-umeti | s-po-raz-umeti
       get/understand   apart-get   over-apart-get   with-over-apart-get
                          | 'to understand' | | 'to agree'

---

[9]Regarding (24): some varieties do exhibit a verb *jeti*, but only with an aspectual meaning 'to start'. While this is the same root, with the aspectual meaning having developed from the root's basic meaning 'grab'/'hold' or 'take' (Snoj 2009), it is not the root's meaning that the prefixed verb is based on, so we mark *jeti* in (24) with a hashtag.

(29) nesti | pri-nesti | do-pri-nesti
      carry   at-carry   to-at-carry
      'to carry' | 'to bring' | 'to contribute'

(30) riniti | *pod-riniti | iz-pod-riniti
      push   under-push   from-under-push
      'to push' | | 'to push out'

The meaning contribution of the outermost prefix suggests that these examples contain more than one lexical prefix. In (22) and (23) the addition of *v-* and *vz-*, respectively, leads to an idiosyncratic, or perhaps spatial meaning; in (24) the prefix *o-* adds a spatial meaning; in (25) the contribution of *po-* is not very clear (little discernible meaning change compared to its singly prefixed input); in (27), *pri-* adds an idiosyncratic meaning; etc. This situation is suprising in view of the idea that lexical prefixes generally do not stack.

The question is, then, how these prefixes should be analyzed. Possible answers include: (i) they are, despite their meanings, VP-external superlexicals; (ii) they fall into one of the additional categories of prefixes described in the literature (cf. Babko-Malaya 2003, Tatevosov 2008, Markova 2011, etc.); (iii) they are indeed VP-internal lexicals, but can be stacked because some special conditions are met. The last option then further opens several possibilities that could be explored, such as the possibility that these examples, in a sense, only include one prefix (and the inner prefix is somehow incorporated into the root), or that these are in fact two prefixes which either appear in a double-VP structure with two independent ResultPhrase positions for lexical prefixes, that they are result modifiers, or that they even require a completely different approach, perhaps one in which all prefixation is introduced above the VP (cf. Biskup 2023). In what follows, we explore these options.

4.1 Option 1: They are superlexical

If the outer prefixes of the *vz-po-staviti*-type verbs were instances of *v*P-external, superlexical prefixes, then one would expect them to exhibit properties typical of superlexical prefixes. One such property is their placement and the ability to stack – since they appear on top of a prefix they could, in principle, be taken as superlexical.

However, there are arguments against this claim. Firstly, they do not carry typical superlexical, adverbial meanings. If we consider the verb *pri-za-nesti* in (27), adding the prefix *pri-* results in an idiosyncratic meaning shift from 'to

carry in' to 'to spare', which cannot be the result of one of the two possible adverbial readings that *pri-* has, according to Šekli (2016), namely, a delimitative or an inchoative reading, as in *pri-preti* 'open a little' and *pri-žgati* 'to light up', respectively.

Also, superlexical prefixes are typically said not to allow secondary imperfectivization (see Section 2). Except for *vpoklicati*[PFV] 'to conscript' in (22), all other verbs given in (22)–(30) have well-attested secondary imperfectives: *vzpostavljati*[IPFV] 'to establish', *oprijemati*[IPFV] 'to hold on to', *povzpenjati*[IPFV] 'to climb', *razodevati*[IPFV] 'to reveal', *prizanašati*[IPFV] 'to spare', *sporazumevati*[IPFV] 'to communicate', *spozabljati se*[IPFV] 'to forget oneself'.[10] It should be emphasized that these do not seem to be cases of a prefix combining with an imperfective base – if this were the case, the resulting verb should be, contrary to fact, perfective. Rather, the imperfectivized verbs match the meaning of the perfective form (except in aspect), suggesting that these are in fact imperfectivizations of the doubly prefixed verbs:

(31) a. Veter je {za-nesel[PFV] / za-našal[IPFV]} listje na dvorišče.
      wind AUX behind-carry behind-carry leaves.ACC on yard
      'The wind carried leaves to the yard.'

   b. *Veter je {pri-za-nesel[PFV] / pri-za-našal[IPFV]} listje na
      wind AUX at-behind-carry at-behind-carry leaves.ACC on
      dvorišče.
      yard
      Literally: 'The wind spared leaves to the yard.'

(32) a. *Sodišče ni { za-neslo[PFV] / za-našalo[IPFV]} osumljencem.
      court NEG.AUX behind-carry behind-carry suspects.DAT
      Literally: 'The court didn't carry to the suspects.'

   b. Sodišče ni {pri-za-neslo[PFV] / pri-za-našalo[IPFV]} kriminalcem.
      court NEG.AUX at-behind-carry at-behind-carry criminals
      'The court didn't spare the criminals.'

And finally, according to Svenonius (2004) superlexical prefixes normally do not appear in nominalizations, in particular root/zero nominalizations (cf. also Caha

---

[10]The fact that at least for many speakers, *vpoklicati*[PFV] 'to conscript' does not have a natural imperfective counterpart is not problematic, given that it is also not the case that every perfective verb with a single prefix has a secondary imperfective counterpart, e.g., *za-bresti*[PFV] 'to get stuck' does not. In fact, the input of *vpoklicati*[PFV], i.e., *poklicati*[PFV] 'to call up', also does not have a secondary imperfective counterpart.

& Ziková 2016). While it should be noted that not all verbs in Slovenian derive root nominalizations, several of these *vz-po-staviti*-type verbs do:

(33) iz-po-staviti | iz-po-stav-a   (/ iz-po-stav-e)
     out-over-stand out-over-stand-F.SG.NOM   out-over-stand-F.SG.GEN
     'to single out' | 'branch'

(34) do-pri-nesti | do-pri-nos-∅   (/ do-pri-nos-a)
     to-at-carry   to-at-carry-M.SG.NOM   to-at-carry-M.SG.GEN
     'to contribute' | 'contribution'

(35) za-pri-seči | za-pri-seg-a   (/ za-pri-seg-e)
     behind-at-reach behind-at-reach-F.SG.NOM   behind-at-reach-F.SG.GEN
     'to pledge' | 'pledge'

(36) v-po-klicati | v-po-klic-∅   (/ v-po-klic-a)
     in-over-call   in-over-call-M.SG.NOM   in-over-call-M.SG.GEN
     'to call in, enlist' | 'conscription'

Root nominalizations are usually assumed not to contain structure above the VP, and following Svenonius (2004), the existence of root nominalizations can be taken as an argument that these prefixes are structurally similar to lexical prefixes, merged inside the verb phrase.

The only reason to consider the outermost prefix in the verbs under discussion to be superlexical, then, would be their placement, whereas their other properties speak against their being superlexical. In what follows, we will therefore further explore the option that they are not superlexical.

4.2 Option 2: They are neither lexical nor superlexical

Whereas a binary split into VP-internal lexical prefixes and a possibly internally diverse group of superlexical prefixes is the most common stance taken in the literature (present also in several cartography-like accounts such as Wiland 2012), some authors have proposed systems with more than two circumscribed groups of prefixes. In this section, we consider whether the stacked prefix in our *vz-po-staviti*-type verbs could belong to one of these additional classes, and conclude that it could not. Note that we will always leave the highest-merging prefix type of these systems out of the discussion: that the stacked prefix in our *vz-po-staviti*-type verbs cannot be any of these highest merging types follows from the discussion in Section 4.1.

Tatevosov (2008) analyzes lexical prefixes as merging in a result phrase inside the VP and superlexical prefixes as merging outside the vP. He suggests that between the lexical and the superlexical prefixes there is a third group – intermediate prefixes, such as the Russian completive *do-* – which merges somewhere above the VP and below the superlexicals.

While *vz-po-staviti*-type verbs share certain properties with verbs with intermediate prefixes (e.g. being able to be imperfectivized), they also have characteristics that set them apart. According to Tatevosov (2008), intermediate prefixes (among other characteristics) yield compositional meanings and never influence argument structure. As we already saw in Section 4.1, the outermost prefix in *vz-po-staviti*-type verbs can lead to non-compositional meanings, such as *vz-* in (23) ('to set' > 'to establish') or *pri-* in (27) ('to carry in' > 'to spare'), which come with concomitant argument structure effects (shown with more detail in Section 4.3). As was also already mentioned in Section 4.1, *vz-po-staviti*-type verbs often serve as the basis for root nominalizations, as in (33), which following Svenonius (2004) also suggests that their prefixes do not originate above the VP. We therefore conclude that our *vz-po-staviti*-type verbs are not simply intermediate prefixes.[11]

In a similar vein, Markova (2011) presents an account in which lexical prefixes, which she merges inside the VP as head adjuncts to V^0, are joined by three groups: outer prefixes, which are above vP; higher inner prefixes, which originate between VP and *v*P; and lower inner prefixes, which originate in a PathP complement to V^0.

Given that Markova's (2011) higher inner prefixes are positionally the same as Tatevosov's (2008) intermediate prefixes, the same arguments that we just presented against viewing the stacked prefix in *vz-po-staviti*-type verbs as Tatevosov's intermediate prefixes will also apply to the possibility that these prefixes would be Markova's higher inner prefixes. At the same time, the stacked prefixes in *vz-po-staviti*-type verbs will also not be Markova's lower inner prefixes, since she reserves this position for spatial and causative prefixes, whereas the stacked prefixes in a number of our *vz-po-staviti*-type verbs are neither spatial nor causative: see again, for example, (27). Also, Markova's lower inner prefixes cannot contribute idiosyncratic meanings, which she reserves for lexical prefixes, but the stacked prefixes in our *vz-po-staviti*-type verbs can contribute idiosyncratic meanings.

Note, however, that somewhat in passing, Markova (2011: 260) also mentions the possibility that a verb hosts two lexical prefixes, in a V^0 combining two

---

[11]In the spirit of Žaucer (2013), an argument could also be made on the basis of relative scope with respect to VP adverbials, the restitutive 'again' and adverbs of completion, all of which scope over the outer prefix. For a demonstration of some of this, see Section 4.3.3 below.

prefixes and a verb, that is, in a V^0 to which two prefix heads have been adjoined. From what we can tell, this structure, which assumes the possibility for idiosyncratic meanings for both prefixes, can actually successfully derive our *vz-po-staviti*-type verbs. Though Markova does not mention this, her account probably also predicts the possibility that a verb hosts a lexical prefix as well as a stacked lower inner prefix, a structure that presumably can derive some of our *vz-po-staviti*-type verbs. We return to this in Section 4.3.1.

Another account that proposes more than two groups of prefixes was put forth in Babko-Malaya (2003). As a version of the superlexical category, Babko-Malaya has Aktionsart-prefixed verbs, in which the prefix merges outside the VP (for which see Section 4.1). In addition, she has lexically prefixed verbs, in which the prefix is adjoined to V^0, and resultatively prefixed verbs, in which the prefix (itself part of a complex head) is adjoined to V^0. As explained by Babko-Malaya (2003: 27) herself, the semantics derived from those structures is such that double prefixation is only possible when a stacked prefix is an Aktionsart prefix (i.e., a superlexical prefix in the terminology from Section 4.1), while it actually prevents double prefixation with either two lexical prefixes, two resultative prefixes, or a combination of the two. So the stacked prefixes in our *vz-po-staviti*-type verbs will clearly be neither the lexical nor the resultative prefixes of Babko-Malaya (2003).

Note, however, that as pointed out by a reviewer, the account from Babko-Malaya (2003) is presumably not incompatible with the existence of stacked prefixes of the type of *vz-po-staviti* if such stacked prefixes are analyzed as result modifiers in the sense of Žaucer (2013) (even though Babko-Malaya herself does not discuss this type of data). This would be a version of the view that these stacked prefixes are VP-internal, lexical prefixes, which is the option we discuss next, having determined now that our *vz-po-staviti*-type verbs can be neither superlexical nor intermediate, or something of the sort.

4.3 Option 3: They are lexical

If prefixes in *vz-po-staviti*-type verbs are VP-internal lexical prefixes, we expect them to exhibit properties typically ascribed to lexical prefixes. Again, an argument against such an analysis is that the prefixes under discussion stack, while for lexical prefixes it is assumed that they generally do not stack, see Section 1 and Section 2. The explanation for this restriction is structural. Because lexical prefixes are generally assumed to be resultative and originate in a VP-internal Result Phrase [RP], as shown in Figure 4 (based on Svenonius 2004: (80)), and because verbal structure is assumed to be able to host only one result/one RP

(Rappaport Hovav & Levin 2001, Ramchand 2008), it should normally not be possible to have more than one lexical prefix per verb.[12]
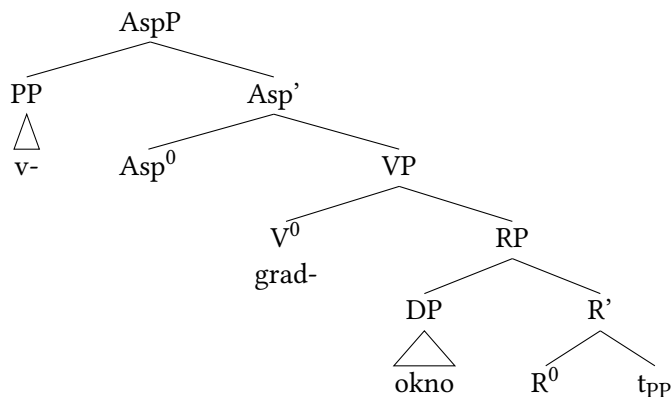
Figure 4: Structure dictates the one-lexical-prefix restriction.

However, as already indicated in Section 4.1, these prefixes display several other properties that can be taken as arguments for a VP-internal, lexical analysis. In addition to non-superlexical interpretations, the availability of secondary imperfectivization and root nominalizations, the outer prefixes in *vz-po-staviti*-type verbs also exhibit some argument-structure effects.

For example, the "singly" prefixed verb *pri-jeti* 'to grab' can select for a single accusative object, the reflexive clitic SE, and an optional genitive object, or an optional reflexive clitic SE and a prepositional phrase, as shown in (37). The "doubly" prefixed *o-pri-jeti* 'to hold on to', on the other hand, is unacceptable (in most modern varieties) with a single accusative object, requires the genitive object with a reflexive clitic SE, and simply does not tolerate prepositional objects, as shown in (38). Differences in the argument structure of the singly- and doubly-prefixed counterparts are observed also in other cases, as shown in (39)–(41).

(37) a. pri-jeti ročaj
          at-grab handle.ACC
          'to grab the handle'

      b. pri-jeti se   (ročaja)
          at-grab REFL handle.GEN
          'to grab (on to the handle)'

---

[12]Though see den Dikken (1995) for a different understanding of the structure used for particles and prefixes and the restrictions it imposes.

    c. pri-jeti (se)   za ročaj
        at-grab REFL for handle.ACC
        'to grab on to the handle'

(38) a. * o-pri-jeti   ročaj
          around-at-grab handle.ACC

    b. o-pri-jeti   se   *(ročaja)
        around-at-grab REFL   handle.GEN
        'to grab on to the handle'

    c. * o-pri-jeti   (se) za ročaj
         around-at-grab REFL for handle.ACC

(39) a. pri-seči (*pričo)
        at-reach   witness.ACC
        'to swear, take an oath'

    b. za-pri-seči   (pričo)
        behind-at-reach witness.ACC
        'to take an oath; to swear in a witness'

(40) a. za-nesti   skrbi   Vidu
        behind-carry worries.ACC Vid.DAT
        'to carry worries to Vid'

    b. * za-nesti   Vidu (s   skrbmi)
         at-behind-carry Vid.DAT with worries
         'to carry worries to Vid'

(41) a. pri-za-nesti   (*skrbi)   Vidu
        at-behind-carry   worries.ACC Vid.DAT
        'to spare Vid'

    b. pri-za-nesti   Vidu (s   skrbmi)
        at-behind-carry Vid.DAT with worries
        'to spare Vid (the worries)'

Given that we seem to be led to the conclusion that the outer prefix in *vz-po-staviti*-type verbs is a lexical prefix, it should be noted that different authors have previously observed that VP-internal prefixes are not a homogeneous group. A natural question to ask, then, is whether the outer prefixes in *vz-po-staviti*-type verbs share any of the properties of those proposed subgroups.

4.3.1 Option 3.1: They are lexical – but these verbs contain only one prefix

This option presents itself as a possibility especially in view of the fact that some of these apparently doubly-prefixed verbs are no longer used without a prefix. For example, while (42) exists in some Slovenian dialects (and in BCMS), it does not exist in standard Slovenian, nor in many other dialects that normally use *vz-po-staviti*. Similarly, (43) does not exist in modern Slovenian (though it does exist in BCMS), and neither does (44).

(42) #*staviti* 'set' (exists in some Western Slovenian dialects)

(43) #*peti* 'pull' (but exists in BCMS)

(44) *jeti 'grab'/'hold'

Given that these simplex forms are not attested (or are at best very limited) synchronically, it could be the case that the innermost prefix, even if historically a prefix, is just a part of the root (cf. Fowler 1996), or in other terms, as suggested in Markova (2011: 260) for all prefixes resulting in idiosyncratic meaning shifts, is adjoined to V^0, forming a complex verbal head. According to this analysis, a verb can have more than one lexical/X^0-adjoined prefix, and since prefixes are adjoined to v^0, they are freely ordered.

On the one hand, it seems to us that Markova's proposal could be seen as consistent with *vz-po-staviti*-type verbs, especially for those built on verbs like *po-staviti* 'to set' or *pri-jeti* 'to grab', whose unprefixed bases are not attested syncronically, as well as for those whose outer prefix seems somehow related to a spatial use, such as in *v-po-klicati* 'to enlist'. On the other hand, for a number of *vz-po-staviti*-type verbs aspect presents an issue. Several of these verbs, such as *vz-peti* 'to climb.PFV', are based on stems that were historically imperfective, and just like most lexically prefixed verbs (and unlike most native unprefixed verbs), these verbs generally form secondary imperfectives, e.g. *po-stavljati* 'to stand.IPFV', *vz-penjati* 'to climb.IPFV', *pri-jemati* 'to hold.IPFV'. This suggests that these inner prefixes trigger perfectivity. It is unclear to us how such adjunction could account for the change of aspect. In Svenonius's (2004) account, for example, the perfectivizing effect arises when a prefix moves from the RP into a VP-external aspect projection; if the prefix is part of a complex V^0, such movement does not seem to be possible. For those *vz-po-staviti*-type verbs which exhibit singly-prefixed counterparts even in modern Slovenian, such as *v-po-klicati* 'to enlist' or *za-pri-seči* 'to take an oath, to swear somebody in', this aspectual concern regarding treating their inner prefix as V^0-adjoined is even more obvious.

In addition, whereas some of these *vz-po-staviti*-type verbs synchronically do not exhibit unprefixed versions, they do occur in a modern Slovenian with several different prefixes, (45)–(47), resulting in forms with either clearly related or with idiosyncratic meanings. We can take this as an argument against an analysis on which the innermost prefixes are simply part of the root: While we agree with Romanova (2004), who considers similar examples of "cranberry roots" in Russian, that these roots are light (according to Romanova they can have no semantics at all), a comparison of the same root with different prefixes implies some common meaning (for (45), this could be paraphrased as 'to place') while the prefixes add a predictable spatial meaning.

(45) *na-staviti* | *po-staviti* | *v-staviti* | *pre-staviti* | *do-staviti* | *od-staviti* ...
     on-set   over-set   in-set   over-set   to-set   from-set
     'set' | 'set' | 'insert' | 'move' | 'deliver' | 'remove'

(46) *na-peti* | *vz-peti* | *v-peti* | *raz-peti* | *pri-peti* | *od-peti*   ...
     on-pull   up-pull   in-pull   apart-pull   at-pull   from-pull
     'stretch'/'string' | 'climb' | 'fasten' | 'spread' | 'attach' | 'detach'

(47) *na-jeti* | *pri-jeti* | *za-jeti* | *ob-jeti* | *vz-eti*   ...
     on-grab   at-grab   behind-grab   around-grab   up-grab
     'hire' | 'grab' | 'scoop' | 'hug' | 'take'

And finally, assuming that the forms in (45)–(47) are unprefixed poses a problem for the varieties in which the simplex forms of the verbs in (45)–(47) do exist, and it also does not account for those *vz-po-staviti*-type verbs that are perfectly normally attested both in standard Slovenian and across Slovenian dialects without the prefix (e.g., *klicati* 'to call', the root of the doubly prefixed verb *v-po-klicati* 'to enlist'). We thus conclude that despite some merits, Markova's account falls short of fully explaining our *vz-po-staviti*-type verbs.

4.3.2 Option 3.2: They are lexical – but these verbs have two VPs (=double resultative structure)

As mentioned in Section 4.3, the restriction to a single lexical prefix per verb has been derived as a consequence of the structural position of lexical prefixes; because the clausal structure can only have one RP, there can normally only be one lexical prefix per verb phrase (and consequently per verb). However, Žaucer (2009) discusses a class of verbs in Slovenian that seem to have two resultative prefixes, and ultimately analyzes these as having a double-VP structure (cf. also

Tatevosov 2022). In the discussion of the cumulative (/accumulative/saturative) prefix *na-*, a crucial piece of support for the double-VP structure is argued to be the two sets of unselected objects, (48) and (49).

(48) *(pre)-igrati[PFV] Maradono
     over-play   Maradona.ACC
  'fake out Maradona'

(49) *(na)-*(pre)-igravati[PFV/IPFV] se   Maradone
     on-over-play   REFL Maradona.GEN
  'get / getting one's fill of faking out Maradona'

As is evident from our examples in Section 4.3, the *vz-po-staviti*-type verbs do not behave like this. They do not appear to introduce two unselected objects.

Furthermore, the outermost prefix in (49) and this type of examples require an imperfective input, which is not the case in *vz-po-staviti*-type verbs. Also, (49) and this type of examples are normally read perfectively, with the outermost prefix there triggering perfectivity; in other words, an example such as (49) does not necessarily get an imperfective reading despite the presence of the imperfective suffix *-ava*. At the same time, though, the imperfective affix *can* be interpreted as scoping over the outermost prefix – in this case the interpretation of (49) is 'getting one's fill of faking out Maradona'. Unlike (49), and as shown in (50), the outermost prefix of *vz-po-staviti*-type verbs never perfectivizes its input and the imperfective affix always scopes over the outermost prefix, which further means that the whole verb is interpreted as imperfective.

(50) a. pri-jeti[PFV] – pri-jemati[IPFV] | o-pri-jeti[PFV]   – o-pri-jemati[IPFV]
      at-grab   at-grab.SI   around-at-grab around-at-grab.SI
      'to grab' | 'to grab on to'

    b. pri-nesti[PFV] – pri-našati[IPFV] | do-pri-nesti[PFV] – do-pri-našati[IPFV]
      at-carry   at-carry.SI   to-at-carry   to-at-carry.SI
      'to carry to'   'to contribute'

    c. po-staviti[PFV] – po-stavljati[IPFV] | iz-po-staviti[PFV] – iz-po-stavljati[IPFV]
      over-stand   over-stand.SI   out-over-stand   out-over-stand.SI
      'to set' | 'to single out'

While Žaucer (2009) discusses other properties of examples that can be analysed as including two VPs, we take these differences as evidence enough to conclude that prefixes in *vz-po-staviti*-type verbs are not similar to the cumulative *na-*.

4.3.3 Option 3.3: They are lexical – result modifiers, not main result predicates

The literature has identified one further group of prefixes that does not fully respect the standard division into lexical and superlexical. As discussed by Žaucer (2013), prefixes such as excessive (*pre-*), repetitive (*pre-*), attenuative (*pri-*, *po-*), and distributive (*po-*) have adverbial, superlexical-like meanings, can stack, and do not affect argument structure at least when stacked, which makes them look like ordinary superlexical prefixes. An example of this type of prefix is given in (51).

(51) pre-na-polniti
      over-on-fill
      'overfill'

However, Žaucer (2013) argues, contrary to what would be expected given the properties listed above, that these prefixes nevertheless merge VP-internally, supporting this claim, for example, with the fact that they scope below VP-adverbials, as shown in (52). The proposed analysis is that these prefixes are result modifiers, thus a sort of adverbial prefixes, but ones that modify the result phrase directly, before it is merged together with the verb.[13]

(52) U-stekleničil sem tole vino sicer na roke, pre-u-stekleničil ga bom pa
      in-bottled   AUX this wine PTCL on hand over-in-bottled   it   will PTCL
      z   mašinco.
      with machine
      'Though I bottled this wine manually, I'll re-bottle it with a machine.'
                                                              (Žaucer 2013: 292)

What (52) says is that the first time the wine was bottled it was bottled manually, while the second time it was bottled this was done with the use of a machine, which indicates that the repetitive *pre-* is inside the scope of the 'with'-adverbial, which, in turn, means that *pre-* does not originate above the VP.

   Interestingly, the same scopal facts can be observed with *vz-po-staviti*-type verbs. As shown in (53) the entire verb *oprijeti* 'to hold on to' is in the scope of the 'with'-adverbial, suggesting that all parts of the verb originate VP-internally.

---

[13]As already mentioned, this is a possibility not considered by Babko-Malaya (2003), whose analysis explicitly rules out stacked lexicals and resultatives, but it is, as pointed out to us by a reviewer, a possibility that is in fact perfectly compatible with that system.

(53) Vejo   sem sicer pri-jel z   roko, o-pri-jel   se   je bom pa
     branch AUX PTCL at-hold with hand around-at-hold REFL it AUX PTCL
     z   rokavico.
     with glove
     'I grabbed the branch with my hand, but I'll hold on to it with a glove.'

The two sets of prefixes also behave the same with respect to the restitutive reading of *spet* 'again'. That is, both the excessive/measure prefix in (54) and the outer prefix in *vz-po-staviti* 'establish' in (55) take narrow scope with respect to the restitutive reading of *spet* 'again'.

(54) Juš je   hladilnik spet pre-na-polnil.
     Juš AUX fridge   again over-on-filled
     'Juš restored the fridge to an overfilled state.'
     Not: Juš was overly involved in filling up the fridge.   (Žaucer 2013: 293)

(55) Miha je   stike   z   očetom spet vz-po-stavil.
     Miha AUX contacts with father again up-over-set
     'Miha restored contacts with his father.'
     (No other interpretation.)

While Žaucer's (2013) result-modifying prefixes have a predictable adverbial interpretation and the outer-most prefixes in *vz-po-staviti*-type verbs do not seem to, both of these types of prefixes behave comparably with respect to scopal tests, suggesting that they share the same structural position.[14]

4.3.4 Option 3.4: They are lexical and parallel to particles

It is well known that there exist parallels between Germanic particles and Slavic prefixes, e.g. Spencer & Zaretskaya (1998), Svenonius (2004). In fact, similarly to doubly-prefixed verbs of the *vz-po-staviti*-type verbs in Slovenian, we can also observe particle recursion in Germanic, see for example den Dikken (1995: 80). den Dikken (1995) claims that particle recursion is structurally possible but, for unclear reasons, rare. He analyzes recursive particles using his basic structural template from Figure 5 by simply having the second particle as the head of XP, as in Figure 6.

---

[14]Žaucer (2013) does not discuss nominalization possibilities, but root nominalizations from verbs with those result-modifying prefixes are not difficult to find, e.g, *pri-vz-dig* 'a partial lift', *pre-u-stroj* 'remodeling', *pre-u-redba* 'reorganization'. The same holds also of our *vz-po-staviti*-type verbs, cf. (33)–(36) above.

(56)   I'll send the letter on over to Grandma's house.

den Dikken (1995: (116b)), quoting Di Sciullo & Klipple (1994)
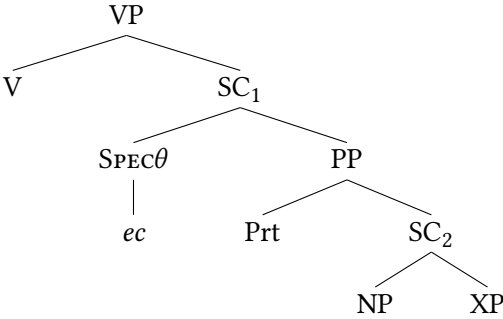
Figure 5: The basic structural template of den Dikken (1995)
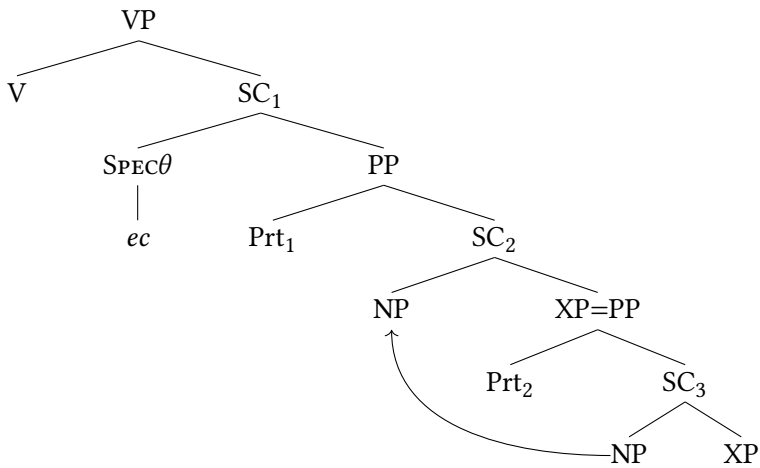
Figure 6: Using den Dikken's (1995) basic structural template to explain unexpected multiple prefixation

4.3.5 Option 3.5: They are some of the lowest projections above VP

There is yet another set of accounts that we have not discussed, namely, accounts that merge all prefixes, including lexical ones, outside the VP. One part of these accounts is represented by systems which at least implicitly still subscribe to two groups, lexical prefixes and a group of higher prefixes, with a single slot for lexical prefixes (e.g. Slabakova 2005, Istratkova 2006, Wiland 2012); like the accounts

discussed above, with lexical prefixes originating VP-internally, these accounts thus generally also end up with a restriction to a single lexical prefix. In addition, it is also not clear to us that such systems can really explain argument structure effects of lexical prefixes well, cf. Žaucer (2009: 16–18). Most recently, Biskup (2023) also develops a system with all prefixes merged outside the VP, but his version presumably allows more flexibility than the previous all-prefixes-outside-the-VP accounts as it does not really seem to subscribe to two groups, and it does not limit the number of lexical prefixes structurally but rather by appealing to conceptual reasons; for a similar case as our *vz-po-staviti*-type verbs, it explicitly allows two lexical prefixes hosted in two separate internal-prefix phrases above the VP. The approach looks promising to us for approaching our *vz-po-staviti*-type verbs, however, in addition to the concern regarding argument-structure effects already stated above, it is also not clear to us – assuming a universal clausal spine – what the nature of the lexical-prefix projections introducing the multiple lexical prefixes could be, and why they could be freely remergeable.

4.4 Instead of a conclusion—a partial proposal

We have shown that the outer prefixes in *vz-po-staviti*-type verbs, even though they are stacked on top of another prefix, do not behave like other superlexical prefixes but rather much more like VP-internal, lexical prefixes. Table 6 presents a comparison of our *vz-po-staviti*-type verbs, or rather, their outer prefixes, lexical prefixes, superlexical prefixes and result-modifying prefixes on the basis of the six most typically considered properties. Some of these properties are clearly related to one another, so for example, a prefix's VP-internal position is related to its ability to form a secondary imperfective, which is merged outside the VP and thus scopes over it. Similarly, as already explained in Section 4.3, placing lexical prefixes in a dedicated VP-internal Result Phrase means that a verb should not host a stack of such prefixes. Additionally, idiosyncratic meaning and argument-structure effects of lexical prefixes also seem to be related to their position inside the VP.

So far we mentioned 12 different *vz-po-staviti*-type verbs that used 10 different prefixes as the outer prefix. Most likely, then, the outer prefixes of *vz-po-staviti*-type verbs do not form a homogeneous class of prefixes, so we actually need not expect to find a single explanation for all of them.

The type of verbs that had been discussed by Žaucer (2002) and Svenonius (2004), *iz-pod-riniti* 'to push out' and *s-pod-makniti* 'to jerk away', are probably just instances of a complex prefix which realizes both Path and Place parts of

Table 6: Lexical, superlexical, and other types of prefixes

	Lexical	vz-po-staviti	result mod.	Superlex.
VP-positioning meaning	internal idiosyn./spati.	internal idiosyn./spati.	internal adverbial	external adverbial
affect arg. struct.	Yes	Yes	No	No
form sec. imperf.	Yes	Yes	Yes	No
form root nomin.	Yes	Yes	Yes	No
stacking	No	Yes	Yes	Yes

the preposition phrase inside a single result phrase, as suggested by Svenonius (2004).[15]

Some prefixes have a relatively clear spatial meaning, such as *o-* in *o-pri-jeti* 'hold on to', which is comparable in meaning to verbs where *o-* is more clearly lexical like *o-kleniti* 'grab on to', *o-graditi* 'to put a fence around', or *o-črtati* 'to draw a line around' (in some cases the (core) spatial meaning got obscured by a more metaphorical interpretation) and *v-* in *v-po-klicati* 'to enlist', which can even be doubled by a preposition phrase with the same prefix, as in (57).

(57) Trener ga je   v-po-klical v reprezentanco.
      coach him AUX in-over-call in national-team
      'The coach called him up into the national team.'

In cases like these, the outer prefix may seem to be a proper lexical prefix that would require a result phrase of its own, which would mean that we need two RPs inside the VP, which seems like a problem – but cf. Markova (2011) and Biskup (2023). Note that even though these verbs have a different argument structure from their unprefixed counterpart, the contribution of the prefix to the argument-structure change is not very clear, suggesting that potentially one of the two prefixes can receive an alternative interpretation.

In many respects, our *vz-po-staviti*-type verbs seem to behave similarly to doubly-prefixed verbs in which the prefixes are "result modifiers", the main difference being the interpretation of prefixes/prefixed verbs – while the "result mod-

---

[15]The two combinations *iz-pod-* and *s-pod-* are synonymous. One can find both versions of these two verbs in written Slovenian – *iz-pod-riniti* and *s-pod-riniti* both with the same meaning 'to push out' and likewise *s-pod-makniti* and *iz-pod-makniti* both meaning 'to jerk away'. Spoken Slovenian hardly makes a distinction between the two pronunciations of these two forms, so we are treating them as just two realizations of the same lexical unit.

ifiers" in Žaucer (2013) have a clear adverbial reading, prefixes in *vz-po-staviti*-type verbs lead to anything between a slight modification in the interpretation of the input to a full-scale idiosyncratic meaning shift compared to the input. Despite this, we propose that the prefixes in *vz-po-staviti*-type verbs should be subsumable under a result-modifier analysis.[16]

Based on Žaucer (2013), we thus propose that the structure in Figure 7 captures the two positions for the prefixes in *vz-po-staviti*-type verbs. Note that the result-modifying prefix (on its own) here cannot introduce an unselected object (perhaps unlike the structure in Figure 6).
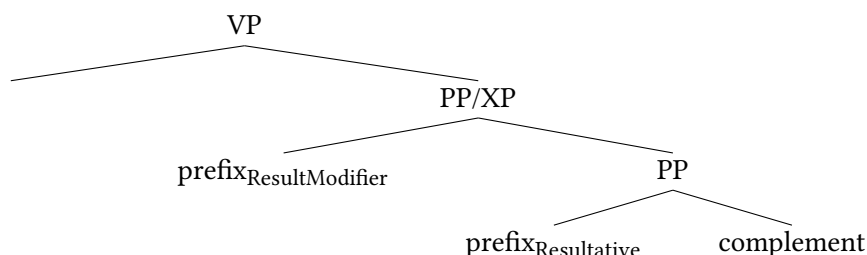
Figure 7: The structure with the two positions of the two prefixes of *vz-po-staviti*-type verbs

5 Conclusions

Our corpus data show that even prefixes which have been claimed to serve (almost) exclusively as lexical prefixes appear stacked over another prefix in up to 20% of their occurrences, which ultimately means that no prefix is used exclusively as a lexical prefix, or that lexical prefixes can sometimes also stack. Our corpus data also confirms a tendency for a hierarchy, but as multiple prefixes have more than one use and since all of them can be used either as lexical or as superlexical prefixes and can appear in more than one position, a true hierarchy of superlexical prefixes could only be determined, perhaps, if prefix occurrences were coded for specific prefix uses – a task that unfortunately seems quite unrealistic, but also one that would inevitably end up drawing in individual researcher's subjective decisions. Our corpus study also showed that whereas prefixed verbs

---

[16]One could say that just like standard lexical prefixes, which sometimes contribute a compositional spatial interpretation and sometimes a non-compositional idiosyncratic interpretation, result-modifying prefixes also have these two options: contributing either a compositional adverbial interpretation or a non-compositional idiosyncratic interpretation, which we observed with many *vz-po-staviti* type verbs.

are very common in Slovenian, verbs with stacked prefixes are very rare, all in all making the use of corpora rather poorly suited for investigating prefix stacking options in Slovenian.

On the other hand, our corpus investigation also turned up a sizeable set of verbs with two prefixes in which the outer prefix does not seem to have any of the typical superlexical characteristics, other than the fact that it occurs stacked over another prefix. Zooming in on these verbs, which we called *vz-po-staviti*-type verbs, we compared their outer prefixes to superlexical prefixes, to intermediate (and other types of in-between) prefixes, and to some types of stacked prefixes that had previously been proposed to instantiate lexical prefixes despite being stacked. We argued that both the inner and the outer prefix in *vz-po-staviti*-type verbs are lexical and cannot be explained away easily. We found that the outer prefixes in these verbs do not seem to form a homogeneous class, and so it is quite likely that it need not be just one explanation that will solve all of these examples. Some of the discussed cases can be explained relatively easily, and at least for a large part of them they seem best treated as (a version of) result-modifying prefixes, though some cases may need alternative approaches, which we leave for future research.

Abbreviations

ACC	accusative		IPFV	imperfective
ATT	attenuative		M	masculine
AUX	auxiliary		NOM	nominative
BCMS	Bosnian/Croatian/ Montenegrin/Serbian		NEG	negation
			PERD	perdurative
COMPL	completive		PFV	perfective
CUML	cumulative		PTCL	particle
DAT	dative		REFL	reflexive
DELIM	delimitative		REP	repetitive
DIST	distributive		SAT	saturative
EXC	excessive		SG	singular
F	feminine		SI	secondary imperfective
GEN	genitive		TERM	terminative
INCP	inceptive		TV	thematic vowel

Acknowledgments

References

Arsenijević, Boban, Katarina Gomboc Čeh, Franc Lanko Marušič, Stefan Milosavljević, Petra Mišmaš, Jelena Simić, Marko Simonović & Rok Žaucer. 2024. *Database of the Western South Slavic Verb HyperVerb 2.0 – WeSoSlaV*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1846.

Babko-Malaya, Olga. 2003. Perfectivity and prefixation in Russian. *Journal of Slavic Linguistics* 11(1). 5–36. http://www.jstor.org/stable/24599703.

Biskup, Petr. 2023. Deriving morphological aspect in Russian: A matryoshka way. In Christina Clasmeier & Julia Golbekand (eds.), *Linguistische Beiträge zur Slavistik: XXIX. JungslavistInnen-Treffen vom 09. bis 10. September 2021 an der Ruhr-Universität Bochum*, 9–30. Berlin: Peter Lang.

Caha, Pavel & Markéta Ziková. 2016. Vocalic alternations in Czech prefixes: Evidence for prefix movement. *Acta Linguistica Hungarica* 63(3). 331–377. DOI: 10.1556/064.2016.63.3.3.

Caha, Pavel & Markéta Ziková. 2022. Prefixes in Czech zero-derived nominalizations and verbs. In Petr Karlík & Lucie Taraldsen Medová (eds.), *Nominalizations and Participles in Czech and beyond*, 3–61. München: Lincom.

Čibej, Jaka, Špela Arhar Holdt, Kaja Dobrovoljc & Simon Krek. 2019. *Frequency lists of words from the Gigafida 2.0 corpus*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1273.

den Dikken, Marcel. 1995. *Particles: On the syntax of verb-particle, triadic, and causative constructions*. Oxford: Oxford University Press.

Di Sciullo, Anna Maria & Elizabeth Klipple. 1994. Modifying affixes. In Sharon Hargus, Gerald R. McMenamin & Vida Samiian (eds.), *Proceedings of the Western conference on linguistics XXIII (WECOL 93)*, 68–80. Fresno: Department of Linguistics, California State University. https : / / wecol . weebly . com / proceedings.html.

Endo, Yoshio & Bartosz Wiland. 2014. The symmetric syntax of Japanese complex verbs and Slavic prefixes. In Anna Cardinaletti, Guglielmo Cinque & Yoshio Endo (eds.), *On peripheries: Exploring clause initial and clause final positions*, 141–159. Tokyo: Hituzi Syobo Publishing.

Filip, Hana. 2003. Prefixes and the delimitation of events. *Journal of Slavic linguistics* 11(1). 55–101. https://www.jstor.org/stable/24599705.

Fowler, George. 1996. An articulated theory of aspect and prefixation in Slavic. In Jindřich Toman (ed.), *Annual workshop on Formal Approaches to Slavic Linguistics: The College Park meeting 1994*, vol. 20, 97–12. Ann Arbor: Michigan Slavic Publications.

Gehrke, Berit. 2008. *Ps in motion: On the semantics and syntax of P elements and motion events* (LOT Dissertation Series 184). Utrecht: LOT.

Istratkova, Vyara. 2006. When prefixes become adverbs. In Tanya Ivanova-Sullivan, Andrea D. Sims & Miriam Whiting (eds.), *Proceedings of the second graduate colloquium on Slavic linguistics*, vol. 6 (Ohio Slavic Papers), 1–31. Columbus: Department of Slavic, East European Languages & Literatures, The Ohio State University.

Klimek-Jankowska, Dorota & Joanna Błaszczak. 2022. The status of secondary imperfectivization in Polish: Evidence from VP idioms. *Journal of Slavic Linguistics* 30(FASL 29 issue). 1–19. DOI: 10.1353/jsl.2022.a923068. http://ojs.ung.si/index.php/JSL/article/view/99.

Łaziński, Marek. 2011. Polish aspectual prefixes, their order and functions: A study based on the National Corpus of Polish. *Word Structure* 4. 231–243. DOI: 10.3366/word.2011.0012.

Markova, Angelina. 2011. On the nature of Bulgarian prefixes: Ordering and modification in multiple prefixation. *Word Structure* 4(2). 244–271. DOI: 10.3366/word.2011.0013.

Matushansky, Ora. 2002. On formal identity of Russian prefixes and prepositions. *MIT Working Papers in Linguistics* 42. 217–253.

Milićević, Nataša. 2004. The lexical and superlexical verbal prefix *iz-* and its role in the stacking of prefixes. *Nordlyd* 32(2). 279–300. DOI: 10.7557/12.70.

Ramchand, Gillian Catriona. 2004. Time and the event: The semantics of Russian prefixes. *Nordlyd* 32(2). 324–360. DOI: 10.7557/12.72.

Ramchand, Gillian Catriona. 2008. *Verb meaning and the lexicon: A first-phase syntax*, vol. 116. Cambridge: Cambridge University Press.

Rappaport Hovav, Malka & Beth Levin. 2001. An event structure account of English resultatives. *Language* 77(4). 766–797. https://www.jstor.org/stable/3086847.

Romanova, Eugenia. 2004. Superlexical versus lexical prefixes. *Nordlyd* 32(2). 255–278. DOI: 10.7557/12.69.

Šekli, Matej. 2016. Pomeni glagolskih predpon v slovenščini. *Filološke studije* 14(1). 273–287. https://hrcak.srce.hr/203716.

Slabakova, Roumyana. 2005. Perfective prefixes: What they are, what flavors they come in, and how they are acquired. In Steven Franks, Frank Y. Gladney & Mila Tasseva-Kurktchieva (eds.), *Formal Approaches to Slavic Linguistics 13: The South Carolina Meeting*, vol. 13, 324–341. Ann Arbor: Michigan Slavic Publications.

Snoj, Marko. 2009. *Slovenski etimološki slovar*. Ljubljana: Modrijan.

Spencer, Andrew & Marina Zaretskaya. 1998. Verb prefixation in Russian as lexical subordination. *Linguistics* 36(1). 1–39. DOI: 10.1515/ling.1998.36.1.1.

Svenonius, Peter. 2004. Slavic prefixes inside and outside VP. *Nordlyd* 32(2). 206–253. DOI: 10.7557/12.68.

Tatevosov, Sergei. 2008. Intermediate prefixes in Russian. In Andrei Antonenko, Christina Bethin & John Baylin (eds.), *Proceedings of the annual workshop on formal approaches to Slavic linguistics 16. the stony brook meeting*, vol. 16, 423–445. Ann Arbor: Michigan Slavic Publications.

Tatevosov, Sergei. 2022. *Multiple prefixation, reflexivization, and event identification*. Paper presented at FASL 31; abstract available at https://fasl.humanities.mcmaster.ca/program/.

Tenny, Carol. 1994. *Aspectual roles and the syntax-semantics interface* (Studies in Linguistics and Philosophy 52). Dordrecht: Springer.

Toporišič, Jože. 2000. *Slovenska slovnica*. Fourth edition. Maribor: Založba Obzorja.

Vidovič Muha, Ada. 1993. Glagolske sestavljenke: Njihova skladenjska podstava in vezljivostne lastnosti. *Slavistična revija* 41. 161–192. https://srl.si/ojs/srl/article/view/COBISS_ID-35424256.

Wiland, Bartosz. 2012. Prefix stacking, syncretism and the syntactic hierarchy. In Markéta Ziková & Mojmír Dočekal (eds.), *Slavic languages in formal grammar: Proceedings of FDSL 8.5, Brno 2010*, 307–324. Berlin: Peter Lang.

Žaucer, Rok. 2002. *The role of verbal prefixes in Slavic: Evidence from Slovenian locative denominal verbs*. Ottawa: University of Ottawa. (MA thesis).

Žaucer, Rok. 2009. *A VP-internal/resultative analysis of 4 "VP-external" uses of Slavic verbal prefixes*. Ottawa: University of Ottawa. (Doctoral dissertation).

Žaucer, Rok. 2013. A case for result-modifying prefixes. In Alexander Podobryaev (ed.), *Formal Approaches to Slavic Linguistics 20: The Second MIT Meeting*, 290–304. Ann Arbor: Michigan Slavic Publications.

Chapter 12

Russian verbal stress retraction as induced unstressability

Ora Matushansky

SFL (CNRS/Université Paris-8/UPL)

This paper analyzes Russian verbal stress through the prism of the 1sg pattern, which characterizes about a third of the productive second conjugation (*i*-verbs), as well as many others. In this pattern the 1sg and a few other present-tense forms surface with inflectional stress, while all other cells of the present-tense paradigm appear with stem-final stress. I propose that this pattern arises as a result of the more general hiatus resolution process that deletes a vowel before another vowel, on the assumption that the accentual specification of the deleted vowel is retained. I propose that the vocalic thematic suffix is post-accenting and the vocalic present-tense suffix is accented. Once the former is deleted, the latter is rendered unstressable because it receives two conflicting accentual requirements: to bear stress (accentuation) and to shift it to the next syllable (post-accentuation). This conflict is resolved by the deletion of the present-tense suffix from the metrical tier, which forces the accent onto the ending if available and onto the final syllable of the stem otherwise.

1 The puzzle: The 1sg present-tense pattern

The Russian verb productively consists of four parts: the lexical stem (henceforth, L-stem), which contains the root and semantically contentful suffixes, the the-

matic suffix, the tense suffix and agreement morphology.[1] The thematic suffix is a cover term for a morpheme that appears between the verbal stem (potentially including verbalizing or imperfective suffixes) and the tense and agreement suffixes. Slavic thematic suffixes have been analyzed as verbalizers or as semantically null morphological glue.[2]

(1) a. léz-   l-   a   (athematic verb)
         L-STEM: climb   THEME: none   TENSE: past   ϕ: F.SG

     a. žértv-ov-   a-   l-   a   (thematic verb)
         L-STEM: sacrifice   THEME: a/i   TENSE: past   ϕ: F.SG

As Russian is a lexical stress language (see Zaliznjak 1985, Melvold 1989, Idsardi 1992, Garde 1998, Alderete 1999, Revithiadou 1999, Butska 2002, and Dubina 2012), each morpheme potentially introduces an accent, which can appear on it (which would make the morpheme ACCENTED), before it (PRE-ACCENTING) or after it (POST-ACCENTING). The position of the surface stress is determined by the Basic Accentuation Principle (2):

(2) THE BASIC ACCENTUATION PRINCIPLE (Kiparsky & Halle 1977):
     Assign stress to the leftmost accented vowel; if there is no accented
     vowel, assign stress to the initial vowel.

An examination of the accentuation of Russian thematic verbs reveals three productive patterns in the present tense correlating with two in the past: consistent stem stress (Table 1-a), consistent post-stem stress (Table 1-b) and variable stress in the present (final stress in the first-person singular, stem-final stress elsewhere, henceforth THE 1SG PATTERN) correlated with post-stem stress in the past (Table 1-c). The pattern in Table 1-d, involving stem-final stress in the present-tense correlating with stress on the thematic suffix in the past, cannot be called productive because it occurs with only four verbal stems, but as it also characterizes the productive verbalizing suffix *-ow-*, it is quite frequent.

The stem-stress pattern in Table 1-a corresponds to an accented L-stem (which, being leftmost, wins over any suffixal accents). The consistent post-stem stress

---

[1]The transcriptions below closely follow Russian orthography and do not indicate: (i) palatalization before front vowels (/Ci/ → [Cʲi], /Ce/ → [Cʲe]), (ii) various vowel reduction phenomena in unstressed syllables, (iii) voicing assimilation and final devoicing. Stress is marked by an acute accent on the vowel. The yers (abstract high lax unrounded vowels) are represented as /ĭ/ (the front yer) and /ŭ/ (the back yer). The letters ч (IPA [t͡ɕ], see Padgett & Żygis 2007), ш (IPA [ʂ]), ж (IPA [ʐ]), щ (IPA [ɕɕ], ц (IPA [t͡s]) are traditionally rendered as č, š, ž, šč, and c.

[2]See Antonyuk et al. (2022–2023) for a range of opinions.

Table 1: Accentual interaction in thematic verbs

		PRS-1SG	PRS-3SG	PST-F.SG	PST-PL
a.	stem: *-žal-* 'sting'	žál^j-u	žál-i-t	žál-i-l-a	žál-i-l-i
b.	post-stem: *-govor-* 'speak'	govor^j-ú	govor-í-t	govor-í-l-a	govor-í-l-i
c.	1SG: *-l^jub-* 'love'	l^jubl^j-ú	l^júb-i-t	l^jub-í-l-a	l^jub-í-l-i
d.	stem-final present: *-koleb-* 'rock'	kolébl^j-u	kolébl^j-e-t	koleb-á-l-a	koleb-á-l-i

in the past tense of both Table 1-b and Table 1-c suggests that the thematic suffix is accented, while the L-stems can be either unaccented or post-accenting. However, the 1SG pattern in Table 1-c is not predicted by the system sketched so far, and neither is the pattern in Table 1-d, which only arises with the class of verbs whose thematic suffix surfaces as /a/ in the past and as /i/ (giving rise to the so-called transitive softening mutation) in the present (henceforth, the *-a-/-i-* class).

In this paper I will link the 1SG pattern to the unstressability of the present-tense suffix, which results from its absence in the metrical tier. I will propose that this absence itself arises from an accentual conflict: that with unaccented L-stems the deletion of the thematic vowel before the present-tense suffix creates an accentual conflict that can only be resolved by the deletion of the problematic position from the metrical tier. I will then hypothesize how post-accenting L-stems can produce both the consistent post-stem stress (Table 1-b) and the stem-final stress in the present (Table 1-d), and link the difference between the two situations to glide deletion and its timing.

The paper is structured as follows. In Section 2 I will introduce the segmental phonology of Russian verbal conjugation and the hiatus resolution mechanism: vowel-before-vowel deletion. I will also discuss the accentuation of the relevant morphemes revealed by their interplay in the athematic verb and show that in the presence of a thematic suffix a stress pattern arises that is not predicted by the interaction of these morphemes.

Section 3 discusses the role of the thematic suffix. I will show that the thematic suffix usually introduces an accent, which should have the double effect of removing the difference between unaccented and post-accenting L-stems and

nullifying the impact of all following suffixes. As this predicts the impossibility of the 1SG pattern and removes the possibility of explaining it in the terms of L-stem accentuation, a special lexical property, that of triggering stress retraction, has been appealed to. I will show that this hypothesis does not explain why some thematic classes are more prone to exhibiting the 1SG pattern than others or why the paradigm cells that fail to undergo retraction are phonologically defined as simple vocalic suffixes. My explanation of the latter fact will be introduced in section Section 4: I will suggest that in the first conjugation the 1SG pattern arises from induced unstressability of the present-tense suffix.

Section 5 examines the 1SG pattern in *-a-/-i-* verbs and argues that verbs manifesting it have an unaccented lexical stem, which further supports a phonological explanation of the 1SG pattern. Section 6 provides such an explanation by ascribing the unstressability of the present-tense suffix to accentual conflict. As will be discussed below, due to the deletion of the thematic suffix before the vocalic present-tense suffix, the same syllable ends up with conflicting instructions: both to bear an accent and to assign it to the next syllable. The need to resolve this conflict will be shown to derive not only the 1SG pattern but also the stem-final stress in the present tense of some *-a-/-i-* verbs. The treatment of the 1SG pattern will be shown to extend to second-conjugation verbs, which have been argued to have a null present-tense suffix.

Section 7 provides the conclusion and discusses potential reasons for the non-productivity of the 1SG pattern in some verb classes.

2 Background: Verbal conjugation and the 1SG pattern

In this section I discuss the conjugation of the Russian verb: first the segmental representation of the two tenses and then their accentual properties. As Table 2 illustrates, Russian has two conjugation classes, distinguished by the vowel appearing before the person-number suffix in the present tense: In the first conjugation it is *-e-* and in the second, *-i-*.

While the first-conjugation *-e-* corresponds to the present-tense suffix, the second-conjugation *-i-* is the thematic suffix (Micklesen 1973, Coats & Lightner 1975, and Itkin 2007: 129–130, though alternative analyses exist, see Section 6.3).

The consideration of the past-tense forms shows that the verb *nestí* 'to carry' is athematic (no vowel appears between the L-stem and the past-tense suffix *-l-*), while the verb *vinít^j* 'to blame' contains the thematic vowel *-i-*, see Table 3.

Table 2: Verbal conjugations, present-tense paradigms: *nestí* 'to carry', *vinít*[j] 'to blame'

	first conjugation		second conjugation	
	singular	plural	singular	plural
1.	nes-e-u → nesú	nes-e-m → nes[j]óm	vin-i-Ø-u → vin[j]ú	vin-i-Ø-m → viním
2.	nes-e-š[j] → nes[j]óš[j]	nes-e-te → nes[j]óte	vin-i-Ø-š[j] → viníš[j]	vin-i-Ø-te → viníte
3.	nes-e-t → nes[j]ót	nes-e-nt → nesút	vin-i-Ø-t → vinít	vin-i-Ø-nt → vin[j]át

The infinitive suffix (surface [ti] under stress, [t[j]] otherwise) shows the same behavior.

Table 3: Verbal conjugations, past-tense paradigms: *nestí* 'to carry', *vinít*[j] 'to blame'

	first conjugation		second conjugation	
	singular	plural	singular	plural
M	nes-l-ŭ → n[j]ós		vin-i-l-ŭ → viníl	
F	nes-l-a → neslá	nes-l-i → neslí	vin-i-l-a → viníla	vin-i-l-i → viníli
N	nes-l-o → nesló		vin-i-l-o → vinílo	

The past tense (historically, the active past participle form) is segmentally un-controversial, and its number-gender suffixes are identical to those of pronouns. While the concatenation of the various morphemes in the past tense is relatively straightforward, in the present tense vowel sequences are created that do not surface as such.[3]

2.1 Verbal conjugation and vowel-before-vowel deletion

While Lightner (1965, 1972) and Halle (1973) propose rather abstract underlying representations for Russian present-tense agreement suffixes, for our purposes the finer details do not matter, and I will follow Melvold (1989) and assume that

---

[3]I will not discuss the details of how the consonant cluster created by the stem-final consonant and the past-tense suffix *-l-* or the infinitive suffix *-tĭ-* is resolved for various consonants (see Lightner 1965, 1972). The alternation between the surface back vowel with a palatalized preceding consonant ([[j]o]) under stress and the phonological /e/ in unstressed syllables in Table 2, Table 3 and elsewhere is allophonic (Lightner 1969, Boyd 1997).

the underlying representations of these suffixes are nearly always identical to their surface forms, as indicated in Table 2.[4] As is easy to see, most but not all agreement suffixes in the present tense are consonantal.

The 1SG form, which will be crucial for the discussion below, shows how hiatus is resolved in Russian. If the vowel preceding another vowel is *i*, like in second-conjugation verbs, it turns into a glide before any vowel distinct from *i*.[5] Otherwise the first vowel is deleted:

(3)   a.   vin-   i-   Ø   u   → vin-j-u → *vinʲú*
           blame TH PRS 1SG

      b.   pros- i-   Ø   u   → pros-j-u → *prošú*
           ask   TH PRS 1SG

      c.   nes- e-   u   → nes-ȩ́-u → *nesú*
           carry PRS 1SG

While in (3c) the deleted vowel belongs to the present-tense suffix, this latter can itself trigger vowel deletion when preceded by a vocalic or vowel-final thematic suffix, such as, for instance, the semelfactive suffix *-nu-*, whose vowel is deleted before the vocalic present-tense suffix, as in Figure 1.

$[[[\sqrt{}\text{-TH}]_1\text{-PRS}]_2\text{-3SG}]_3$ (stomp-PRS-3SG)

$[[[\text{top-nu}]_1\text{-e}]_2\text{-t}]_3$
        ↓
    [tópnet]   Vowel-before-vowel deletion
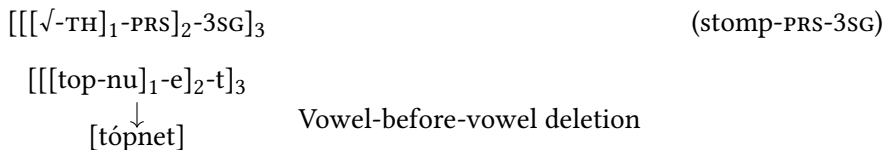
Figure 1: 3SG derivation

---

[4]The surface representations of the 3PL endings, *-ut-* and *-at-* for the first and second conjugations respectively, arise from the morphologically conditioned merger of the present-tense suffix *-e-* (for *-ut-*) or the thematic vowel *-i-* (*-at-*) with the nasal of the ending (cf. Lightner 1969, Kayne 1967). The same VN-modifications occur in the active present participle, inside some verbal roots and in the declension of the ten nouns in [mʲa] (Lightner 1967, Halle 2004). Melvold (1989: 237) assumes this representation for the second conjugation but not for the first one (where she postulates the surface [ut] as the underlying representation), yet the behavior of this ending with respect to stress suggests a consonantal ending in both conjugation classes.
[5]Examples (3b–3c) illustrate the fact that the consonant–glide sequence undergoes a mutation known as TRANSITIVE SOFTENING, or IOTATION (Jakobson 1929, Meillet 1934, Kortlandt 1994, Townsend & Janda 1996, *inter alii*; see Halle 1963, Lightner 1972, Coats & Lightner 1975, Bethin 1992, Brown 1998 and Rubach & Booij 2001 for generativist analyses), which will not be directly relevant here.

The vowel of the thematic suffix is deleted before the vowel of the present-tense suffix also in the 1SG, and then the present-tense suffix is deleted before the vocalic 1SG ending *-u-*, as in Figure 2.[6]

$$[[[\sqrt{\text{-TH}}]_1\text{-PRS}]_2\text{-1SG}]_3 \qquad\qquad (\text{stomp-PRS-1SG})$$

$$[[[\text{top-nu}]_1\text{-e}]_2\text{-u}]_3$$
$$\downarrow$$
$$[[\text{topn-e}]_2\text{-u}]_3 \qquad \text{Vowel-before-vowel deletion}$$
$$\downarrow$$
$$[\text{tópnu}] \qquad \text{Vowel-before-vowel deletion}$$
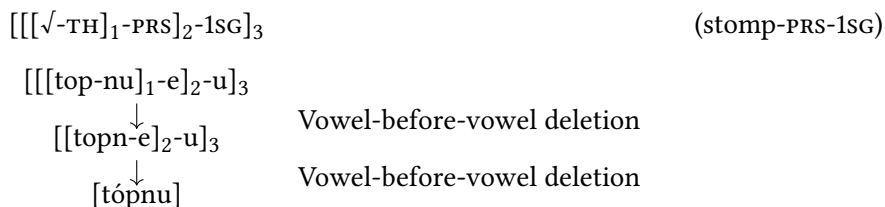
Figure 2: 1SG derivation

The hypothesis that the derivation of the Russian verb involves the deletion of vowels before other vowels was originally proposed by Jakobson (1948), who suggested that the longer form of the verbal stem is always the underlying one (see also Lightner 1965, Halle 1973, Melvold 1989, etc.). The natural question to ask here is what happens to the accents when a vowel is deleted or turns into a glide, and this will turn out to be the clue to the 1SG pattern. However, before this issue can be addressed (in Section 6), it is necessary to establish the underlying accentuation of Russian verbal suffixes.

I will begin with the closed class of verbs lacking the thematic suffix (1a) and on their basis I will show that the 1SG pattern is indeed problematic for the assumptions made so far.

2.2 The Halle–Idsardi stress theory and accent interaction

To illustrate accent interactions I use the autosegmental metrical structure notation introduced by Halle & Vergnaud (1987a,b) and further developed in Melvold

---

[6]Melvold (1989: 83–86) points out that there are two arguments for the absence of the present-tense suffix *-e-* in the 1SG and 3PL: the lack of Velar Palatalization and the position of the stress (which she predicts to retract after hiatus resolution). She proposes therefore that the present-tense suffix is null in the 1SG and 3PL, and the two endings are accented. The correct result ensues, yet the fact that the 1SG and the 3PL endings behave differently in 1SG-pattern verbs then requires an additional stipulation. Since I assume that the present-tense suffix is realized in the entire first-conjugation paradigm, my alternative explanation is that the underlying representation of the present-tense suffix is *-o-* and the source of (Velar) Palatalization is a floating [−back] feature on the 2SG, 3SG, 1PL and 2PL endings. Conversely, palatalization in (3a) results from the consonant-glide sequence *nj* that has undergone transitive softening (see fn. 5). I will not develop the argument further here.

(1989), Idsardi (1992), Halle & Idsardi (1995) and Halle (1997), where each sylla-
ble projected to the metrical tier is indicated by an asterisk and foot edges are
marked by parentheses. Feet are unbounded from one accent to the next or to the
end of the phonological word, and left-headed, which means that lexical accents
can be encoded as underlying left parentheses. The head of each foot is projected
to the next line:

```
           (∗    ∗)
(4)    a.  zim  u
               ↓

            ∗
           (∗    ∗)
       b.  zim  u        →        zímu 'winter.SG.ACC'
```

In the Halle (1997) version, which I will be using here, the fact that unaccented
words surface with initial stress is implemented by the addition of the right paren-
thesis at the right edge. The Basic Accentuation Principle (2) is implemented by
the assumption that feet are left-headed on all lines of the metrical tier, which
ensures that only the head of the leftmost foot projects to the next line:

(5)   Indo-European stress rules (after Halle 1997):

    a. Accents are notated in vocabulary representations with left
        parentheses on line 0

    b. Line 0 is subject to the edge-marking rule RRR

    c. Line 0 is subject to the head-marking rule L

    d. Line 1 is subject to the edge-marking rule LLL

    e. Line 1 is subject to the head-marking rule L

    f. Stress is assigned to the head of the word

Thus when an accented stem is combined with an unaccented suffix, as in (6a),
stress falls on the stem. Conversely, when the stem is unaccented and the suffix,
accented, stress surfaces on the suffix (6b). Finally, when both the stem and the
suffix are unaccented, the first syllable is stressed (6c):

```
            ∗
           (∗      ∗)
(6)    a.  lez  l  i

                   ∗
            ∗    (∗)
       b.  klad l  a                               (surface klalá)
```

```
      *
      *     *)
  c.  klad  l  i                                          (surface kl* áli)
```

Empirically, the combination of a post-accenting stem with an accented suffix does not give rise to a clash: Stress surfaces where both morphemes assign it, i.e., on the suffix. This is illustrated in (7) for the nominal domain: The nominative ending is accented and bears the main stress with both an unaccented and a post-accenting stem (the unaccented accusative ending provides the control distinguishing accented, unaccented and post-accenting stems):

(7) a. ruká/rúku 'hand.SG.NOM/ACC'   (unaccented stem)
     b. čertá/čertú 'line.SG.NOM/ACC'   (post-accenting stem)

In the Halle-Idsardi framework this result is obtained by postulating that whenever a sequence of two parentheses obtains that do not group any stress-bearing material, one of them is deleted:

```
           *(   (*)
(8)  a.  ruk   a
             ↓

                *
           *(   *)
     b.  ruk   a
```

While for examples like (8) the choice of the parenthesis to be deleted makes no difference, the interaction between the left parenthesis introduced by post-accenting morphemes and the right parenthesis introduced by (5b) makes it clear that it is the second parenthesis in a sequence that is deleted, as will be now shown.

Empirically, when a post-accenting morpheme is not followed by any stress-bearing material, stress surfaces on the final syllable. Examples can be readily drawn from nominal declension, where post-accenting nouns surface with stress on the stem-final syllable if the case ending is an unstressable non-vocalized yer, like the genitive plural in (9a) and the nominative singular in (9b). The same happens in adjectives, as in (10):

(9) a. bulavá/bulavámi/buláv 'mace.SG.NOM/PL.INS/PL.GEN'
     b. sekretárʲ/sekretarʲá/sekretarʲámi 'secretary.SG.NOM/SG.GEN/PL.INS'

(10) a. zdoróv/zdorová/zdorovó/zdoroví 'robust.F/M/N/PL'
     b. tʲažʲól/tʲaželá/tʲaželó/tʲaželí 'heavy.F/M/N/PL'

Several ways of accounting for this effect are possible and I will not choose between them.[7] Importantly, under all approaches this process, distinguishing as it does between vocalized and non-vocalized yers, is a late one. What is crucial, however, is that the representation of such cases in the Halle-Idsardi framework involves two parentheses on the right edge:

(11) a. nominative singular

       *   * (*)
     bu lav a

    b. genitive plural

       *   * ( )
     bu lav ŭ

If the first parenthesis in the sequence were deleted, the outcome would be identical to that for an unaccented stem and stress would be incorrectly predicted to be initial. If, on the other hand, the rightmost parenthesis is deleted, the resulting configuration can be repaired as suggested in fn. 7.[8] It will be later demonstrated that the deletion of the second one in the sequence of two immediately adjacent parentheses leads to a correct prediction in another situation where such a configuration arises.

2.3 The underlying accentuation of Russian verbal suffixes

Following Halle (1973) and Melvold (1989), four main accentual classes of athematic verbs can be established, depending on the accentuation of the root, with the positions of the underlying accents indicated by underlining in Table 4.[9] As discussed above, systematic stem stress (Table 4-a) is a sign of an accented root, and variable stress (Table 4-c) is an indicator of an unaccented root. I follow Melvold (1989) and treat (b) and (d) in Table 4 as post-accenting roots, but differ from her in their analysis, as will be seen below.

---

[7]The assumption that the nominative singular and genitive plurals endings are underlyingly back yers makes it possible to capitalize on the fact that word-internally an accent assigned to a yer surfaces on the preceding syllable. To capture this, Halle (1997: 284) inserts a left parenthesis on the syllable preceding an accented yer. Alternatively, these stress retraction phenomena have been accounted for by an appeal to iambic feet in Russian (Crosswhite 1999, 2000, Gouskova 2010, and Dubina 2012, among others). I will not attempt to address this discussion here.

[8]Yet another alternative would be to move the final left parenthesis before the insertion of the right parenthesis. I reject this option since it requires the same repair strategy with an additional assumption about ordering, and the need to delete one of the two immediately adjoining parentheses is motivated independently.

[9]The fifth class consists of just two verbal roots, *-mog-* (*močʲ* 'to be able') and the cranberry root *-im-/-nʲa-* (e.g., *prinʲátʲ* 'to accept') and their derivatives, which exhibit the 1sg pattern. I return to this matter in Section 7.1.1.

Table 4: Accentual interaction in athematic (√-T-ɸ) verbs

			accented PST-F.SG	unaccented PST-PL	accented PST-3SG
a.	stem: -*lez*- 'climb'	A	léz-l-a	léz-l-i	léz-e-t
b.	post-stem: -*nes*- 'carry'	PA+	nes‿-l-á	nes‿-l-í	nes‿-ʲó-t
c.	variable (past): -*klad*- 'put'	UA	kla-l-á	klá-l-i	klad-ʲó-t
d.	retracting (past): -*griz*- 'gnaw'	PA−	gríz‿-l-a	gríz‿-l-i	griz‿-ʲó-t

I also follow Melvold in assuming that the past-tense exponent -*l*-, as expected from a consonantal affix, does not introduce an accent. As a result, the contrast between the feminine and the plural in the past of Table 4-c is derived by treating the plural suffix -*i*- as unaccented, while the feminine ending -*a*- is accented. The masculine and neuter endings are unaccented as well.

The post-stem pattern in Table 4-b, with consistent final stress in the past, results, Melvold argues, from a post-accenting root, whereas the pattern in Table 4-d involves the special rule of retraction triggered by a subclass of verbal roots. While Melvold (1989) implements this process by moving the relevant parenthesis one syllable to the left, Halle (1997) handles it by inserting a parenthesis before the preceding syllable.

Lexically conditioned retraction does not, however, explain the facts discussed in Matushansky (to appear), namely, that verbs following the pattern in Table 4-b also violate the Basic Accentuation Principle (2) in the infinitive and in the passive past participle. Despite the fact that both these suffixes behave as pre-accenting in other environments, stress is final:

(12) a. nestí 'to carry',
      cf. léztʲ 'to climb', klástʲ 'to put', gríztʲ 'to gnaw'
   b. unesená 'carried away.F.SG',
      cf. perelézena 'climbed over', progrízena 'gnawed through', sprʲádena 'spun'

The Basic Accentuation Principle (2) predicts that in a sequence of a post-accenting and a pre-accenting morpheme the stress assigned by the latter should win

(13) (indices are added to indicate which morpheme introduced which parenthesis). Such is in fact the case in other instances of such morpheme sequences.

(13)
a.
$$* \ (^1 +(^2 \ *$$
nes   tĭ

b.
$$(^2 \ * \qquad (^1 \ *$$
nes   tĭ   wrongly winning accent

To explain the facts in (12), Matushansky (to appear) argues that the pattern in Table 4-d should be analyzed as involving unaccentable roots, i.e., roots that cannot bear a parenthesis anywhere but at the right edge.[10] As a result, the accent is forced rightwards, yielding word-final stress in passive past participles and the realization of the yer in the infinitive suffix. To explain the pattern in Table 4-d Matushansky (to appear) proposes that it involves post-accenting stems and that forcing stress retraction is the general property of the past-tense suffix.[11] As is easy to see, under this approach accented stems will retain stress on themselves, unaccented stems will be unaffected, unaccentable stems will still force post-stem stress, and only in Table 4-d stress will be retracted:

(14)
a.
$$*(\quad (*)$$
grɨz l a
$$\downarrow$$

b.
$$(* \quad (*)$$
grɨz l a
$$\downarrow$$

c.
$$*$$
$$(* \quad (*)$$
grɨz l a

---

[10]This is a novel notion introduced to explain the fact that both the unaccentable PPP suffix *-en-* and unaccentable roots cannot bear an accent but, as a last resort, can bear stress when not followed by stress-bearing material. This ability to bear stress distinguishes unaccentability from unstressability (to be discussed further). See Matushansky (2023b) for a proposal distinguishing the two in a different framework treating Russian accent as tone (cf. Dubina 2012): Unaccentable roots in it are absent from the tonal tier, and unstressable ones, from the metrical tier.

[11]The accent introduced by the feminine ending *-a* is not affected by this retraction. This is naturally achieved if stress is assigned cyclically, but I will not pursue this line of inquiry here, leaving it for future research.

Turning now to the present tense, only two patterns can be detected (modulo fn. 9): systematic stress on the stem (Table 4-a) if it is accented, and on the present-tense suffix (Table 4-b–d) otherwise. This means (Halle 1973, Melvold 1989) that the present-tense suffix has to introduce an accent: If it were unaccented, the Basic Accentuation Principle (2) would predict stem stress both for accented stems (due to the accent of the stem) and for unaccented stems (stress on the leftmost syllable). Conversely, if the present-tense suffix is accented (as assumed by Melvold 1989), post-stem stress is correctly predicted for the entire present-tense paradigm for both unaccented (15) and post-accenting (16) roots:

(15)   a.   klad- e̲-   m   → kladʲóm
              put   PRS 1PL

       b.   klad- e̲-   te   → kladʲóte
              put   PRS 2PL

(16)   a.   nes‾ - e̲-   m   → nesʲóm
              put   PRS 1PL

       b.   nes‾ - e̲-   te   → nesʲóte
              put   PRS 2PL

Given that the present-tense suffix is accented and deleted before the 1SG ending -*u*-, the fact that this ending remains stressed with unaccented verbs (e.g., *kladú* 'put.1SG') demonstrates, *ceteris paribus*, that the accent of a deleted vowel is neither deleted nor shifted to the left, and this is also what is predicted by the Halle-Idsardi system:

```
              *  ( *    *)
(17)   a.   klad- e-   u
              put    PRS 1SG
                    ↓
              *  (      *)
       b.   klad- e̸-   u
              put    PRS 1SG
                    ↓    *
              *  (      *)
       c.   klad-      u
              put    PRS 1SG
```

A possible alternative would be that the accent is deleted together with the vowel but the 1SG ending is accented, drawing the stress. I will argue, however, that the interaction of accents surviving after hiatus resolution can account for the 1SG pattern that would be inexplicable otherwise.

2.4 Intermediate summary

In this section I have discussed and motivated my background assumptions about the segmental and accentual properties of Russian tense and agreement morphemes. Segmentally, Russian tense and agreement markers were taken to coincide with their surface forms except for the present-tense suffix -*e*-, which surfaces as -*^jo*- (palatalizing [o]) under stress, and the 3PL suffix (which I take to be -*nt*-). Sequences of two vowels are resolved, following Jakobson (1948), by the deletion of the first one (unless the first vowel is an *i*, which turns into a glide before a vowel other than *i*).

The examination of the finite paradigms of athematic verbs, alongside with their infinitive and passive past participle forms, makes it possible to determine the accentual properties of various inflectional suffixes:

- the present-tense suffix -*e*- and the feminine singular suffix -*a*- are accented

- the plural suffix -*i*- is unaccented (and the same is true for the masculine (-*ŭ*-) and neuter (-*o*-) suffixes, which show the same accentual behavior; for minor lexically-conditioned variation see Melvold 1989 and Marklund Sharapova 2000)

- the past-tense suffix -*l*- is unaccented but retracting (forcing the realization of the stress of a post-accenting stem on the stem-final syllable)

Since in the Halle-Idsardi system feet are left-headed, the deletion of an accented vowel yields rightward stress shift. This prediction is correct for the class of verbs in Table 1-b, characterized by the post-stem stress pattern. I will now argue that the accentual patterns in Table 1-c and d cannot be explained by the mechanisms postulated so far.

3 Thematic verbs and the 1SG pattern

Except for the two athematic stems in fn. 9, the 1SG stress pattern in the present is only attested in thematic verbs. Importantly, it can be found with several thematic suffixes, as shown in Table 5 and Table 6.

Table 5: Accentual interaction in thematic verbs, illustrated for the semelfactive suffix *-nu-*

		accented PRS-3SG	accented PRS-1SG	accented PST-F.SG	unaccented PST-PL
a.	stem: *-top-* 'stomp'	tóp-n-e-t	tóp-n-u	tóp-n-u-l-a	tóp-n-u-l-i
b.	post-stem: *-max-* 'wave'	max-n-nʲ-ó-t	max-n-ú	max-n-ú-l-a	max-n-ú-l-i
c.	1SG: *-obman-* 'cheat'	obmá-n-e-t	obma-n-ú	obma-n-ú-l-a	obma-n-ú-l-i

Table 6: Accentual interaction in thematic verbs, illustrated for the thematic suffix *-e-*

		accented PRS-3SG	accented PRS-1SG	accented PST-F.SG	unaccented PST-PL
a.	stem: *-vid-* 'see'	víd-i-t	víž-u	víd-e-l-a	víd-e-l-i
b.	post-stem: *-vel-* 'order'	vel-í-t	velʲ-ú	vel-é-l-a	vel-é-l-i
c.	1SG: *-vert-* 'spin'	vért-i-t	verč-ú	vert-é-l-a	vert-é-l-i

The fact that first-conjugation verbs (*-ʲo-* in the present tense, exemplified by Table 5) and second-conjugation verbs (/i/ in the present tense, exemplified by Table 6) can both exhibit the 1SG pattern suggests that it is linked not to a given concrete present-tense suffix, but to the morphological feature [−PST]. While I will not make such an assumption, the discussion of the present-tense allomorphs will be postponed until Section 6, and in the remainder of this section I will address the thematic suffix, arguing that it plays a crucial role in the emergence of the 1SG pattern.

3.1 The accentuation of thematic suffixes and the 1SG pattern

Since the vowel of the thematic suffix either is deleted or turns into a glide before the vowel of the present-tense suffix, the underlying accentuation of the thematic suffix must be established on the basis of the past tense, where it is left intact. The Basic Accentuation Principle (2) means that accented L-stems can be identified by systematic stem stress in Table 5-a and Table 6-a. If the thematic suffix were unaccented, we would expect to find the varying pattern in the past tense of some verbs, indicating unaccented L-stems, as in Table 4.[12] The fact that this pattern is unattested in the past tense of verbs exhibiting the 1SG pattern strongly suggests that the thematic suffix must introduce an accent.

Because a sequence of two parentheses without any stress-bearing elements between them is simplified to a single parenthesis (see Section 2.2), post-accenting stems in (18b) are simplified to the same representation as unaccented stems (18a) by the time a tense suffix is added, so the difference between the 1SG pattern and the post-stem pattern in the present is not expected to follow from the accentuation of the L-stem.

(18) a. unaccented L-stem

> * (*   (*)
> √- TH PST F.SG

 b. post-accenting L-stem

> *((*   (*)   * (*   (*)
> √- TH PST F.SG   →   √- TH PST F.SG

It therefore seems reasonable to assume (Halle 1973: 328, Melvold 1989: 291, Idsardi 1992: 124, Gladney 1995: 114–117, Feldstein 2015, among others) that the 1SG present-tense pattern is due to something not considered so far.

3.2 The role of the thematic suffix

While Red′kin (1965) and Zaliznjak (1985) claim that there is no correlation between the thematic suffix and stress, Slioussar (2012) shows that the three stress patterns in Tables 1, 5, and 6 are not equally productive in all verb classes and that some thematic suffixes do not produce the 1SG pattern at all, as shown in Table 7. (Examples are provided for all thematic classes that can give rise to the

---

[12]One thematic suffix, surfacing as *-a-* in the past tense and undetectable in the present, is unaccented. Evidence for this comes from the variable position of the stress in its past tense (e.g., *lgalá/lgáli* 'lied.F.SG/PL'). In the present the suffix is undetectable (*lgu/lžʲot* 'lie.1SG/3SG') due to hiatus resolution before the vocalic present-tense suffix.

1SG pattern; for non-productive thematic classes the numbers given represent the number of unprefixed verbs in that class.[13])

Table 7: Stress and thematic suffixes

	theme (PST/PRS)	PRS.1SG	PRS.2SG	INF	gloss	1SG pattern
a.	**a/aj**	čit-áj-u	čit-áj-e-š[j]	čit-á-t[j]	'read'	0/∞
b.	**e/ej**	bel-éj-u	bel-éj-e-š[j]	bel-é-t[j]	'be white'	0/∞
c.	**nu/n (semelfactive)**	tolk-n-ú	tolk-n[j]-ó-š[j]	tolk-nú-t[j]	'push'	6/∞
d.	none or Ø	mog-ú	móž-e-š[j]	móč[j]	'be able'	2/84
e.	a/Ø	ser-ú	sér-e-š[j]	sr-á-t[j]	'shit (dial.)'	1–2/20 (39)
f.	a/i	piš-ú	píš-e-š[j]	pis-á-t[j]	'write'	60/103 (84)
g.	o/i	kol[j]-ú	kól-e-š[j]	kol-ó-t[j]	'stab'	5/5
h.	nu/n (mutative)	gíb-n-u	gíb-n-e-š[j]	gíb-nu-t[j]	'perish'	0/60
i.	**i**	proš-ú	prós-i-š[j]	pros-í-t[j]	'ask'	23%
j.	e/Ø	verč-ú	vért-i-š[j]	vert-é-t[j]	'turn'	6/83

Several empirical generalizations can be established based on the patterns in Table 7. Of the four productive verb classes in Russian (a, b, c, i; the unproductive

---

[13]For the 21 *j*-final verbs with the theme -*a*- in the past, the shape of the stem makes it impossible to determine if in the present this theme is deleted before the present-tense suffix (cf. fn. 12) or undergoes a readjustment rule (cf. Matushansky 2023a) turning it into [i] (which would then turn into a glide). The same issue arises for the two verbs with OCS palatalization of the final consonant cluster ([žd] arising from underlying [dj]), *žáždat[j]* 'to thirst' and the non-standard *stráždat[j]* 'to suffer'.

(i)   a.   tájat[j]/táju/táet 'melt.INF/1SG/3SG'

          b.   stráždat[j]/strážduu/stráždet 'suffer.INF/1SG/3SG' (Modern Russian *stradát[j]/stradáju/stradáet*, literary variant with the -*a*-/-*i*- thematic suffix *stradát[j]/strážduu/stráždet*)

Nineteen of them have stem stress and, though I have assigned them to the -*a*-/-*i*- class, their uncertain status is indicated by parentheses in the table. The two *j*-final verbs with systematic post-stem stress, *smeját[j]s[j]a* 'to laugh' and the archaic *vopiját[j]* 'to clamor', have been assigned to the -*a*-/-*Ø*- class because no verb with a detectable -*a*-/-*i*- thematic suffix shows the post-stem stress pattern in the present tense.

class f also contains all the verbs derived with the productive suffix *-ow-*) the 1SG pattern is productive in one (1556 out of the 6875 *i*-verbs in Zaliznjak's 1977 dictionary, according to the calculations in Slioussar 2012). It never occurs with the thematic suffixes surfacing as *-aj-* and *-ej-* in the present, which suggests that it is dependent on the deletion of a vowel. However, verbs derived with the pre-accenting mutative suffix *-nu-* (Table 7-h) or with the unaccented thematic suffix *-a-* that is deleted in the present tense (Table 7-e) also do not give rise to the 1SG pattern.

Given these facts it is reasonable to assume that the 1SG pattern is linked to the deletion of a vowel that introduces an accent. Support for this hypothesis comes from the fact that with the accent-bearing vocalic suffixes *-i-* and *-a-/-i-* (as well as its allomorph *-o-/-i-*) the 1SG pattern is systematic, and with two more accent-bearing thematic suffixes it is marginally possible: with the semelfactive first-conjugation *-nu-* (six verbs to be discussed in section Section 7.1.1) and with the second-conjugation *-e-* (five verbal roots, see Section 7.1.3).

The most important empirical generalization to be drawn from Table 7 is that the 1SG pattern is systematically available with some thematic suffixes (*-i-*, *-a-/-i-*) and exceptional with others (*-nu-*, *-e-*), which suggests that the properties of the thematic suffix play a role in determining which stress pattern the verbal stem (L-stem + thematic suffix) gives rise to.[14] Nonetheless, as will be presently shown, it cannot be the thematic suffix itself that is responsible, since in no verb class is the 1SG stress pattern the only one available.

3.3 Treating retraction as the lexical property of the stem

As discussed in Section 2.2, when two parentheses appear in a sequence with no asterisk in between, one is deleted, reflecting the fact that a post-accenting stem and an unaccented stem followed by an accented suffix yield the same surface outcome (18). This is also the configuration that arises when the vowel of the thematic suffix is followed by the present-tense suffix, as in (19). When the thematic vowel is deleted, a sequence of two accents is created that should be resolved into one:[15]

---

[14]Zaliznjak (1985: 28, 380) offers a number of lexical generalizations over both patterns and points out that the systematic post-stem pattern is characteristic of the more archaic strata of the vocabulary, providing such near-minimal prefixed verb pairs as the standard *razbužú/razbúdit* 'awaken.1SG/3SG', *počinʲú/počínit* 'repair.1SG/3SG' (1SG pattern) vs. the literary *učinʲú/učinít* 'initiate.1SG/3SG', *vozbužú/vozbúdit* 'arouse.1SG/3SG' (post-stem stress). The prefixes themselves, however, cannot be regarded as the reason for these contrasts.

[15]I use the semelfactive suffix *-nu-* for an example despite the exceptionality of the 1SG pattern with it because the consonant remaining after hiatus resolution makes it easier to abstract away from the L-stem. While the suffix is represented here as accented (since it also makes for easier representations), I will revise this assumption later.

```
      *  (*(*)
(19)  a.  √- nu  e    t
          TH  PRS 3SG
                ↓
      *  (  (*)
      b.  √- nʹ́  e    t
          TH  PRS 3SG
                ↓
      *     (*)
      c.  √- n   e    t
          TH  PRS 3SG
```

It is easy to see that systematic post-stem stress is predicted here, and the addition of a vocalic suffix (such as the 1SG -*u*-) instead of a consonantal one (like the 3SG -*t*-) does not change the outcome. This is why Melvold (1989: 291), following Halle (1973), proposes that the stems giving rise to the 1SG pattern are marked to undergo retraction in all forms of the present tense except 1SG, where the present-tense suffix is null (see fn. 6). Idsardi (1992: 124) improves upon this by proposing that retraction fails in the 1SG because its trigger, the present-tense marker, is deleted before another vowel. Halle (1997) encodes retraction by inserting an additional parenthesis before the trigger morpheme:

```
      *  (*(*
(20)  a.  √- nu  e    t
          TH  PRS 3SG
                ↓
      *     (*
      b.  √- nʹ́  e    t
          TH  PRS 3SG
      *     ↓
      (*   (*
      c.  √- n   e    t
          TH  PRS 3SG
```

Even though the present-tense suffix is deleted before the 1SG ending only in first-conjugation verbs (in second-conjugation verbs there is glide formation in the 1SG (fn. 5)), the connection between the stress failing to retract and a vocalic ending is real and supported by independent evidence. As Feldstein (2015) points out, there exist two more forms with the same final stress as in the 1SG: the imperative (surface [i]) and the present tense gerund (surface [ʲa]):[16]

---

[16]Stress in the active present participle generally patterns with non-1SG, but sometimes doesn't (e.g., *učúsʲ*/*účitsʲa* 'study.1SG/3SG' vs. *učáščijsʲa* 'studying.M.SG', see also Zaliznjak 1985: 29, 77).

(21) a. vert-í 'spin.IMP', vertʲ-á 'spin.GER' (cf. verčú/vértit 'spin.1SG/3SG')

   b. obman-í 'cheat.IMP' (cf. obmanú/obmánet 'cheat.1SG/3SG')

   c. lʲub-í 'love.IMP', lʲubʲ-á 'love.GER' (cf. lʲublʲú/lʲúbit 'love.1SG/3SG')

While Feldstein simply points out that non-retracting forms all have a simple vowel ending of the type -V#, Idsardi's proposal makes retraction failure phonologically predictable, deriving it from hiatus resolution. For this proposal to succeed, however, it is necessary for retraction to happen after hiatus resolution:

(22)
```
        *  (* (*   *)
   a.  √- nu   e    u
          TH   PRS 3SG
                ↓
        *  (    (*   *)
   b.  √- nu̸    e    u
          TH   PRS 3SG
                ↓
        *  (    (    *)
   c.  √- nu̸   e̸    u
          TH   PRS 3SG
```

Assuming that the deletion of a vowel removes it from the metrical tier but retains the accent, it is to the representation in (22c) that stress rules apply. While the first parenthesis is deleted by regular processes (since no metrical element follows), Idsardi's claim is that the deletion of the present-tense suffix makes it impossible for it to trigger retraction.

   The assumption that the past-tense suffix -*l*- is retracting (Section 2.3) makes it impossible to explain retraction failure in the 1SG by the fact that a deleted suffix is removed from the metrical tier: The asyllabic past-tense suffix is not present on the metrical tier either. Furthermore, the restrictions both on the verbal classes exhibiting the pattern (only with deleted accent-bearing thematic suffixes) and on the pattern itself (failing before simple vocalic endings) suggest that it is not due to an arbitrary lexical property of the stem. In the next section I will introduce an explanation for retraction failure with simple vocalic endings: I will propose that the 1SG pattern results from induced unstressability.

4 The 1sg pattern as induced unstressability of the present-tense suffix

I begin this section with an assumption. Suppose that with some verbs the present-tense suffix is not represented on the metrical tier. Once again I use a *-nu-* verb to illustrate the matter and I will assume an unaccented L-stem because, as discussed in Section 3.1, the combination of a post-accenting stem and an accented suffix produces the same result as that of an unaccented stem and an accented suffix.

Starting out with the 1sg form, the assumption that the present-tense suffix is absent from the metrical tier (23a) gives rise to word-final stress once the vowel of the thematic suffix is deleted before the vowel of the present-tense suffix (23b). The deletion of the present-tense suffix (23c) yields the correct surface form:

(23)　a.
```
      *  (*        *)
   √- nu   e    u
      TH   PRS  3SG
            ↓
```
　　b.
```
      *  (         *)
   √- nú   e    u
      TH   PRS  3SG
            ↓
```
　　c.
```
      *  (         *)
   √- nú   e̸    u
      TH   PRS  3SG
```

Asyllabic endings are predicted to exhibit different behavior. After the deletion of the thematic vowel (24b) the two parentheses at the right edge of the word are not followed by any metrical material. As discussed in Section 2.2, this configuration yields leftward stress shift, which I implement, like in (20), by doubling the last left parenthesis. Stem-final stress (24c) is therefore correctly predicted with asyllabic endings, and Feldstein's generalization (stress retraction in the absence of a vocalic ending) is explained:

(24)　a.
```
      *  (*        )
   √- nu   e    t
      TH   PRS  3SG
            ↓
```

```
        *  (            )
  b.  √- nʉ́   e    t
        TH   PRS 3SG
                ↓

      (*  (
  c.  √- nʉ́   ɇ́   t
        TH   PRS 3SG
```

The final issue to be resolved is that of the 2PL ending *-te-*, which is wrongly predicted to be stressed:

```
          *  (          *)
(25)  a.  √- nʉ́   e    te
            TH   PRS 2PL
                  ↓        *
        ✗   * (          *)
      b.  √- nʉ́   e    te
            TH   PRS 3SG
```

I hypothesize, in accordance with historical evidence (Zaliznjak 1985: 316–322), that the 2PL ending is either extrametrical (i.e., not represented on the metrical tier) or retracting. As a result, stress ends up on the stem-final syllable in all cells of the finite paradigms except for the 1SG.

The hypothesis that the 1SG pattern is due to induced unstressability of the present-tense suffix rather than to retraction explains why stress is stem-final with consonantal suffixes and final with vocalic ones but does not explain how unstressability is induced. Yet if the 1SG pattern is lexically triggered, it is not expected to be productive, contrary to fact (Section 3.2); moreover, *i*-verbs with post-stem stress are being continually shifted into it (see Zaliznjak 1985: 29, 2019, Feldstein 1986: 57–59, Choi 1996: 108, Marklund Sharapova 2000: 132, and Esʹkova 2008: 469, 2014: 469). It is therefore desirable to derive induced unstressability from some independently motivated property of the L-stem, and stem accentuation (i.e., the lack of an accent vs. post-accentuation) seems the best candidate. In the next section I will provide some evidence linking the 1SG pattern and the lack of an accent on the L-stem.

5 The accentuation of the verbal stem

As shown by Halle (1973, 1975, 1997) and Melvold (1989), stem-conditioned stress retraction is also attested in the nominal declension, where some nouns undergo

it in the plural, and in adjectival inflection, where it is triggered for most adjectival stems by the long-form suffix. Melvold (1989) further argues that both post-accenting and unaccented stems can undergo retraction:

(26) unaccented feminine stem: final stress in the singular, except in the accusative

    a. sg: ruká/rúku 'hand.sg.nom/acc', pl: rukámi 'hand.pl.ins'   (regular)

    b. sg: dušá/dúšu 'soul.sg.nom/acc', pl: dúšami 'soul.pl.ins' (retracting)

(27) post-accenting feminine stem: final stress in the singular

    a. sg: čertá/čertú 'line.sg.nom/acc', pl: čertámi 'line.pl.ins'   (regular)

    b. sg: stroká/strokú 'text line.sg.nom/acc', pl: strókami 'text line.pl.ins'

             (retracting)

Verbs are different. While Melvold (1989: 291) proposes that verbal retraction occurs in post-accenting L-stems, in this section I will argue that the 1sg pattern correlates with unaccented L-stems.

5.1 The verbs in *-a-/-i-* and L-stem accentuation

The class of first-conjugation *-a-/-i-* verbs is a semi-closed one: This thematic suffix combines with a finite set of stems (103, to the best of my knowledge, as well as five *-o-/-i-* verbs)[17] and is also used with the denominal verbalizing suffix *-ow-* (on which more below). Instead of the post-stem stress pattern (Table 8-b), this class contains, in addition to verbs exhibiting the stem and 1sg patterns, five verbal roots (cf. Gladney 1995: 115) with the pattern in Table 8-d, where stress is retracted to the stem-final syllable throughout the present-tense paradigm: *kolebátʲ/koléblʲu* 'rock.inf/1sg', *kolixátʲ/kolíšu* 'sway.inf/1sg', *alkátʲ/álču* 'crave.inf/1sg', the archaic variant *stradátʲ/stráždu* 'suffer.inf/1sg' and the two equally archaic prefixed derivatives of the cranberry root *-im-*, *vnimátʲ/vnémlʲu* 'heed.inf/1sg' and *primátʲ/priémlʲu* 'accept.inf/1sg'; in modern spoken Russian the last three take the thematic suffix *-aj-*. The post-stem pattern, on the other hand, is not attested in this thematic class.

    Inside this class there are two subclasses of derived stems: the non-productive class derived by the suffix *-ot-* and the productive class in *-ow-*. While the former

---

[17]The five verbs in *-o-/-i-* (*kolótʲ* 'to stab', *molótʲ* 'to grind', *polótʲ* 'to weed', *borótʲ* 'to fight', and *porótʲ* 'to whip') all have stems ending in [olo] or [oro], which are, respectively, pleophonic allomorphs of *-la-* and *-ra-* in Russian (on pleophony in Slavic see, e.g., Sussex & Cumberley 2006: 36–37, 207).

Table 8: Accentual interaction with the 1st conjugation suffix *-a-/-i-*

		accented PRS-3SG	accented PRS-1SG	accented PST-F.SG	unaccented PST-PL
a.	stem: *-maz-* 'smear'	máž-e-t	máž-u	máz-a-l-a	máz-a-l-i
b.	post-stem: N/A; potential candidates among *j*-final stems can be assigned to the *-a-/-Ø-* class				
c.	1SG: *-vʲaz-* 'tie'	vʲáž-e-t	vʲaž-ú	vʲaz-á-l-a	vʲaz-á-l-i
d.	stem-final present: *-koleb-* 'rock'	koléblʲ-e-t	koléblʲ-u	koleb-á-l-a	koleb-á-l-i

creates the 1SG pattern (Table 8-c), the latter gives rise to the stem-final one (Table 8-d).

Starting with the former, all ca. 20 stems ending in *-ot-* form sound-emission verbs. While for most verbs in this category no meaningful root can be identified before *-ot-*, at least the verbs *vorkotátʲ* 'to grumble', *topotátʲ* 'to stamp' and *trepetátʲ* 'to tremble' can be argued to be built on the roots *-vork-*, *-top-* and *-trep-*, given the verbs *vorkovátʲ* 'to coo', *tópatʲ* 'to stamp, tramp' and *trepátʲ* 'to pull, flutter'. The fact that the accented root of (28a) is not stressed in (28b) could indicate that *-ot-* is accentually dominant (and either post-accenting or unaccented, since it is never stressed itself), and this is confirmed by the fact that all verbs with this suffix give rise to the 1SG pattern.

(28)   a.   tópatʲ 'to stamp, tramp': tópaju (1SG)/tópajet (3SG)
         b.   topotátʲ 'to stamp': topočú (1SG)/topóčet (3SG)

Can it be determined if *-ot-* stems are unaccented or post-accenting? Unfortunately, the answer is no, because the thematic suffix *-a-/-i-* introduces an accent, and the fact that action nouns null-derived from *-ot-* verbs exhibit initial stress (29), while suggestive of an unaccented L-stem, could also be the artefact of conversion, which favors initial stress.[18] However, the fact that the same suffix uniformly gives rise to the same accentual behavior strongly indicates that the 1SG pattern depends on the accentuation of the L-stem.

---

[18]None of these nouns naturally forms a plural, which excludes this way of checking their accentuation. The fact that the post-accenting diminutive suffix *-ŭk-* derives a post-accenting noun (*xoxotók*) is also non-indicative (cf. Halle 1973: 340).

(29) a. groxotát^j 'to bang': groxočú (1sg)/groxóčet (3sg)
      gróxot 'a bang'
   b. xoxotát^j 'to laugh loudly': xoxočú (1sg)/xoxóčet (3sg)
      xóxot 'laughter'

The same conclusion can be drawn from the denominal verbalizer *-ow-*, which, as (30) shows, surfaces as [ov] before the surface [a] in the past and in the infinitive and as [u] (followed by the surface [j]) in the present.[19]

(30) a. vračevát^j 'to treat, heal': vračúju (1sg)/vračújet (3sg)   (retraction)
   b. kritikovát^j 'to critique': kritikúju (1sg)/kritikújet (3sg)

The accentual behavior of *-ow-* verbs is strikingly different from that of *-ot-* verbs (as well as from that of *i*-verbs): Unless they have systematic stem stress (e.g., *komándovat^j* 'to command'), they are stressed on the thematic suffix in the past and exhibit stem-final stress in the present (Table 8-d). The dependence of this stem-final pattern on the accentuation of the thematic suffix is confirmed by the fact that the *-a-/-i-* subclass contains no verbs that exhibit post-stem stress in the present (modulo fn. 13).

Given that the *-a-/-i-* thematic suffix introduces an accent, the accentuation of the *-ow-* stems in (30) is difficult to determine: They can be unaccented or post-accenting. Since *-ow-* verbs are denominal, the accentuation of their L-stems should be linked to their nominal bases. However, as noted by Red′kin (1965), Halle (1973: 344–347), Zaliznjak (1985: 107), and Gladney (1995), among others, the relation between the accentuation of a noun and that of the verb that is derived from it is not straightforward, as can also be shown by the following *i*-verbs:

(31) accented nouns
   a. razžálobl^{j}ú/razžálobit 'move to pity.1sg/3sg'   (stem)
      (cf. žáloba/žálobu 'complaint.nom/acc ')
   b. bešú/bésit 'enrage.1sg/3sg'   (1sg)
      (cf. bésa/bésami 'devil.sg.gen/pl.ins'
   c. bombl^{j}ú/bombít 'bomb.1sg/3sg'   (inflection)
      (cf. bómba/bómbu 'bomb.nom/acc ')

---

[19]See Melvold (1989) for the assumption that *-ow-* combines with the thematic suffix *-a-/-i-* and a demonstration how its surface realization is determined by the resulting syllable structure. Systematic treatments of (some other instances of) the surface [u] as an underlying /ow/ before consonants are presented in Lightner (1965) and more recently in Itkin (2007: 147–148).

(32) post-accenting nouns

    a. kónču/kónčit 'finish.1SG/3SG'   (stem)
       (cf. koncá/koncámi 'end.SG.GEN/PL.INS')

    b. žen^jú/žénit 'marry.1SG/3SG'   (1SG)
       (cf. žená/ženú 'wife.NOM/ACC'

    c. strujú/struít 'stream.1SG/3SG'   (inflection)
       (cf. strujá/strujú 'stream.NOM/ACC')

(33) unaccented nouns

    a. prizeml^jú/prizemlít 'land.1SG/3SG'   (stem)
       (cf. zeml^já/zéml^ju 'ground.NOM/ACC')

    b. poručú/porúčit 'entrust.1SG/3SG'   (1SG)
       (cf. ruká/rúku 'hand.NOM/ACC'

    c. boron^jú/boronít 'harrow.1SG/3SG'   (inflection)
       (cf. boroná/bóronu 'harrow.NOM/ACC')

Derivation with the suffix *-ow-* may preserve the accent of the base noun (34a), or may override it (34b). Yet if the suffix *-ow-* were unaccented, we would expect to find at least some verbs derived from a post-accenting noun that would end up with stress on the suffix itself. The fact that instead in the past we find stress on the thematic suffix, as in (34c), strongly suggests that the suffix *-ow-* is post-accenting. Indeed, as shown by Garde (1998: 126), in the sequence of two post-accenting morphemes the second accent wins (thus violating the Basic Accentuation Principle (2)).[20]

---

[20]Garde (1998: 126) illustrates this with the derivation in (34c), which can also be accounted for by the assumption that *-ow-* cannot bear an accent (except by retraction). Garde (1998: 131) further provides an example of a post-accenting root followed by the post-accenting diminutive suffix *-ĭc-*, where stress is final (i.a), and the same result is obtained with the post-accenting diminutive suffix *-ŭk-* (i.b). Since, however, both suffixes contain yers, which are known to be unstressable, these cases are also non-definitive.

(i)   a. dvor/dvorí 'yard.SG.NOM/PL.NOM'
       dvoréc/dvorcí 'palace.SG.NOM/PL.NOM'

    b. kazák/kazakí 'Cossack.SG.NOM/PL.NOM'
       kazačók/kazačkí 'boy-servant.SG.NOM/PL.NOM'

Garde's generalization, however, is supported also by the suffix *-ič-* (cf. *moskvičí* 'denizens of Moscow' from the post-accenting *Moskvá*/*Moskvú* 'Moscow.NOM/ACC'). Conversely, it should also be noted that a sequence of two post-accenting diminutive suffixes *-ŭk-* yields stress on the first one (e.g., the unaccented root *-vetĭr-* 'wind' yields a post-accenting simple diminutive *veterók*/*veterká* 'wind.DIM.NOM/GEN' and a stem-accented double diminutive *veteróček*/*veteróčka* 'wind.DIM.NOM/GEN').

(34) a. accented:
      uród/uródɨ 'ugly person.SG.NOM/PL.NOM' → uródovat^j 'to disfigure'

   b. accented:
      krítika/krítiki 'critique.SG.NOM/PL.NOM' → kritikovát^j 'to critique'

   c. post-accenting:
      vrač/vračí 'doctor.SG.NOM/PL.NOM' → vračevát^j 'to treat, heal'

If the stem-final pattern (Table 8-d) is associated with a post-accenting L-stem, as suggested by verbs in *-ow-*, it seems reasonable to hypothesize that the 1SG stress correlates with an unaccented L-stem, and stress-initial nouns in *-ot-* provide further tentative evidence in favor of this view for verbs in *-ot-*. In the next subsection I will offer additional support for the assumption that the 1SG stress pattern corresponds to an unaccented L-stem.

5.2 Two 2nd conjugation *-a-/-i-* verbs

There exist two second-conjugation verbs with the thematic suffix *-a-* in the past: *gnát^j* 'to chase' and *spát^j* 'to sleep', which both exhibit accentual variability in the past:

(35) a. gnalá/gnáli 'chase.PST.FSG/PL'
   b. spalá/spáli 'sleep.PST.FSG/PL'

As demonstrated in Section 2, accentual variability in the past is a diagnostic of the lack of a preceding accent, i.e., both the roots and the thematic suffix in (35) are unaccented. While *gnát^j* 'to chase' shows the 1SG pattern in the present, the root of the verb *spát^j* 'to sleep' is asyllabic, so its accentuation could conform to any of the three patterns:

(36) a. gon^jú/gónit 'chase.1SG/3SG'
   b. spl^jú/spít 'sleep.1SG/3SG'

The 1SG present-tense stress pattern can therefore be taken as an indication that the L-stem is unaccented. The post-accenting stem can then be assumed to give rise either to the systematic post-stem stress (with second-conjugation *i*-verbs and occasional *e*-verbs, as well as with six semelfactive *nu*-verbs) or to consistent stem-final stress (with *-a-/-i-* verbs).

The two questions to address next are (i) how the 1SG pattern is derived, and (ii) why *i*-verbs, *nu*-verbs and *e*-verbs also give rise to the post-stem stress pattern while *-a-/-i-* verbs surface with stem-final stress. I will propose that the 1SG

pattern is due to an accentual conflict arising from the hiatus resolution with an accented vowel, and that post-accenting L-stems create two ways of avoiding this conflict, depending on the fate of the thematic vowel.

6 The 1sg pattern as an accentual conflict

I have suggested (Section 4) that the 1sg pattern arises when the present-tense suffix is absent from the metrical tier (being either not projected there or removed from it). I have also provided evidence (Section 5) that the 1sg pattern is associated with unaccented L-stems and arises when the thematic vowel is deleted before the vowel of the present-tense suffix. The lack of accentual variability in the present tense of athematic verbs as contrasted with their past tense shows that the present-tense suffix -*e*- introduces an accent (Section 2.3). Likewise, the lack of accentual variability in the past tense of most thematic verbs (Section 3.1) entails the same for thematic suffixes, and with the unaccented thematic suffix -*a*- and in athematic verbs the 1sg pattern is exceptional.

While so far I have been tacitly assuming that the present-tense suffix and the thematic suffix are accented, I am now going to revise this assumption and propose that the thematic suffixes giving rise to the 1sg pattern are post-accenting. Since the past-tense suffix -*l*- has been argued to be retracting (Section 2), this assumption makes no difference in the past tense of thematic verbs, as illustrated in (37); for the sake of intelligibility the foot boundary introduced by post-accentuation is indicated by a square bracket:

(37)   Past-tense retraction

```
        *   *[    (*)
    a.  √- nu   l    a
           TH   PST F.SG
                ↓
            *
        *  [*     (*)
    b.  √- nu   l    a
           TH   PST F.SG
```

The situation is different in the present, where the vowel of the post-accenting thematic suffix is deleted before the vowel of the accented present-tense suffix. My intuition here is that the removal of the present-tense suffix from the metrical tier is due to a conflict that is created by this deletion. On the assumption that the accent of a deleted vowel remains after deletion and is associated to the same syllable (defined from the left by the consonant(s) of the first syllable, and

from the right, by the nucleus of the second one), this syllable would receive contradictory instructions: to project an accent on the metrical tier and to shift this accent one syllable to the left. I represent this conflict in (38a–38b), with the right parenthesis at the right edge of the word deleted in (38b) because it is immediately preceded by another parenthesis (see Section 2.2):[21]

$$
\begin{array}{lll}
* & *[(*) \\
\sigma & \sigma & \sigma \\
| & | & | \text{\textbackslash} \\
\end{array}
$$

(38) a. √- nu   e ʼt
          TH   PRS 3SG
                ↓

        *   (*[
        σ   σ
        | ⟋ | \
    b. √- nú   e ʼt
         TH   PRS 3SG
                ↓

        *   ([
        σ   σ σ
        | ⟋ | \
    c. √- nú   e ʼt
         TH   PRS 3SG
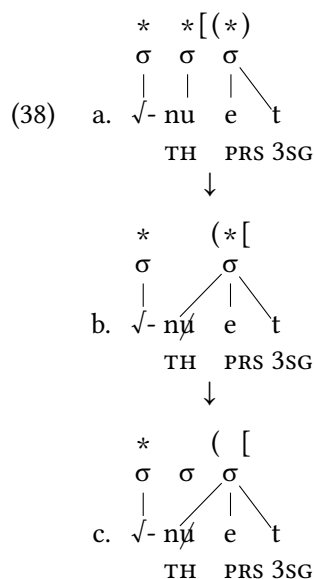
The representation in (38b–38c) makes explicit the relation between thematic vowel deletion and an accentual conflict: Since the deletion of the thematic vowel triggers resyllabification of the resulting phonological string while retaining the lexically specified instructions for the metrical tier, the rebuilt syllabic structure is subject to conflicting instructions. This is shown in (38b): The same syllable cannot be simultaneously accented and post-accenting. I propose that the problematic position is deleted from the metrical tier (38c).

Once again, as no stress-bearing elements are contained between the two parentheses in (38c), the second parenthesis is deleted (39a). Because the remaining parenthesis ends up word-final, stress, like with post-accenting stems, will surface on the final syllable of the stem (39b):

---

[21]The order of the two parentheses is changed to emphasize that "[" forces the accent on the next syllable, but the deletion of the present-tense suffix from the metrical tier means that the accent would be assigned to the agreement suffix also if the order is maintained. Crucially, the deletion of an extra parenthesis has to follow hiatus resolution.

```
          *     (
          σ     σ
          |    / |  \
(39)   a. √- n  e    t
          TH PRS 3SG
               ↓
         (*     (
          σ     σ
          |    / |  \
       b. √- n  e    t
          TH PRS 3SG
```
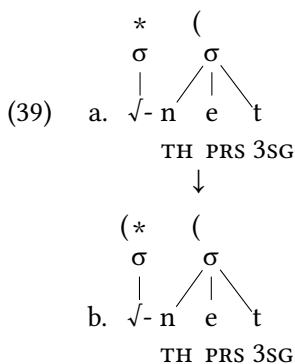
If, on the other hand, the present-tense suffix is followed by a syllabic suffix, i.e., the 1SG -*u*-, as in (40a), the present-tense gerund -*ʲa*- or the imperative -*i*-, the thematic vowel is deleted (40b). After this deletion the present-tense suffix is removed from the metrical tier (40c), and then the present-tense suffix is deleted before another vowel (40d) and its accent is realized on the vowel of the 1SG suffix:

```
           *    *[(*   *)
           σ    σ  σ    σ
           |    |  |    |
(40)   a.  √- nu    e    u
           TH     PRS 1SG
                 ↓
           *       (*[  *)
           σ       σ    σ
           |      / |   |
       b.  √- nu̶    e    u
           TH     PRS 1SG
                 ↓
           *      (     *)
           σ      σ     σ
           |     / |    |
       c.  √- nu̶    e    u
           TH     PRS 1SG
                 ↓        *
           *(            *
           σ            σ
           |           /  |
       d.  √- n   e̶     u
           TH  PRS 1SG
```
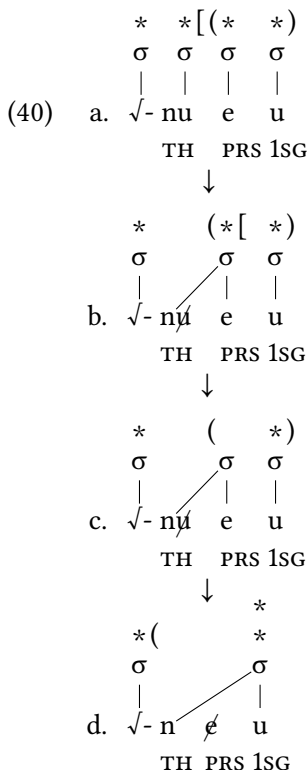
Problematically, the Halle-Idsardi model does not have the means to express the intuition that unstressability results from an accentual conflict. This is not purely a matter of notation: In this model, parentheses on line 0 of the metrical tier represent foot boundaries and post-accentuation is implemented by placing a parenthesis on the next asterisk. There is therefore no difference between the illicit (*ex hypothesi*) structures in (38b) or (40b) and the licit structures created by a sequence of two accented (41a) or post-accenting (41b) morphemes:

(41) a.
```
      *
    (*    (*)
    Maš    a
```
Masha SG.NOM

   b.
```
            *
    *[    *[    *)
    Moskv  ič    a
```
Moscow NMLZ SG.GEN

As discussed in fn. 20, the Basic Accentuation Principle (2) incorrectly predicts initial stress in cases like (41b). This fact suggests that post-accentuation is indeed a process, as proposed by Garde (1998) and Melvold (1989): In Melvold's approach post-accentuation is represented as a diacritic forcing post-cyclic movement of the appropriate parenthesis one syllable to the right, which is why it yields the correct outcome for (42). Nonetheless, neither Melvold's approach nor Garde's predict that the structure resulting from (38b) should be in any way problematic.

(42) a.
```
    (*p   (*p    *)
    Moskv  ič    a
```
Moscow NMLZ SG.GEN

```
              ↓
    *    ((*p    *)
    Moskv  ič    a
```
   b. Moscow NMLZ SG.GEN

```
              ↓    *
    *    *    (*)
    Moskv  ič    a
```
   c. Moscow NMLZ SG.GEN

I will continue to use the enriched representation with square brackets because I believe that it not only encodes a valuable intuition about the source of the un-stressability of the present-tense suffix, but also makes it possible to explain how

this effect is nullified when the L-stem is post-accenting (see Section 5 on the cor-relation between unaccented -*ot*- verbs with the 1sg pattern and post-accenting -*ow*- verbs with the retracting pattern). More specifically, in section Section 6.1 I will show how post-stem stress is correctly predicted for post-accenting L-stems, and in section Section 6.2, how -*a*-/-*i*- verbs with post-accenting L-stems give rise to the stem-final pattern alternating with the 1sg pattern.

6.1 The role of a post-accenting stem

As discussed in Section 5.1, the sequence of two post-accenting morphemes, as in (41b), does not obey the Basic Accentuation Principle (2): Whereas in Halle's (1997) framework the surface stress is expected to coincide with the first accent, the real outcome is the same as if the first morpheme were unaccented. However, as will be shown now, the hypotheses I developed so far give rise to the correct outcome in a structure like (43a), where a post-accenting L-stem is followed by a post-accenting thematic suffix.

As the vowel of the thematic suffix is followed by the vocalic present-tense suffix, the former is deleted, (43b). Once again, a sequence of two parentheses with no metrical elements between them (set in a box) is simplified to a single parenthesis and here it is crucial that the one deleted is the second one, yielding (43c). Note that the derivation proceeds left to right, so clash resolution precedes and bleeds the creation of an accentual conflict:

(43)   a.
```
        *[  *[(*)
        σ   σ   σ
        |   |   |\
   √- nu      e    t
        TH   PRS 3SG
              ↓
```

   b.
```
        *[⬚(|*[
        σ        σ
        |        |\
   √- nⱮ     e    t
        TH   PRS 3SG
              ↓
```

   c.
```
        *[        *[
        σ         σ
        |         |\
   √- nⱮ     e    t
        TH   PRS 3SG
```
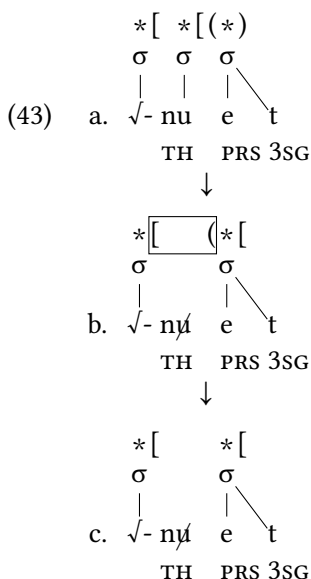
The structure in (43c) is clearly distinct from that in (38b): Here no conflicting instructions are associated to the same syllable. As a result, nothing is deleted from the metrical tier and the final post-accentuation is resolved to final stress.[22]

To recap, with a post-accenting L-stem the deletion of the thematic vowel and the subsequent reassignment of its bracket to the present-tense suffix creates a metrical structure (43b) that is identical to the combination of a post-accenting stem with an accented suffix (8a), which is resolved by the deletion of the second accentual mark (43c) in a manner fully parallel to (8b). In other words, a post-accenting stem prevents the creation of an accentual conflict and stress is thus correctly predicted to fall on the present-tense suffix.

While *e*-verbs and *i*-verbs (combining with the null present-tense suffix but yielding the same metrical structure as *nu*-verbs) will be discussed in Section 6.3, in the next subsection I turn to the derivation of the stem-final present-tense pattern of *-a-/-i-* verbs.

6.2 The accentual patterns of *-a-/-i-* verbs

The class of *-a-/-i-* verbs is characterized by the thematic suffix *-a-* in the infinitive and the past tense and by transitive softening (fn. 5) in the present. As transitive softening is known to arise from an underlying consonant-glide sequence, the thematic suffix is assumed to surface as [i] in the present tense, either as a result of a readjustment rule (Bethin 1992) or due to ablaut triggered by the present-tense suffix (Matushansky 2023a).

As discussed in Section 5.1, *-a-/-i-* verbs lack the post-stem pattern, which seems to be replaced by the stem-final one, as shown in Table 8, repeated below as Table 9.

The main difference between the thematic suffixes *-a-/-i-* and *-nu-* lies in the fate of the vowel: While the thematic vowel of *-nu-* is deleted in the present tense, the *-a-/-i-* suffix (or rather, its *-i-* allomorph) turns into a glide. I propose that this difference can derive the observed stress retraction with post-accenting stems.

As discussed above, when the thematic suffix *-nu-* follows an unaccented L-stem, the deletion of the thematic vowel before another vowel gives rise to an accentual conflict (38b). The problematic position (the present-tense suffix) is then deleted from the metrical tier (38c), and the resulting post-accenting stem is realized with stem-final stress. The same outcome is correctly expected to arise with the thematic suffix *-i-* (44).

---

[22]If a left parenthesis is inserted before the final syllable, as in (39b), it will be deleted, as in (43b–43c). This is why I skip these steps in the derivation in (43).

Table 9: Accentual interaction with the 1st conjugation suffix -*a*-/-*i*-

		accented PRS-3SG	accented PRS-1SG	accented PST-F.SG	unaccented PST-PL
a.	stem: -*maz*- 'smear'	máž-e-t	máž-u	máz-a-l-a	máz-a-l-i
b.	post-stem: N/A; potential candidates among *j*-final stems can be assigned to the -*a*-/-Ø- class				
c.	1SG: -*v^{j}az*- 'tie'	v^{j}áž-e-t	v^{j}až-ú	v^{j}az-á-l-a	v^{j}az-á-l-i
d.	stem-final present: -*koleb*- 'rock'	kolébl^{j}-e-t	kolébl^{j}-u	koleb-á-l-a	koleb-á-l-i

(44)

a.
```
*  *[(*)
σ  σ  σ
|  |  |  \
√- i  e   t
   TH PRS 3SG
        ↓
```

b.
```
*     (*[
σ      σ
|    / |  \
√- j  e    t
   TH PRS 3SG
        ↓
```

c.
```
*    (  [
σ     σ
|   / |  \
√- j  e    t
   TH PRS 3SG
```

To obtain the desired outcome (i.e., stem-final stress) for post-accenting stems, I capitalize on the difference between vowel deletion and glide formation. While the accent of the thematic suffix (indicated by shading in (45)) remains on the same syllable in both cases, I hypothesize that this is not true for the accent of the post-accenting stem (set in a box). I propose that if a vowel turns into a glide, the accent that would be assigned to it behaves like a word-final accent in

that it is realized on the assigning syllable, i.e., on the final syllable of the stem (concurrently with the creation of a conflicting position, as in (45b–45c)):

(45) a.
```
    *[* [ (*)
    σ σ   σ
    | |   | \
    √- i   e   ˋt
       TH  PRS 3SG
```
↓

b.
```
    *[]  [(*[
    σ      σ
    |    /| \
    √- j   e   ˋt
       TH  PRS 3SG
```
↓

c.
```
    []*    (*[
    σ      σ
    |    /| \
    √- j   e   ˋt
       TH  PRS 3SG
```
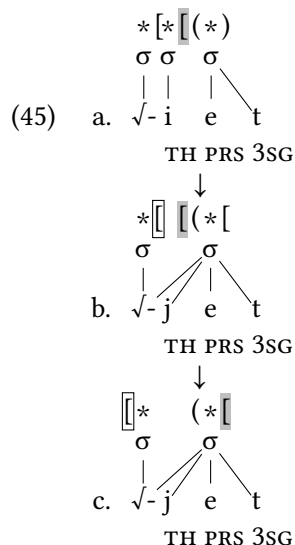
As a result, even though the present-tense suffix is deleted from the metrical tier (45d), stress surfaces on the final syllable of the L-stem (45e):

(45) d.
```
    [*   ( [
    σ     σ
    |   /| \
    √- j  e  ˋt
       TH PRS 3SG
```

e.
```
    *    ↓
    [*   ( [
    σ     σ
    |   /| \
    √- j  e  ˋt
       TH PRS 3SG
```
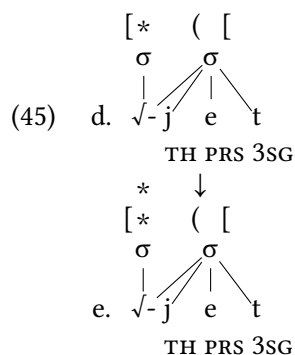
The natural question arises why glide formation makes a post-accenting stem become accented, while vowel deletion does not. Beyond noting that both strategies (accent retraction and accent advancement) seem equally valid outcomes for the disappearance of an accented vowel, I can provide no answer for the choice of strategy, it could also be lexically determined. Importantly, it has to be the post-accentuation of the preceding syllable that is affected by glide formation, since, as I will now show, the accent of the thematic suffix *-i-* is not retracted when it forms a glide in the 1SG.

6.3 Second-conjugation verbs and the derivation of the 1sg pattern

As noted above, the difference between the first and the second conjugations in Russian lies in the realization of the present-tense suffix: While in the first conjugation it is *-e-*, in the second conjugation it is zero. The paradigms in Table 1 and Table 6, repeated in Table 10 and Table 11, illustrate two facts: firstly, that second-conjugation verbs manifest two thematic suffixes in the past tense, *-e-* and *-i-*, both corresponding to /i/ in the present, and secondly, that in the present both classes of verbs exhibit the same three stress patterns as *nu*-verbs: stem stress, post-stem stress and the 1sg pattern.

Table 10: Accentual interaction in thematic verbs, illustrated for the thematic suffix *-i-*

		PRS-1SG	PRS-3SG	PST-F.SG	PST-PL
a.	stem: *-žal-* 'sting'	žál^j-u	žál-i-t	žál-i-l-a	žál-i-l-i
b.	post-stem: *-govor-* 'speak'	govor^j-ú	govor-í-t	govor-í-l-a	govor-í-l-i
c.	1sg: *-l^{j}ub-* 'love'	l^{j}ubl^j-ú	l^{j}úb-i-t	l^{j}ub-í-l-a	l^{j}ub-í-l-i

Table 11: Accentual interaction in thematic verbs, illustrated for the thematic suffix *-e-*

		accented PRS-3SG	accented PRS-1SG	accented PST-F.SG	unaccented PST-PL
a.	stem: *-vid-* 'see'	víd-i-t	víž-u	víd-e-l-a	víd-e-l-i
b.	post-stem: *-vel-* 'order'	vel-í-t	vel-ú	vel-é-l-a	vel-é-l-i
c.	1sg *-vert-* 'spin'	vért-i-t	verč-ú	vert-é-l-a	vert-é-l-i

Two types of explanations have been given for the lack of the thematic suffix *-e-* in the present tense. One proposal (Jakobson 1948, Melvold 1989) is that the

thematic vowel *-e-* is deleted before the present-tense suffix *-i-* (46a). The alternative (Micklesen 1973, Coats & Lightner 1975, Itkin 2007: 129–130, Matushansky 2023b) is that the second-conjugation present-tense suffix is null, and the thematic vowel *-e-* is raised to [i] in the present tense (46b).

(46)   a. $[[[gor\text{-}e]_2\text{-}i]_3\text{-}t]_4 \rightarrow [[[gor\text{-}é]_2\text{-}i]_3\text{-}t]_4 \rightarrow$ [gorit]   (vowel deletion)

     b. $[[[gor\text{-}e]_2\text{-}\emptyset]_3\text{-}t]_4 \rightarrow [[[gor\text{-}i]_2\text{-}\emptyset]_3\text{-}t]_4 \rightarrow$ [gorit]   (vowel change)

With the former approach the derivation of the three accentual patterns proceeds along exactly like for *nu*-verbs. In the latter approach to obtain the 1SG pattern and its nullification with post-accenting L-stems it is necessary to assume that the null present-tense suffix also introduces an accent.

I begin with a concrete verb (Table 10c) exhibiting the 1SG pattern: stem-final stress before consonantal suffixes and final stress on vocalic ones. I propose that, as before, the thematic suffix is post-accenting while the null present-tense suffix introduces an accent (47a). Since the present-tense suffix is non-segmental, its accent is assigned to the syllable of the thematic suffix (47b), and the resulting accentual conflict leads to the deletion of the thematic suffix from the metrical tier (47c). As in (38c) and (44a), superfluous parentheses are removed and stress is realized on the final syllable of the stem:

(47)   a.
```
       *  *[(
       σ  σ
       /|  /|
    lʲub i   ∅   t
       √   TH PRS 3SG
              ↓
```
     b.
```
       *(*[
       σ  σ
       /|  /|
    lʲub i   ∅   t
       √   TH PRS 3SG
              ↓
```
     c.
```
       *(  [
       σ  σ
       /|  /|
    lʲub i   ∅   t
       √   TH PRS 3SG
```

In the 1SG, on the other hand, the problematic position is followed by another vowel (48a), which should (and does) turn the thematic suffix into a glide. Recall that in Section 6.2 the thematic suffix -*a*-/-*i*- formed a glide in the present tense, which was taken as the reason why -*a*-/-*i*- verbs exhibit stem-final stress: I proposed that if a vowel turns into a glide, the accent assigned to it is shifted one syllable to the left (45). Is (48) incorrectly predicted to also give rise to stem-final stress?

```
              * ( * [        * )
              σ  σ           σ
              /| /|          |
(48)   a.  lʲub i    ∅    u
              √    TH PRS 1SG
                      ↓

              * (   [        * )
              σ  σ           σ
              /| /|          |
       b.  lʲub i    ∅    u
              √    TH PRS 1SG
```

The prediction is avoided because glide formation is timed differently in the two derivations. In (45) glide formation both precedes and causes the creation of a problematic position, while in (48) the accentual conflict deleting the thematic suffix from the metrical tier arises before its conversion into a glide: First the accentual conflict is resolved by the deletion of the problematic position from the metrical tier (48c), and then a glide is formed (48d). As this glide formation cannot affect accentuation, stress falls on the vowel of the 1SG ending (48e). Stress assignment in *i*-verbs therefore provides an argument for a cyclic approach to Russian accentuation.

```
              * (            * )
              σ  σ           σ
              /| /|          |
(48)   c.  lʲub i    ∅    u
              √    TH PRS 1SG
                      ↓
```

```
        * (        *)
        σ          σ
        /|  _____/
d.  lʲub j    ∅    u
    √    TH  PRS  1SG
              ↓
                   *
        * (        *)
        σ          σ
        /|  _____/
e.  lʲub j    ∅    u
    √    TH  PRS  1SG
```
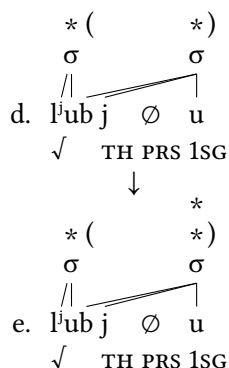
If the L-stem is post-accenting, the derivation proceeds as in (43): The accentual conflict is prevented because, before the present-tense suffix can influence the outcome, the second of the two accents not divided by metrical material is deleted.

6.4 Summary

I have proposed that the 1SG pattern arises from induced unstressability: An accented suffix is deleted from the metrical tier when it receives two conflicting accentual specifications, which is what happens when the vowel of a post-accenting suffix turns into a glide or is deleted before an accented vowel. When the L-stem is post-accenting, the accent of the thematic suffix has to be deleted, which straightforwardly derives the post-stem stress with *-nu-* verbs. For *-a-/-i-* verbs an additional assumption is required (cf. Melvold 1989: 254) that when a glide is formed, the accent assigned to it shifts to the preceding syllable.

The advantage of this view is that it derives the three stress patterns from the independently motivated property of L-stem accentuation: Unaccented stems exhibit the 1SG pattern and post-accenting stems surface with consistent stress position unless a glide is formed.

7 Conclusion and questions for future research

I have proposed that the accentual behavior of thematic verbs in the present tense can be linked transparently to the accentual specification of the L-stem and to the accentuation of the thematic suffix. The combination of an unaccented L-stem and a post-accenting thematic suffix creates a configuration that makes the present-tense suffix unstressable by forcing it off the metrical grid. A post-accenting L-stem is hypothesized to remove the problematic thematic accent and

so not to create such a problem, yielding post-stem stress for all thematic suffixes, except *-a-/-i-*, which yields stem-final stress because of glide formation. Since both the first- and second-conjugation present-tense suffixes come into conflict with a post-accenting thematic suffix, it is apparently not the concrete vowel that has this property, but rather the abstract morpheme.[23] The natural question is whether nominal stress retraction (Halle 1973, 1975, 1997, Melvold 1989, Revithi-adou 1999, Alderete 1999, Butska 2002, Dubina 2012, Osadcha 2019, etc.) can be accounted for by the same mechanism. Given that nouns can exhibit retraction in the singular and in the plural and that both unaccented and post-accenting nouns can trigger it, more work is needed to determine if nominal retraction is the same phenomenon. The same issue arises for adjectival retraction.

The empirical contributions of this study include the facts that the 1sG pattern is dependent on the deletion of an accent-bearing thematic suffix, that it is not equally frequent with different thematic suffixes and that it correlates with an unaccented L-stem. This approach can explain why the thematic suffixes *-aj-* and *-ej-* do not give rise to the 1sG pattern: As their vowels are not deleted before the present-tense suffix, their accent will not shift. The reason why the non-productive mutative suffix *-nu-* and the thematic suffix *-a-/-Ø-* do not yield the 1sG pattern is that the latter is not accented, and the former is pre-accenting, so accentuation is not affected by the deletion of their thematic vowel.[24]

The alternation of the stem-final stress pattern for the *-a-/-i-* suffix with the post-stem pattern for all other thematic vowels triggering the 1sG pattern has allowed us to determine the thematic suffix for some *j*-final verbs.

While the intuition that the 1sG pattern arises from induced unstressability can be accounted for in the terms of the Halle-Vergnaud framework, the hypothesis that this unstressability is due to an accentual conflict between post-accentuation and accentuation cannot be expressed with the tools of this framework: Accented and post-accenting morphemes in it have the same effect, the only distinction be-ing the position of the accent. Though I have adjusted the notation to encode the postulated difference between accentuation and post-accentuation, this change goes against the core principles of the framework, where a parenthesis indicates a foot boundary rather than an instruction to include or not include the carrier syllable into the foot created. Since I believe that this enrichment makes it pos-

---

[23]While I have chosen to present this analysis as a series of representations, it can be equally easily cast in a rule-based framework and in OT.

[24]Intuitively, pre-accentuation operates on the already existing structure, unlike post-accentuation, which is an instruction for the structure to be built, so it is reasonable to assume that at the present-tense cycle the accent of a pre-accenting suffix has already been assigned to the stem-final syllable.

sible to account for rather complex phenomena, the question arises whether the Halle-Vergnaud framework can be made compatible with this more complex notation, or another framework should be used, where the simultaneous placement of a bracket and a parenthesis on the same metrical position can be represented as a conflict of instructions, e.g., with post-accentuation representing the tail of an iambic foot or by treating the two types of accents as tones (Matushansky 2023b).

A number of empirical questions remain. Some, like the interaction of vowel deletion and glide formation with secondary imperfective suffixes, will be left for future research. Others, like the persistence of stem accentuation across verb classes and the derivation of the 1SG pattern in verb classes where it is an exception, will be discussed in the remainder of this section.

7.1 Unproductive 1SG pattern: Athematic verbs, *e*-verbs, *nu*-verbs, and *a*-verbs

In this subsection I will discuss verb classes for which the 1SG pattern is attested only with a small number of verbs. As in one of these classes (Section 7.1.1) the 1SG pattern is manifested only when the stem is prefixed with two specific prefixes, I will propose that in all these cases the 1SG pattern arises from idiosyncratic lexical specification.

7.1.1 Two athematic verbs

As mentioned in fn. 9, there are two athematic stems giving rise to the 1SG pattern in the derived verbs: *-mog-* (*moč^j* 'to be able', *pomóč^j* 'to help') and the cranberry root *-im-/-n^{j}a-* (*prin^{j}át^j* 'to accept', *podn^{j}át^j* 'to raise', *obn^{j}át^j* 'to hug', etc.).

Table 12: Two athematic 1SG pattern verbs

		PST-F.SG	PST-PL	PRS-1SG	PRS-3SG
a.	*-pri.m/pri.n^{j}a-* 'accept'	pri.n^{j}a-l-á	prí.n^{j}a-l-i	pri.m-_-ú	prí.m-e-t
b.	*-mog_-* 'be able'	mog_-l-á̱	mog_-l-í	mog_-_-ú	móž_-e-t

The verb *prin^{j}át^j* 'to accept' in Table 12a exhibits accentual variability in the past, as expected from an unaccented stem, while the verb *moč^j* 'to be able' in Table 12b surfaces in the past with consistent word-final stress that is indicative of a post-accenting stem. While at first blush it might seem that these facts argue

against the link between an unaccented stem and the 1SG pattern, there is no thematic vowel deletion here to create an accentual conflict. These verbs can be regarded as lexically specified to delete the present-tense suffix from the metrical tier.

7.1.2 Six *-nu-* verbs

There are only six *nu*-stems that exhibit the 1SG stress pattern in the present tense. Four of them form perfective verbs only (*pom^{j}anút^j* 'to remember', *obmanút^j* 'to cheat', *vzgl^{j}anút^j* 'to glance', and *minút^j* 'to elapse') and can therefore be regarded as semelfactive, while two are imperfective (*tonút^j* 'to drown' and *t^{j}anút^j* 'to pull') and may involve the same suffix *-nu-* as mutative verbs.

As discussed above, the imperfective suffix *-nu-* is pre-accenting, so its deletion does not give rise to an accentual conflict. I propose that the reason why the semelfactive *-nu-* does not give rise to the 1SG pattern is that it is accented. For both types of *-nu-* I propose that the six stems above force the suffix to become post-accenting. As (49) shows, such exceptional behavior can target some prefixed derivatives of a particular root:

(49)   *-gl^{j}ad-* 'look'
      a.   i. gl^{j}ánu/gl^{j}ánet 'will glance.1SG/3SG (semelfactive suffix *-nu-*)
                                                                (stem)
          ii. progl^{j}ánet 'will glance through.3SG, impers.' (ibid.)
      b. gl^{j}ažú/gl^{j}adít 'look.1SG/3SG' (suffix *-e-*)   (post-stem)
      c.   i. vzgl^{j}anú/vzgl^{j}ánet 'will glance.1SG/3SG'   (1SG)
          ii. zagl^{j}anú/zagl^{j}ánet 'will look in on.1SG/3SG', etc.

The stem stress in (49a) suggests that the root is accented, while the post-stem stress in (49b) is explained by the fact that *-e-* is dominant (see Section 7.1.3). However, the behavior of (49c) is unexpected both for an accented root and for the accented *-nu-*.[25] While the stipulation that the thematic suffix *-nu-* is exceptionally post-accenting in the prefixed verbs in (49c) accounts for their stress pattern, it cannot be independently motivated. Nonetheless, given that the combination of a prefixed stem and a thematic suffix can be semantically non-compositional or idiomatic, phonological unpredictability can also be accommodated.

---

[25]There is no clear difference in meaning between (49a-i) and (49c-i), but (49a-i) is either archaic or dialectal.

7.1.3 Five *e*-verbs

The same issue arises when the exceptional character of accentual variance with *e*-verbs is considered. Only five out of the ca. 80 second-conjugation *e*-verbs surface with stem stress (*slíšat^j* 'to hear', *zavíset^j* 'to depend', *vídet^j* 'to see', *ne-navídet^j* 'to hate', and *obídet^j* 'to offend', with the last three diachronically derived from the same root *-vid-*), which strongly suggests that the thematic suffix *-e-* is accentually dominant. Support for this claim comes from the fact that, on the basis of all *e*-verbs that have corresponding semelfactives (16 verbs) or mutatives (4 verbs), stem stress in *-nu-* verbs systematically corresponds to post-stem stress in *-e-* verbs. If the thematic suffix *-e-* is accented and dominant, it will remove the underlying accent of the L-stem:

(50) *-krik-* 'shout'
    a. kríknu/kríknet 'will give a shout.1SG/3SG' (semelfactive suffix *-nu-*)
                                                   (stem)
    b. kričú/ kričít 'shout.1SG/3SG' (suffix *-e-*)   (post-stem)

(51) *-perd-* 'fart' (vulgar)
    a. p^jórdnu/p^jórdnet 'will give a fart.1SG/3SG' (semelfactive suffix *-nu-*)
        (stem)
    b. peržú/perdít 'fart.1SG/3SG' (suffix *-e-*)   (post-stem)

(52) *-molk-* 'be silent'
    a. mólknu/mólknet 'be silent.1SG/3SG' (pre-accenting mutative suffix
      -nu-)   (stem)
    b. molčú/molčít 'be silent.1SG/3SG' (suffix *-e-*)   (post-stem)

If the dominant suffix *-e-* were post-accenting, we would wrongly expect systematic accentual variance, as in Table 11c: if the L-stem accent is removed, it becomes unaccented. However, only five *e*-verbs show the 1SG pattern (*deržát^j* 'to hold', *terpét^j* 'to tolerate', *smotrét^j* 'to look', *vertét^j* 'to turn', and *dišát^j* 'to breathe'). The prevalence of the post-stem pattern Table 11b in *e*-verbs (ca. 70 verbs out of 80) therefore strongly suggests that the suffix *-e-* is accented. No accentual conflict arises with the null accented present-tense suffix, and the Basic Accentuation Principle (2) predicts systematic surface stress on the thematic vowel, barring the five accented stems.

To derive the 1SG pattern the same analysis can be appealed to as that proposed for the thematic suffix *-e-* in the preceding subsection: Suppose that these five *e*-verbs take the post-accenting allomorph of the thematic suffix (or force it to become post-accenting).

7.1.4 Two 1SG *-a-*/*-∅-* verbs

To complete the empirical picture, the unproductive thematic suffix *-a-*/*-Ø-* is unaccented, as shown by the fact that it permits accentual variability in the past tense (fn. 12). Like athematic verbs though, this class also includes two verbs with the 1SG pattern:[26]

(53)   a.   srat[j]/serú/séret 'shit.INF.1SG/3SG'
       b.   stonát[j]/stoní/stónet 'moan.INF1SG/3SG'

Following the reasoning suggested above, I hypothesize that these roots are lexical exceptions triggering post-accentuation of the thematic suffix.

7.1.5 Summary

Given that four classes of verbs exceptionally give rise to the 1SG, which is regular in two other thematic classes, an appeal to lexical exceptions appears to be inevitable. As it does not seem reasonable to postulate a post-accenting allomorph for each of the three thematic suffixes for which the 1SG pattern constitutes an exception, I hypothesize that these stems can force post-accentuation of the thematic suffix.

Of the fifteen verbs in question (2 athematic verbs, 5 *e*-verbs, 6 *nu*-verbs and 2 *a*-verbs) only two have counterparts in other thematic classes that could have given rise to the 1SG pattern:

(54)   a.   dɨšát[j]: 'to breathe':
           -nu-: dɨxnú/dɨxn[j]ót 'provide a breathing sample.1SG/3SG'
       b.   minút[j]: 'to elapse':
           -ow-: minúju/minúeš[j] 'elapse.IPFV.1SG/3SG' (minovát[j] INF)

Though these two stems do not yield the 1SG pattern with other thematic suffixes, as shown in (54), they may be expected not to: *-ow-* is post-accenting, and the *-nu-* derivation may involve a different, if related, root.

---

[26]The 1SG and the gerund forms of the verb *stonát[j]* 'to moan' are ineffable (on paradigm gaps in the 1SG of Russian verbs see Sims (2006), Daland et al. (2007), Pertsova (2016), etc.). The form of its imperative is also compatible with the *-a-*/*-i-* theme, which may be the reason why it exists. The verb *srat[j]* 'to shit' has several conjugational variants, (53a) is merely one of them.

7.2 L-stem accentuation across verb classes

The evidence (Section 5.1) linking the 1SG pattern to unaccented L-stems is rather tenuous, but for each L-stem its accentuation, once determined for one verb class, is predicted to persist in another. To exclude some potential lines of further inquiry, I would like to report that I have found no correlation between the 1SG pattern and the form of the secondary imperfective. The accentual relation between semelfactive *nu*-verbs and their imperfective counterparts in *-i-* does not seem to be predictable either: Although all 1SG *i*-verbs that I have looked at have post-stem stress in the semelfactive, other stress patterns do not appear to be linked to each other (though (55d) seems to be exceptional in that it involves a valency change):

(55) a. katít^j 'to roll' (1SG), katnút^j (final)
      b. skol^jzít^j 'to slide' (post-stem), skol^jznút^j (post-stem)
      c. číllit^j 'to chill out' (stem), čil^jnút^j (post-stem)
      d. voskresít^j 'to resurrect' (post-stem), voskrésnut^j 'to be resurrected' (stem)

A brief examination of stress patterns in minimal pairs composed of semelfactive *nu*-verbs and their imperfective counterparts in *-a-/-aj-* also suggests that one form cannot be predicted from the other:

(56) a. bríznut^j (stem)/brízgat^j (stem) 'to spatter'
      b. zevnút^j (post-stem)/zevát^j (post-stem) 'to yawn'
      c. šmɨgnút^j (post-stem)/šmɨgat^j (stem) 'to dart'
      d. kínut^j (stem)/kidát^j (post-stem) 'to toss'

Even though the first two patterns with stress retention are the most frequent, the existence of the latter two requires an explanation, which does not seem to be provided by postulating any type of accent or lack thereof on the L-stem.

Derivational morphology is just as inconclusive: As shown by examples (31–33), there does not seem to be a transparent relation between the accentuation of a noun and that of the verb that it is derived from. I leave the issue of apparently indeterminate accentuation of thematic L-stems for future research.

Ora Matushansky

Abbreviations

1	first person	IPFV	imperfective
2	second person	M	masculine
3	third person	N	neuter
ACC	accusative	NMLZ	nominalizer
DIM	diminutive	NOM	nominative
F	feminine	PL	plural
GEN	genitive	PRS	present tense
GER	gerund	PST	past tense
INF	infinitive	SG	singular
INS	instrumental	TH	thematic suffix

Acknowledgments

References

Alderete, John. 1999. *Morphologically governed accent in Optimality Theory.* Amherst: University of Massachusetts, Amherst. (Doctoral dissertation).

Antonyuk, Svitlana, Predrag Kovačević & Stefano Quaglia (eds.). 2022–2023. *Thematic formatives and linguistic theory. Glossa special collection.* DOI: 10.16995/glossa.collection.777.

Bethin, Christina Y. 1992. Iotation and gemination in Ukrainian. *The Slavic and East European Journal* 36(3). 275–301. DOI: 10.2307/308583.

Boyd, Michael S. 1997. *Palatalization and coronalization in Russian and Czech: A non-linear approach.* Columbus: The Ohio State University. (Doctoral dissertation).

Brown, Dunstan. 1998. Stem indexing and morphonological selection in the Russian verb. In Ray Fabri, Albert Ortmann & Teresa Parodi (eds.), *Models of inflection*, 196–221. Tübingen: Niemeyer. DOI: 10.1515/9783110919745.196.

Butska, Luba. 2002. *Faithful stress in paradigms: Nominal inflection in Ukrainian and Russian.* New Brunswick: Rutgers University. (Doctoral dissertation).

Choi, Sung-ho. 1996. A study on Russian inflectional accentology. 러시아연구 [Russian studies] 6. 81–115. https://s-space.snu.ac.kr/handle/10371/87976.

Coats, Herbert S. & Theodore M. Lightner. 1975. Transitive softening in Russian conjugation. *Language* 51(2). 338–341. DOI: 10.2307/412859.

Crosswhite, Katherine. 1999. *Vowel reduction in Optimality Theory*. Los Angeles: University of California, Los Angeles. (Doctoral dissertation).

Crosswhite, Katherine. 2000. Vowel reduction in Russian: A unified account of standard, dialectal, and "dissimilative" patterns. In Katherine Crosswhite & Joyce McDonough (eds.), *University of Rochester working papers in the language sciences*, 107–171.

Daland, Robert, Andrea D. Sims & Janet Pierrehumbert. 2007. Much ado about nothing: A social network model of Russian paradigmatic gaps. In Annie Zaenen & Antal van den Bosch (eds.), *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 936–943. Prague: Association for Computational Linguistics. https://aclanthology.org/P07-1118.

Dubina, Andrei. 2012. *Towards a tonal analysis of free stress* (LOT Dissertation Series 313). Utrecht: LOT.

Esʹkova, Natalja A. 2008. *Normy russkogo literaturnogo jazyka XVIII–XIX vekov: Slovar'. Pojasnitel'nye stat'i* [Norms of the Russian literary language of the XVIII–XIX centuries: Dictionary. Clarifying articles]. Moskva: Jazyki slavjanskoj kul'tury.

Esʹkova, Natalja A. 2014. *Slovar' trudnostej russkogo jazyka: Udarenie. Grammatičeskie formy* [The dictionary of Russian language difficulties: Stress. Grammatical forms]. Moskva: Jazyki slavjanskoj kul'tury.

Feldstein, Ronald F. 1986. The Russian verbal stress system. *International Journal of Slavic Linguistics and Poetics* 33. 43–61.

Feldstein, Ronald F. 2015. The stress of the Russian verb: A new interpretation. Paper presented at *Dr. Ronald Feldstein Web Lecture Series on Slavic linguistics*, Duke University, August 24, 2015. https://slaviccenters.duke.edu/programs/lectures/dr-ron-feldstein-russian-verb-stress.

Garde, Paul. 1998. *Grammaire russe: Phonologie et morphologie*. 2nd edition. [First published in 1980]. Paris: Institut d'études slaves.

Gladney, Frank Y. 1995. The accent of Russian verbforms. *Journal of Slavic Linguistics* 3(1). 97–138. https://www.jstor.org/stable/24598997.

Gouskova, Maria. 2010. The phonology of boundaries and secondary stress in Russian compounds. *The Linguistic Review* 27(4). 387–448. DOI: 10.1515/tlir.2010.015.

Halle, Morris. 1963. O pravilax russkogo sprjaženija [About the rules of Russian conjugation]. In *American contributions to the Fifth International Congress of Slavists 1, September 1963, Sofia*, 113–132. The Hague: Mouton.

Halle, Morris. 1973. The accentuation of Russian words. *Language* 49(2). 312–348. DOI: 10.2307/412457.

Halle, Morris. 1975. On Russian accentuation. *The Slavic and East European Journal* 19. 104–111. DOI: 10.2307/306217.

Halle, Morris. 1997. On stress and accent in Indo-European. *Language* 73. 275–313. DOI: 10.2307/416020.

Halle, Morris. 2004. *Russian phonology: The core.* Ms. MIT. http://www.ai.mit.edu/projects/dm/featgeom/halle-russianverb.pdf.

Halle, Morris & William J. Idsardi. 1995. General properties of stress and metrical structure. In John Goldsmith (ed.), *Handbook of phonological theory*, 403–443. Oxford: Blackwell.

Halle, Morris & Jean-Roger Vergnaud. 1987a. *An essay on stress.* Cambridge, MA: MIT Press.

Halle, Morris & Jean-Roger Vergnaud. 1987b. Stress and the cycle. *Linguistic Inquiry* 18. 45–84. http://www.jstor.org/stable/4178524.

Idsardi, William J. 1992. *The computation of prosody.* Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Itkin, Ilja B. 2007. *Russkaja morfonologija* [Russian morphonology]. Moskva: Gnozis.

Jakobson, Roman. 1929. Remarques sur l'évolution phonologique du russe comparée à celle des autres langues slaves. *Travaux du Cercle Linguistique de Prague* 2.

Jakobson, Roman. 1948. Russian conjugation. *Word* 4(3). 155–167. DOI: 10.1080/00437956.1948.11659338.

Kayne, Richard S. 1967. *Against a cyclic analysis of Russian segmental phonology.* Ms. MIT.

Kiparsky, Paul & Morris Halle. 1977. Towards a reconstruction of the Indo-European accent. In Larry M. Hyman (ed.), *Studies in stress and accent* (Southern California Occasional Papers in Linguistics 4), 209–238. Los Angeles: Department of Linguistics, University of Southern California.

Kortlandt, Frederik. 1994. From Proto-Indo-European to Slavic. *Journal of Indo-European Studies* 22. 91–112.

Lightner, Theodore M. 1965. *Segmental phonology of Contemporary Standard Russian.* Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Lightner, Theodore M. 1967. On phonetic nasal ~ *a* alternations in modern Russian verb forms. In *To honor Roman Jakobson: Essays on the occasion of his 70th birthday, 11. October 1966*, vol. 2, 1183–1187. The Hague: Mouton. DOI: 10.1515/9783111349121-032.

Lightner, Theodore M. 1969. On the alternation *e~o* in Modern Russian. *Linguistics* 7(54). 44–69. DOI: 10.1515/ling.1969.7.54.44.

Lightner, Theodore M. 1972. *Problems in the theory of phonology, Vol. I: Russian phonology and Turkish phonology*. Edmonton: Linguistic Research, Inc.

Marklund Sharapova, Elisabeth. 2000. *Implicit and explicit norm in contemporary Russian verbal stress* (Studia Slavica Upsaliensia 40). Uppsala: Acta Universitatis Upsaliensis.

Matushansky, Ora. 2023a. Ablaut and transitive softening in the Russian verb. In Noah Elkins, Bruce Hayes, Jinyoung Jo & Jian-Leat Siah (eds.), *Supplemental Proceedings of the 2022 Annual Meeting on Phonology*, 1–12. Linguistic Society of America. DOI: 10.3765/amp.v10i0.5446.

Matushansky, Ora. 2023b. *Russian verbal stress clash as a tonal conflict*. Ms. SFL (CNRS/Université Paris-8). https://www.trees-and-lambdas.info/matushansky/Downloads/1sg%20pattern%20via%20tone%202.7.pdf.

Matushansky, Ora. to appear. Two BAP violations in Russian verbal stress. *Journal of Slavic Linguistics* FASL issue.

Meillet, Antoine. 1934. *Le slave commun*. Paris: Champion.

Melvold, Janis L. 1989. *Structure and stress in the phonology of Russian*. Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Micklesen, Lew R. 1973. The structure of the Russian verb stems. In Dean S. Worth (ed.), *The Slavic word*, 261–282. The Hague: Mouton.

Osadcha, Iryna. 2019. *Lexical stress in East Slavic: Variation in space and time*. Toronto: University of Toronto. (Doctoral dissertation).

Padgett, Jaye & Marzena Żygis. 2007. The evolution of sibilants in Polish and Russian. *Journal of Slavic linguistics* 15(2). 291–324. https://www.jstor.org/stable/24599670.

Pertsova, Katya. 2016. Transderivational relations and paradigm gaps in Russian verbs. *Glossa: a journal of general linguistics* 1. 13. 1–34. DOI: 10.5334/gjgl.59.

Red'kin, Valerij A. 1965. Ob akcentnyx sootnošenijax imeni i glagola v sovremennom russkom literaturnom jazyke [About accentual relationships of nomina and verbs in the Modern Russian language]. *Voprosy jazykoznanija* 14(3). 111–117.

Revithiadou, Anthi. 1999. *Headmost accent wins: Head dominance and ideal prosodic form in lexical accent systems* (LOT Dissertation Series 015). The Hague: Holland Academic Graphics.

Rubach, Jerzy & Geert Booij. 2001. Allomorphy in optimality theory: Polish iotation. *Language* 77(1). 26–60. https://www.jstor.org/stable/3087027.

Sims, Andrea D. 2006. *Inflectional defectiveness in a paradigmatic theory.* Columbus: The Ohio State University. (Doctoral dissertation).

Slioussar, Natalia. 2012. *Nekotorye svedenija o formoobrazovatel'nyx klassax russkix glagolov* [Some data on the inflectional classes of Russian verbs]. Ms. Utrecht Institute of Linguistics OTS/СПбГУ. http : / / www . slioussar . ru / resources/VerbDatabase/SlioussarVerbs.pdf.

Sussex, Roland & Paul Cumberley. 2006. *The Slavic languages.* Cambridge: Cambridge University Press.

Townsend, Charles E. & Laura A. Janda. 1996. *Common and comparative Slavic: Phonology and inflection.* Columbus, OH: Slavica.

Zaliznjak, Andrej A. 1977. *Grammatičeskij slovar' russkogo jazyka* [Grammatical dictionary of Russian language]. Moskva: Izdatel'stvo Russkij Jazyk.

Zaliznjak, Andrej A. 1985. *Ot praslavjanskoj akcentuacii k russkoj* [From Proto-Slavic accentuation to the Russian one]. Moskva: Nauka.

Zaliznjak, Andrej A. 2019. K istorii odnogo nefonetičeskogo izmenenija [On the history of one non-phonetic change]. In Anatolij F. Žuravlev (ed.), *Slavjanskoe i balkanskoe jazykoznanie: Rusistika. Slavistika. Komparativistika. Sbornik k 64-letiju S. L. Nikolaeva* [Slavic and Balkan Linguistics: Rusistics, slavistics, comparative studies. Festschrift for the 64th birthday of S.L. Nikolaev], 164–204. Moskva: Institute of Slavic Studies.

Chapter 13

Perfectivity in Russian, Czech and Colloquial Upper Sorbian

Olav Mueller-Reichau

Leipzig University

Addressing the topic of inner-Slavic variation in aspect, the present paper discusses the issues raised by the seemingly un-Slavic distribution of perfective and imperfective forms in Colloquial Upper Sorbian. An analysis is offered according to which Upper Sorbian perfectives have a weaker semantics than Czech perfectives, which in turn are weaker than Russian perfectives, with the imperfective in all three languages being radically underspecified. It is shown that this approach can successfully model the observed difference in aspect choice between the three languages.

1 Introduction

All Slavic languages have the grammatical category of verbal aspect, contrasting perfective forms with imperfective ones. Language textbooks usually claim that the perfective aspect expresses an action that is (or will be) completed, whereas the imperfective aspect expresses that the action is (or was) not yet completed. From a linguistic point of view, there are many problems with such simplified statements (cf. Comrie 1976: 18).

One problem for the analysis that considers perfectivity as completion and imperfectivity as incompletion is that it is simply false, as there are well-known occurrences of imperfectives with reference to completed events, as demonstrated in detail for Russian in Grønn (2004). Such cases have thus to be listed as exceptions. Another problem is that such a simple characterisation for each language, useful as it might be from a pedagogical perspective, suggests that the aspectual systems of different Slavic languages all work alike. This, however, is not the

case, as we know at least since the foundational work of Dickey (2000). Other authors addressing inner-Slavic aspectual differences include Alvestad (2013), Arregui et al. (2014), Breu (2000b), Klimek-Jankowska (2022), Petruchina (2000), Rivero & Arregui (2010), Wiemer (2008). As for studies seeking to explain deviating aspectual behaviours between two Slavic languages, most have addressed differences between Czech and Russian, e.g. Berger (2013), Berger (2016), Gehrke (2022), Heck (2018), Mueller-Reichau (2018), Stunová (1991, 1993).

In the present paper, I would like to draw attention to the aspectual system of Upper Sorbian, more specifically to Colloquial Upper Sorbian. Papers that include this language into the theoretical discussion on aspectual variation within the Slavic family are rare (but see Breu 2000b, Toops 1998, Wiemer 2008). A proposal for a formal analysis of aspect in Upper Sorbian has, to my knowledge, never been carried out. My motivation is also driven by the belief that a proper theoretical treatment of perfectivity and imperfectivity in Sorbian will shed new light on notions and categories that we as aspectologists believe to know well, and that we, therefore, do not think of questioning. Central in this respect is the notion of perfectivity, which is why I cannot avoid repeating the following often-quoted statement in this introduction (the author conducted field work on the Sorbian dialect of Mužakov): "Mne kažetsja, čto perfektivnost' v tom smysle, kak my ee sebe predstavljaem v russkom jazyke, vovse ne suščestvuet v mužakovskom" [It seems to me that perfectivity in the sense we understand it in Russian does not exist at all in the dialect of Mužakov; own translation] (Ščerba 1973: 121).[1] The present paper argues for a nuanced view of perfectivity in Slavic.

The paper is structured as follows. In Section 2 I will first present the relevant facts about Colloquial Upper Sorbian, with special emphasis on where it differs from Russian and Czech patterns. In Section 3 I will then introduce some theoretical background assumptions that I take for granted in the line of argumentation to follow. Section 4 discusses the proposal made in the literature according to which perfectivity in Colloquial Upper Sorbian should be analysed as terminativity. In view of the empirical shortcomings of this kind of approach, Section 5 moves on to assess the explanatory power of the second available proposal, which holds that perfectives in Colloquial Upper Sorbian encode determinateness. I will conclude that this analysis is on the right track, but it has to be made more precise to fully capture the data. To this end, I will offer a proposal in Section 6, showing that it correctly models the distribution of aspectual forms. Having so far laid out my proposal in prose, I will pin it down formally in Section 7. Section 8 serves to place my analysis into the broader context of

---

[1]Quoted after Scholze (2008: 230) and Werner (2018: 144).

scalarity-based approaches to Slavic aspect. Section 9 introduces two more data points that one might suppose do not fall under the theory developed, and I will explain why they in fact do. Finally, Section 10 concludes by integrating the findings based on Colloquial Upper Sorbian into the general picture of inner-Slavic aspectual variation.

2 What is special about aspect in Upper Sorbian?

Besides being proficient in German, speakers of Upper Sorbian live in a situation of diglossia (Breu 2000a, Lewaszkiewicz 2002; see Scholze 2008: 39ff. for discussion). On the one hand, there is the codified standard language, often referred to as the literary language (e.g. Stone 1993), which is spoken in formal contexts, for instance in school. The codification was oriented towards the Polish and Czech models (Werner 2003: 168). On the other hand, speakers employ the colloquial language in everyday speech. These two varieties differ quite strongly, not the least with respect to verbal aspect.

The colloquial language deviates from the literary language in many respects (Breu 2000a, Fasske 1981, Lewaszkiewicz 2002, Scholze 2008, Stone 1993). Besides aspect, this concerns, inter alia, the sporadic expression of personal masculine gender, the obligatory presence of articles in noun phrases, a particularly high number of lexical borrowings from German, a passive construction with the German loan verb *wordować* 'become', the use of adverbial preverbs as verbal prefixes like in *hrómadźestajeć* 'put together' (together+put) or *nutřčinć* 'put in' (inside+do).

With respect to aspect, standard Upper Sorbian has preserved the categories of Aorist and Imperfect. In addition to this "old" opposition, the literary language also displays the "new" opposition between perfective and imperfective verb forms (e.g. Fasske 1981). Colloquial Upper Sorbian (henceforth: CUS), by contrast, only possesses the distinction between perfective and imperfective aspect. In her extensive description of CUS, Scholze (2008) devotes a whole chapter to the category of verbal aspect, see also Scholze (2023). Breu (2000a) also discusses the functioning of perfective and imperfective forms in CUS in some detail. Unless indicated otherwise, the examples that I discuss in this paper are taken from these two sources.

What is *not* special about Upper Sorbian is the formal coding of the two aspectual categories. To this end, CUS exploits the same system of stem alternations as other Slavic languages do (Breu 2012: 248). As in, for instance, Russian (e.g. Švedova et al. 1980) or Czech (e.g. Karlík et al. 1995), unprefixed verbs are

mostly imperfective. When they are prefixed, they become perfective. When a secondary imperfective suffix is attached, the formerly perfective verb becomes imperfective and, additionally, we also find suppletive pairs.[2]

(1) shows examples from CUS (Scholze 2008: 230):[3]

(1) a. *fönwać* (IPFV) – *sfönwać* (PFV) 'blow-dry'
     b. *wotućić* (PFV) – *wotućwać* (IPFV) 'wake up'
     c. *kipwać* (IPFV) – *kipnć* (PFV) 'tip'

What sets CUS aside from the other Slavic languages is the way perfective and imperfective verb forms are used. In many respects, CUS resembles Czech (Breu 2000a: 45). Consider the following examples:

(2) *Wón je   husto jenož jednu knihu předał.*
     he AUX often only one   book sell.PST.PFV
     'He often sold only one book.'   [CUS]

(3) *Často prodal   jen jednu knihu.*
     often sell.PST.PFV only one   book
     'He often sold only one book.'   [CZ]

As with speakers of Czech, speakers of CUS will choose a perfective form[4] to convey the message that "he often sold only one book", whereas speakers of Russian would select the imperfective form in this context:

(4) *On často {prodaval   / *prodal}   tol'ko odnu knigu.*
     he often sell.PST.IPFV   sell.PST.PFV only   one book
     'He often sold only one book.'   [RU]

In other respects, however, CUS is known to be the odd one out in showing "unslavic" use of aspectual forms. The most striking fact (from the point of view

---

[2]This is, of course, a very simplified summary. Since questions of morphology lie outside of the goals of the present paper, however, it should suffice for our purposes here. See Werner (2003) for a much more detailed picture of Upper Sorbian verbal affixation.

[3]Not surprisingly, Standard Upper Sorbian aspectual morphology works in the same way, see Scholze (2008: 230).

[4]Throughout this text I will speak of "perfective forms" in CUS as if the language had a grammatical category signalling perfectivity understood sensu stricto, i.e. as completed event denotation. As will become clearer soon, however, it has not. My decision might cause confusion, but I think it would be even more confusing if I chose a different label for those forms that have perfective morphology from the point of view of other Slavic languages.

of what one would expect from a "well-behaved" Slavic language; Breu 2000a: 54) is perhaps the possibility of referring to an ongoing event by means of a perfective verb:

(5) *Jurij jo   rune jen text šełožił,   hdyž sym ja nutř šišoł.*
    J.   AUX now a   text translate.PST.PFV when AUX I in   come.PST.PFV
    'When I came in, Jurij was translating a text.'

That sentence (5) can be interpreted so that the translation event temporally includes the moment when the speaker enters, stands in sharp contrast to what its Czech or Russian counterparts would allow for, cf. (6) and (7). Pay attention to the translations enforced by the use of the perfective forms *přeložil* and *perevel*, respectively.

(6) *Když jsem přišel,   přeložil   jeden text.*
    when AUX come.PST.PFV translate.PST.PFV one   text
    'When I came, he had translated a text.'   [CZ]

(7) *Kogda ja prišel,   on perevel   tekst.*
    when I come.PST.PFV he translate.PST.PFV text
    'When I came, he had translated a text.'   [RU]

The phenomenon is well-known in the literature dedicated to verbal aspect in Upper Sorbian. Apart from Breu (2000a) and Scholze (2008), documentation and discussion can also be found in Werner (2003, 2013). The following is quoted from Werner (2003: 43).

(8) *Sym runje při tym, krótke powědančko přełožić.*
    AUX now at that short tale   translate.INF.PFV
    'I am at the moment engaged in translating a short story.'

Within the present paper, the possibility of using perfectives to refer to ongoing events will be the main focus of interest, but let me add two other contexts in which CUS shows remarkable (in the sense of "unslavic") aspect use. Both are well discussed in the relevant literature.

First, CUS has a compound future with a perfective infinitive (Stone 1993: 637), as illustrated in (9) from Scholze (2008). Secondly, there is no ban on perfective infinitives after phase verbs, as can be seen in (10) from Breu (2012).

(9) *A   potom budu jej pokazać!*
    and then   AUX her show.INF.PFV
    'And then I'll show her!'

(10)   *Tón jo   započał   jowo zaćišće   napisać.*
      he   AUX start.PST.PFV his   impressions write.INF.PFV
      'He began to write down his impressions.'

3 Some theoretical background assumptions

In what follows, I will revisit the data on aspect in CUS and discuss the analyses that have been presented so far to account for these data. Jumping ahead, I will argue for a theoretical account that follows the intuitive explanation suggested in Toops (2001). My proposal will build on three background assumptions. First, I consider the category of aspect to relate reference times and event times (Klein 1994). Second, following Grønn (2004), I take the imperfective operator IPF to introduce a radically underspecified aspectual relation; all that an imperfective form requires is that the reference time overlaps with the event time, whatever the overlap relation will ultimately look like. Third, I assume that different Slavic languages may vary with respect to the content of the perfective category, i.e. with respect to the precise truth-conditional impact of perfectivity on interpretation (Mueller-Reichau 2018).

Informally, perfective aspect in Russian and other East Slavic languages has been characterised as expressing "connectedness", which roughly means that the event encoded by a perfective verb will always have to be understood as grounded ("connected") within the chain of particular events preceding and following it (Barentsen 1995, 1998, Dickey 2000, Dickey 2015, Stunová 1991, 1993). Grønn (2004) offers a way to model the intuition behind connectedness in truth-conditional terms. He proposes that, in the case of Russian perfectives, the reference time has to end when the target state of the event is in force.

Let me elaborate on that. For a state to be in force at a given moment t, it has to be valid immediately before t and immediately after t. Now let t be the final moment of the reference time interval introduced by a Russian perfective, as proposed in Grønn (2004). It follows that there has to be a state *after* the reference time interval connected to the state inside the reference time. The kind of connection between the two is trivial: they are parts of one and the same state. In other words, a Russian perfective will almost always[5] denote a state-changing event, at the same time entailing the existence of an eventuality subsequent to the

---

[5]"Almost always" because there is a systematic exception. Perfectives formed by the delimitative prefix *po-* denote no change-of-state (Filip 2000, Dickey 2006). In these cases, where the prefix carves out a chunk of the process delivered by the verbal base, connectedness is trivially given because the reference time cannot end but when the process is in force.

denoted change of state. Typically, the subsequent eventuality is a state, which in turn may provide the occasion for further events in the course of the world talked about. The event introduced by a perfective may, however, also be "connected" to a process, as with ingressive verbs like *pojti* 'start going', for instance. In any event, a Russian perfective will always have to be interpreted as embedded within a chain of eventualities.

Compared to the situation in Russian, the perfective category in Czech has a weaker content. What is required by a perfective here is that the event is referred to in its totality (Stunová 1991, 1993). Formally, this can be modelled by making use of Filip's maximality operator (Filip 2008, 2017). In Czech, then, perfectives impose the condition that the reference time has to include the endpoint of the event ("maximality requirement"), but not that the time of the eventuality brought about by it was partially covered by the reference time.

Now how about Sorbian, more precisely, Colloquial Upper Sorbian (CUS)? In (5), we already saw that perfectives in CUS do not require the reference time to include the endpoint of the event. Thus, perfectivity in CUS seems to be an even weaker semantic notion compared to perfectivity in Czech, which, as we saw, is semantically weaker than perfectivity in Russian.

4 Terminativity

Breu (2000a, 2012) and Scholze (2008) provide an insightful discussion of how speakers of CUS distribute aspectual forms over contexts. When interpreting the presented facts, the authors arrive at the conclusion that a perfective form will be used in CUS if the sentence predicate specifies an "inherent goal", independently of whether the goal will be reached. Here is Scholze's (2008) definition (own translation from German):[6]

> Terminativity, i.e. the existence of a goal of the action, no matter whether it is reached or not, is verbalised in CUS by the perfective aspect. In contrast, the imperfective aspect expresses aterminativity, i.e. actions without an inherent goal. (Scholze 2008: 232-233)[7]

---

[6]A more precise definition is not offered. The term "terminative" seems to be used interchangeably with the terms "telic" and "bounded" (cf. Breu 2012: 264).

[7]"Terminativität, d.h. das Vorhandensein eines Ziels der Handlung, gleichgültig ob dieses erreicht wird oder nicht, wird in der SWR [= CUS] durch den perfektiven Aspekt versprachlicht. Im Gegensatz hierzu drückt der ipf. Aspekt Aterminativität aus, also Handlungen ohne inhärentes Ziel" (Scholze 2008: 232-233).

The proposal appears to be simple and straightforward. It predicts the non-use of perfective aspect when reference is made to single events on the basis of atelic predicates, i.e. of predicates not specifying an inherent goal. The following would be a case in point:

(11) *Hdyž hromadźé za blidom sedźeštaj,* *hladaštaj* *sej do*
when together at table sit.3DU.PST.IPFV look.3DU.PST.IPFV REFL to
wočow.
eyes
'As the two sat together at the table, they looked each other in the eye.'

At the same time, the possibility of perfectives expressing ongoingness, which was illustrated in (5), is directly accounted for. (12) shows a further example of that kind: the predicate is telic/terminative, so verbal aspect is perfective. That the event has not yet finished at the end of the reference time is no obstacle for using the perfective.

(12) *Wón rune jenu kniw šeda.*
he now a book sell.PRS.PFV
'He is now selling a book.'

Scholze (2008: 240) reports that the imperfective form *fönwe* 'blow-dry' in (13) may be replaced by the prefixed perfective *sfönwe*, as shown in (14), but that this replacement induces a meaning shift. Her informants state that, unlike (13), which describes the pure activity of blow-drying, (14) is about an event which is directed toward the goal of having dry hair. One easy way of accounting for this intuition is to associate perfective aspect with the existence of an inherent goal of the action.

(13) *Na tón so wěšći fönwe* *něke.*
well he REFL surely blow-dry.PRS.IPFV now
'He surely is blow-drying his hair right now.'

(14) *Na tón so wěšći sfönwe* *něke.*
well he REFL surely blow-dry.PRS.PFV now
'He surely is blow-drying his hair right now (until it will be dry).'

Successful as it seems at first sight, the proposal that perfectivity in CUS should be analysed as terminativity runs into problems. Look at the following example. Note that the predicate clearly specifies a goal. The goal will be reached when a series of events of the same kind (folding a piece of cloth) will have been performed over a limited set of objects (pieces of cloth):

(15) *Ja kładu   rune tón wešu   hromadze.*
     I   put.PRS.IPFV now the laundry together
     'I'm just folding the laundry.'

In this case, although the predicate describes an event which has an inherent goal (that the laundry is folded), the imperfective is used, and not the perfective, as one might have expected.

Example (16) shows a similar case, which is discussed in Breu (2000a).

(16) *Ja šedawam   rune peć knijow.*
     I   sell.PRS.IPFV now five books
     'I'm selling five books now.'

Here, too, the imperfective form is used despite the fact that the reported event has an "inherent goal" (that five books are sold). Noticing that examples like these run counter to his generalisation, Breu (2000a: 64) writes that "[a] certain complication arises from that the ipf. seems to have another function apart from the expression of aterminativity, namely the expression of distributivity".

This "complication" is not the only problem that arises from identifying perfectivity with terminativity in CUS. Another one is that the proposal makes the wrong prediction for perfective generics. Consider the following, where the predicate is aterminative, but the perfective is used nevertheless.

(17) *Tón basne chětř   nawukne.*
     he poems quickly learn.PRS.PFV
     'He quickly learns poems.'

Learning a poem is an activity which aims at a specific goal. The predicate in (17) is about learning poems, however, and *this* activity has no specific goal, or telos. The addition of "quickly" does not change that: learning poems quickly is still atelic. Despite that, the verb form used in (17) is not imperfective, but perfective, which needs to be explained.

Interestingly, we also find the mirror image to (17), i.e. sentences where the predicate is telic/terminative, but aspect is imperfective. Consider the predicate in (18). It describes an event that has an inherent goal (that the poem is learned). If Breu's and Scholze's proposal was correct, we would expect a perfective verb form. The verb form that appears in (18) is, however, imperfective.

(18) *Tón wukne   rune tón basejn.*
     he learn.PRS.IPFV now the poem
     'He is just learning the poem.'

To conclude so far, analysing perfective verb forms in CUS as expressing terminativity can account for many, but not all data. Let us therefore move on to see in how far the proposal made by Toops (2001) fares better than Breu's and Scholze's.

5 Determinateness

According to Toops (2001), perfectives in CUS encode the notion of determinateness, by which he means a one-time action heading for a goal: "Determinate forms denote a goal-oriented action occuring either once or irregularly" (Toops 2001: 132). Toops' perfectivity condition is more specific than the one of Breu (2000a) and Scholze (2008). Roughly speaking, determinateness is terminativity ("goal-oriented") plus uniqueness ("once"). As we will see now, Toops (2001) can indeed explain aspect choice in CUS for almost all contexts of use.

The first thing to note is that Toops (2001) can account for those data points that Breu and Scholze can also account for. Recall example (5). The denoted event is not completed, but it is goal-oriented and single, and this alone suffices for licensing the perfective.

Moving on to (19), the imperfective version of (12), where the non-use of the perfective implies that either the condition of goal-orientation, or the condition of singularity is not fulfilled (or that both are not). This is indeed the case because the direct object *jenu kniw* 'a book' is to be interpreted as type-referring (Rachilina 2000: 69). The sentence describes several events of selling a copy of the same (kind of) book, leaving open the precise number of events. Due to the absence of a goal to which the repetition of book-sellings would be directed, the example is in line not only with Breu (2000a, 2012) and Scholze (2008), but also with Toops (2001). Both proposals correctly predict the use of an imperfective verb form.

(19)   *Wón rune jenu kniw šedawa.*
       he   now a   book sell.PRS.IPFV
       'He is now selling a book (copy by copy).'

Next, consider (20). This example is in harmony with interpreting perfectivity in CUS as terminativity, as do Breu and Scholze, because the predicate is aterminаtive (it is generic). It is also in line with linking the use of a perfective to the two properties of being terminative ("goal-oriented") and being singular ("one-time"), as in Toops's (2001) account.

(20)   *Wón šedawa   knije.*
    he   sell.PRS.IPFV books
    'He sells books (= is a bookseller).'

As a very nice minimal pair, Breu (2000a: 59) provides (21) and (22):

(21)   *Dyš   sem ja do lodna šišła, jo   sej   wona rune jabuka brała.*
    when AUX I to shop come AUX REFL she   now apples take.PST.IPFV
    'When I came to the shop, she was taking apples.'

(22)   *Dyš   sem ja do lodna šišła, jo   sej   wona rune jabuka zała.*
    when AUX I to shop come AUX REFL she   now apples take.PST.PFV
    'When I came to the shop, she was taking a pack of apples.'

Both examples are identical, except for the fact that the first one ends in an imperfective, whereas the second one ends in a perfective verb. This difference results in the following difference in interpretation. (21) is about the taking of an unspecific set of apples. By contrast, (22), the perfective version, is understood such that what is being taken is a definite set of apples, most likely a package of apples.[8]

Breu (2000a) argues that this contrast provides support for his assumption that CUS perfective forms express terminativity, because packaging the apples into a single unit furnishes the predicate with an inherent goal (that the apples are taken). Toops (2001) could build on this to argue that interpreting the apples as a single unit not only leads to terminativity, but also to understanding the apple-taking as a single event.

So far, we have looked at data that can be accounted for by appealing to terminativity as well as by appealing to determinateness for perfectivity in CUS. If that was all, the former kind of approach would be preferable because, as we saw, terminativity is a weaker semantic notion than determinateness. However, there are examples that the terminativity approach cannot explain, but the determinateness approach can. We already came across such examples above in (15) and (16). Let me repeat one of them for ease of exposition:

(23)   *Ja kładu   rune tón wešu   hromadze.* (=(15))
    I   put.PRS.IPFV now the laundry together
    'I'm just folding the laundry.'

---

[8]Note that both main predicates are translated into English by means of the progressive ('was taking').

Cases like these represent the "complications" that force Breu (2000a) to assume that imperfective forms not only express aterminativity, but also distributivity. Now we can convince ourselves that, if we follow Toops (2001), the complication disappears. Since distributivity is a manifestation of pluractionality (Müller & Sanchez-Mendes 2020), and since perfectives are linked to singularity according to the proposal made in Toops (2001), the non-use of the perfective is correctly predicted.

Above I have argued with respect to (19) that the use of the perfective is excluded because of the missing goal of the predicate. Now we can add that also Toops' second condition for the use of a perfective, singularity, is not met in (19), due to the pluractionality of the predicate.

There are, nevertheless, two examples figuring in the discussion about aspect use in CUS that Toops (2001), as it seems, cannot handle. The first issue arises in connection to the following generic sentence:

(24)   *Tón basne chětř   nawukne.*   (=(17))
        he   poems quickly learn.PRS.PFV
        'He learns poems quickly.'

The problem here is that by uttering (24) the speaker does not refer to a unique event (note the plural form of the direct object). According to the logics of Toops (2001), this excludes the use of a perfective verb form, yet the verb *is* perfective.

A second apparent counterexample is (25):

(25)   *Tón wukne   rune tón basejn.* (=(18))
        he   learn.PRS.IPFV now the poem
        'He is learning the poem.'

This is not a generic, but an episodic sentence which is used to refer to an ongoing event. Since the event is a singleton and oriented towards a goal (that the poem is learnt), the choice of a perfective is licensed on the account of Toops (2001), but the verb actually used is *not* perfective.

In what follows I want to propose a modification of Toops (2001), or rather a specification of that theory. My aim is to integrate the counterexamples presented above into an overall approach in keeping with the spirit of the proposal, and to show that this will give us an account that captures the data correctly.

6 Proposal

Let us define determinateness by exploiting the notion of a path (Gehrke 2008, Krifka 1998, Zwarts 2005).

(26) An event predicate *P* is DETERMINATE iff it is unidimensional, directed and bounded, whereby:

    a. *P* is UNIDIMENSIONAL iff the events in its denotation set share a path structure such that all elements of this path structure are parts of a common path which likewise belongs to the part structure.
    ("all paths are parts of a common path")

    b. *P* is DIRECTED iff the events in its denotation set share a path structure such that there are no two non-overlapping elements in this path structure that occupy the same space.
    ("no return to a formerly traversed region")

    c. *P* is BOUNDED iff the events in its denotation set share a path structure which includes an element that cannot be concatenated by another element of the same path structure such that the resulting path likewise belongs to this path structure.
    ("there is a maximal path")

With this definition in mind, we will now reconsider the examples presented above. My claim is that perfective verbs are used in CUS in those cases where the predicate is determinate in the sense of (26). For ease of exposition I will not refer back to the initial presentation of an example, but instead repeat it with new numbering.

(27)   *Wón rune jenu kniw šedawa.*   (=(19))
      he   now a   book sell.PRS.IPFV
      'He is now selling a book (copy by copy).'

The non-use of the perfective in (27) can be explained by the fact that the predicate does not meet the criteria for being determinate. It does not because the paths belonging to different individual selling events do not have a common path. So (27) violates (26a), the condition of unidimensionality. What has been argued here with respect to (27) generalises to all cases of event repetition, i.e. to all predicates that describe a plurality of events of the same kind.[9]

In contrast to (27), (28) is about a single selling-event. The noun phrase *jenu kniw* 'a book' is understood as referring to a particular book as the article being sold.

---

[9]A reviewer wonders whether this "generalisation" also covers distributive readings with plural subjects performing the same kind of action at the same time along different paths. As I understand it, the question boils down to the issue of whether or not the subject expression belongs to the event predicate. I will have to leave this difficult though important topic for future research.

(28) *Wón rune jenu kniw šeda.*   (=(12))
      he   now a   book sell.PRS.PFV
      'He is now selling a book.'

Since reference is made to a single event, there is no violation of unidimensionality. Moreover, there is no reason to assume that an event of selling a particular book would traverse along a path that runs through the same space twice. Thus, there is no violation of directedness. Finally, such an event has an inherent goal as it develops along a maximal path. A path is maximal, or upper-bounded, if it cannot be concatenated by a subpath of itself without the product of this concatenation falling outside of the path structure of the predicate. Accordingly, there is also no violation of boundedness. We may therefore attest that the predicate is determinate. This predicts the use of a perfective form, and this is what we observe.

So far, so good. Let us now move on to those examples that have turned out to be problematic for Toops (2001). One of them was (25), repeated here:

(29) *Tón wukne   rune tón basejn.* (=(25))
      he   learn.PRS.IPFV now the poem
      'He is learning the poem.'

The predicate in (29) describes a one-time, goal-oriented event. According to Toops (2001), this should license the use of the perfective form. The form actually used is imperfective, however.

The extension, or specification, of the approach of Toops (2001) that I have proposed above solves this problem. Recall that under my definition of determinateness, the described event has to be not only unidimensional ($\approx$ "one-time") and bounded ($\approx$ "goal-oriented"), but also directed. Directedness is thereby defined such that there must not exist two or more pieces (subpaths) of the overall event path that would occupy the same region. Very loosely speaking: at no time during the course of the event should the event return to where it was before.[10]

With respect to (29), two facts need to be noted. First, the movement along a path associated with the learning of a poem is not motion in the physical sense, but rather a metaphorical motion along the words of the poem as they are considered by the learner. Second, the learning of a poem does not usually proceed linearly from the first word to the last word, but rather in cycles. If I

---

[10]I believe that what Toops (2001) had in mind when speaking about "goal-oriented action" was actually an amalgam of directedness and boundedness. This is why I consider my approach to be an elaboration rather than a correction of Toops (2001).

want to memorise a poem, I may start by reading through the whole poem first, then I will perhaps return to the first verse, will read it again through, will maybe return to the first line, will reread it, and so on. As should be obvious, this kind of event structure violates directedness, and therefore the use of the perfective is *not* to be expected, thus explaining the use of the imperfective in (29).

What remains to be discussed are the generic sentences that seem to run counter to the idea that perfectives express determinateness, at least if one takes determinate events to be "one-time goal-oriented", as proposed in Toops (2001). The knowledge about generics that has been accumulated is rich (Cohen 2022; Krifka et al. 1995; Leslie & Lerner 2022). Here I can obviously only scratch the surface. Generally speaking, two kinds of generics have been identified in the literature. The first one goes under the expression "descriptive generics" (e.g. Krifka 2013), or "inductivist generics" (e.g. Carlson 1995). The labels are motivated by the fact that the generalisation expressed by such a generic *describes* the way certain individuals behave in the world, or that the behaviour of the individuals allows for *inducing* the generalisation. Sentence (30) is a case in point.

(30) *Wón šedawa   knije.* (=(20))
      he   sell.PRS.IPFV books
      'He sells books (= is a bookseller).'

Whatever the ultimate analysis of these generics, it is widely agreed that their meaning involves quantification over events, and that the generalisation expressed by them results from quantifying over "sufficiently many" (Cohen 2022) such events. Disagreement concerns the question as to what should count as "sufficiently many" and how to model this factor (see Krifka et al. 1995 and Cohen 2022 for surveys of different types of inductivist approaches).

In light thereof, the generalisation conveyed by uttering (30) is based on the observation that 'he' has acted as seller in sufficiently many events of selling a book. Since the meaning of the predicate *šedawa knije* 'sells books' accordingly entails more than one book-selling event, we may note that unidimensionality is violated in (30). This leads us to expect the imperfective form, in line with the facts.

(31) *Tón basne chĕtř   nawukne.*   (=(24))
      he   poems quickly learn.PRS.PFV
      'He learns poems quickly.'

(31) shows an instance of the second kind of generic sentences. In cases like these, which I refer to as "dispositional generics" in Mueller-Reichau (2020), the

predicate describes an event which the subject referent is expected to be able to perform given the need to do so. Dispositional generics are logically linked to "definitional generics" (Krifka 2013; Seres & Espinal 2019) or "in virtue-of generics" (Greenberg 2003), since recently also known as "normative generics" (Hesni 2022, Leslie & Lerner 2022). The following Polish examples, which are taken from Klimek-Jankowska (2008), may serve as illustration. (32) is a definitional generic.

(32) *Przyjaciel pomoże   w potrzebie.*
      friend   help.PRS.PFV in need
      'A friend will help in need.'   [PO]

For the purposes of the present paper, we may safely ignore details of ongoing discussion about the best analysis of definitional generics. Let us simply note that a definitional generic will be uttered to express that members of the kind named by the subject nominal have the property described by the predicate because they have a principled connection to the kind named by the subject nominal (Prasada & Dillingham 2006). Thus, (32) expresses that one who qualifies as member of the kind/category 'friend' has the property of helping you in case that help is needed *because* he is a friend.

   Related to the definitional generic (32) is the dispositional generic (33), which says that Janek has the property of helping in case that help is needed. Given the truth of (32), (33) silently conveys the additional message that Janek is one who deserves being counted as a friend (Klimek-Jankowska 2008, 2012, Mueller-Reichau 2020).

(33) *Janek pomoże   w potrzebie.*
      Janek help.PRS.PFV in need
      'Janek will help in need.'   [PO]

Returning to Sorbian, the generic sentence (31) is a dispositional generic, like (33). Just like Janek is said to help in case help is needed, 'he' in (31) is said to learn a poem quickly in case a poem needs to be learned. And similar to the way (33) communicates the additional information that Janek deserves being called a friend, (31) silently coarticulates that 'he' qualifies for being called a remarkably smart person.

   Importantly, since the predicate in (31) is about a single learning of a poem relative to the modal context providing the task of learning a poem, just a single event is in question, and the condition of unidimensionality is satisfied. Now, what about directedness and boundedness? Above I said that the path in *learning*

a poem goes in cycles, and that this violates directedness, implying the exclusion of perfective aspect. If so, why is directedness not violated in (31)?

The reason is, arguably, the impact of the adverb *chětř* 'quickly', which triggers an abstraction away from the "real" profile of the path to the temporal distance between the initial point of the path $p(0)$ and the final point of the path $p(1)$. It seems plausible to analyse *chětř* as the focus-bearing constituent in (31). Given this, the sentence will be understood as an answer to the implicit question about the speed at which 'he' learns poems.

Speed is distance divided by time ($v = s/t$). The adverb *chětř* 'quickly' denotes a property of how much time it takes to memorise a poem, i.e. to move along a path from not knowing any word ($p(0)$) to knowing the whole text ($p(1)$). The less time is used to proceed on this path, the faster the learning of the poem; the faster the learning of the poem, the shorter the path of the event of learning the poem. Figure 1 shows different speeds compared with each other.
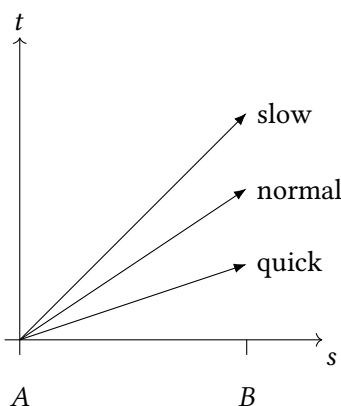
Figure 1: Speeds of learning a poem

Under the assumption that the adverbial *chětř* is the focus constituent, the following propositions constitute the (simplified) set of alternatives relevant for interpretation:

(34)　Focus alternatives to (31):

　　　a.　that he learns poems very slowly

　　　b.　that he learns poems slowly

　　　c.　that he learns poems at normal speed

　　　d.　that he learns poems quickly

　　　e.　that he learns poems very quickly

One proposition out of these, namely (34d), will be asserted as true if (31) is uttered. The set of alternatives in (34) implies a comparison of different speeds. This comparison presupposes a normalised distance $\overline{AB}$.

Crucial for the present discussion is the fact that the paths of the events compared with each other in the informational background of (31) all support directedness, as can be easily read from Figure 1. This should then also hold for the path of the event denoted by (31): there are no non-overlapping subpaths of this event path that would occupy the same region. Moreover, the paths depicted in Figure 1 also support boundedness, as there is an upper-bound determined by *B*. Since the predicate of (31) satisfies boundedness, directedness, and unidimensionality, the use of perfective aspect is called for.

7 Formalisation

The starting point of this paper was the hypothesis, argued for in the literature referred to above, that it is two different kinds of perfectivity that figure in the aspectual systems of Czech and Russian. While the Czech perfective category encodes maximality (reminiscent of the concept of "totality" in the traditional literature), the Russian perfective encodes target state validity (loosely related to the traditional notion of "resultativity"). Having investigated the use of perfective and imperfective verb forms in CUS, I conclude that this language introduces a third kind of perfectivity into the overall picture of aspect in Slavic.

CUS features the weakest perfective category within the Slavic family, so weak indeed that the question arises as to whether it should be called "perfective" at all. The constraint that the use of a CUS perfective form imposes on interpretation is merely that the event property has to be unidimensional, directed and bounded. These properties, summarised under the label "determinateness", are defined in terms of the path that the denoted event is described as traversing.

Czech perfectives come with a more specific requirement. Here, the denoted event is described as a maximal event. The notion of maximality may be defined mereologically in terms of event stages, but also temporally by requiring the reference time to include the final moment of the event. The final moment is the moment at which the upper bound of the event path is reached.

Russian has the most specific perfective category. By using a Russian perfective, the speaker refers not only to an event that has been fully realised (up to the upper bound of its path), but in addition to an eventuality (often a state) that the realisation of the event has brought about. The pragmatic effect is that the particular conditions of said event's successor are relevant for the further discourse.

In the remainder of this paper, I want to integrate these observations and claims into the overall picture. To do so, I take for granted that in every Slavic language, the imperfective category is semantically underspecified ("unmarked"), whereas the perfective category comes with specific content ("marked"). To be concrete, I take the following to be the imperfective operator appearing in every Slavic language:

(35) **Imperfective operator**:
$IPFV \Rightarrow \lambda P \lambda t \exists e. P(e) \wedge t \circ \tau(e)$

As should have become clear, the specific content contributed by the perfective category may differ from language to language. (36) shows the perfective operator of CUS:

(36) **Perfective operator in CUS**:
$PFV_{CUS} \Rightarrow \lambda P \lambda t \exists e. P(e) \wedge \text{DET}(P) \wedge t \circ \tau(e)$

In contrast to that, the perfective operator of Czech may formally be pinned down as in (37).

(37) **Perfective operator in Czech**:
$PFV_{CZ} \Rightarrow \lambda P \lambda t \exists e. P(e) \wedge \text{DET}(P) \wedge t \circ \tau(e) \wedge f_{end}(\tau(e)) \subseteq t$

As for the Russian perfective operator, I propose the following:

(38) **Perfective operator in Russian**:
$PFV_{RU} \Rightarrow \lambda P \lambda t \exists e. P(e) \wedge \text{DET}(P) \wedge t \circ \tau(e) \wedge f_{end}(\tau(e)) \subseteq t \wedge f_{end}(t) \subseteq f_{target}(e)$

Of course, for these formulae to be understandable, I have to state precisely what the property DET is supposed to mean. The following summarises the informal discussion presented above:

(39) **Determinateness**:
$\forall P. \text{DET}(P) \leftrightarrow \text{UNI}(P) \wedge \text{DIR}(P) \wedge \text{BND}(P)$

According to (39), a property will fulfill DET if it fulfills UNI, DIR, and BND. So we move on to state the semantics of the latter three predicates:

(40) **Unidimensionality**:
$\forall P. \text{UNI}(P) \leftrightarrow \forall e \forall e' \forall q. P(e) \wedge e' \leq e \wedge q = \text{TRACE}(e') \rightarrow \exists p. p = \text{TRACE}(e) \wedge q \leq p$

This basically says, to repeat from above, that a predicate is unidimensional iff the events in its denotation set each have a path structure such that all paths within it are parts of a common path within it.

(41) **Directedness**:
$\forall P.\text{DIR}(P) \leftrightarrow \forall e \forall e' \forall e'' \forall p \forall q.P(e) \wedge e' \leq e \wedge e'' \leq e \wedge \text{TRACE}(e') = p \wedge \text{TRACE}(e'') = q \wedge \neg(p \circ q) \rightarrow \text{SPACE}(p) \neq \text{SPACE}(q)$

(41) expresses that a predicate is directed iff the events in its denotation set each have a path structure such that there are no two non-overlapping paths within it that occupy the same space.

(42) **Boundedness**:
$\forall P.\text{BND}(P) \leftrightarrow \forall q \forall e.P(e) \wedge q \leq \text{TRACE}(e) \rightarrow \exists p.p \leq \text{TRACE}(e) \wedge \neg(p + q \leq \text{TRACE}(e))$

According to (42), a predicate is bounded iff the events in its denotation set have a path structure which includes a path that cannot be concatenated by another path within it such that the resulting path would belong to the same path structure.

8 Paths and scales

A reviewer asked me whether I understand paths as scales, and how my proposal relates to Kagan's (2015) Scale hypothesis. I am grateful for these questions as they give me the opportunity to place my approach in a broader context. The answer to the first question is yes, I do. As for the second question, I will answer it carefully in the following short section.

Kagan (2015) is concerned with the role of prefixation in the grammar of Russian. According to her, the semantic contribution of a prefix is such that it effectively fixes a point on a scale. In the words of the author, the prefix "imposes a relation between two degrees on a scale, one of which is a degree associated with the event denoted by the verbal predicate, and the other, the standard of comparison" (Kagan 2015: 24). The second degree, the one that serves as the standard of comparison, is contributed by external sources, that is to say, by means other than the prefix: "the standard of comparison can be contributed either by a linguistic expression that appears in the sentence, or by the context" (Kagan 2015: 24). The scale itself is supplied by the verbal predicate to which the prefix attaches, possibly including the direct object (Kagan 2015: 25). More precisely, it is

contributed by a gradable property "associated" (Kagan 2015: 26) with the verbal predicate. One and the same predicate may be "associated" with different gradable properties, so that different prefixes applying to the same base may operate on different scales (provided by different gradable properties). Since the prefix operates on an independently given scale with an independently given degree that serves as standard of comparison, we may note that, in its core, Kagan's (2015) approach boils down to the claim that prefixes, by relating the comparison degree to some specific degree on the scale, produce upper-bounded scales.

The conclusion that prefixation amounts to creating upper-bounded scales has been arrived at by others as well (e.g. Filip 2008, Gehrke 2008). Upper-bounded scales are what I call bounded paths in the present paper. I assume that Slavic-style aspect works with the two operators PFV and IPFV.[11] These operators apply to verbal predicates. While IPFV is semantically underspecified in its input conditions, PFV calls for predicates that involve (upper-)bounded scales which are at the same time unidimensional and directed. Whenever at least one of these three conditions is not met, PFV will give way to IPFV. And here is now the link to Filip (2008), Gehrke (2008), and Kagan (2015). Following these authors, I hold the view that prefixation produces event descriptions with upper-bounded scales. Therefore, prefixed verbal predicates are well-prepared for serving as the input to PFV. Well-prepared, but not fully prepared, because boundedness is no sufficient condition for perfectivity alone. In addition to boundedness, perfectivity also entails directedness and unidimensionality. Whether or not unidimensionality and directedness are met in addition to boundedness can be read from the presence or absence of secondary imperfective morphology. Secondary imperfective markers on the predicate signal that either unidimensionality or directedness are not met.

Pay attention to a certain flexibility which is built into Kagan's (2015) theory. A verbal predicate will be "associated" with one gradable property (and hence scale) or the other depending on the element in the context of which it appears, i.e. the prefix. In the next section, I will generalise this to other contextual elements in order to explain cases that otherwise would run counter to the predictions.

9 Two intricate cases

Above I have argued for three notions of perfectivity, namely perfectivity as determinateness (the case of CUS), perfectivity as maximality (the case of Czech),

---

[11]See Mueller-Reichau (forthcoming) for discussion of attempts to reduce the two covert operators to one.

and perfectivity as connectedness (the case of Russian). I suggested the following entailments to hold: (i) maximality implies determinateness, and (ii) target state validity (connectedness) implies maximality. While the latter seems uncontroversial, the former is surely not. Whether maximality really fully implies (that is: entails) determinateness is questionable.

Scholze (2008: 244) presents the following dialogue:

(43) A: *Što wó jow činiće?*
     what you.DU here do.PRS.IPFV.DU
     'What are you two doing here?'
   B: *Ja nawuknem rune ka so rajfn wekslwe.*
     I learn.PRS.PFV now how REFL tyre change.PRS.IPFV
     'I am learning how to change a tyre.'

This is one more case where a perfective verb form, *nawuknem* 'am learning', is used to refer to an ongoing event (recall (5) from above). What is remarkable about (43) in the context of the present discussion is a comparison to (29), repeated here one more time.

(44) *Tón wukne rune tón basejn.* (=(29))
   he learn.PRS.IPFV now the poem
   'He is learning the poem.'

With respect to (44), Scholze (2008: 245) notes that the replacement of imperfective *wukne* with perfective *nawukne* is excluded ("ausgeschlossen"). Why, then, is it not excluded in (43)? Scholze writes that the use of perfective *nawukne* to refer to an ongoing learning event will be possible only if the learner has the intention to acquire a certain skill.[12] But this is hardly convincing, because one would not want to think that the learner in (44) does not intend to acquire the skill to recite the poem. What, then, is the critical difference between the contexts (43) and (44)?

Although both learning events in (43) and (44) are about the acquisition of knowledge, they differ in that the acquisition of knowledge about how to change a tyre follows a standardised and thus pre-given plan (it follows a script in the sense of Schank & Abelson 1977), while the acquisition of knowledge about how to recite the poem does not. I propose that (43) presupposes a plan along the lines of which the learning proceeds, and that this plan supplies a directed path.

---

[12]"Terminative Prozessualität setzt bei dem Lexem *wuknć – nawuknć* offensichtlich den beabsichtigten Erwerb einer Fähigkeit voraus" (Scholze 2008: 245).

This is why the perfective form is licensed. In contrast to that, no directed path is available in the context of (44), because there is no standard way of how to learn a poem. The technique that you choose will always be based on your personal preferences and individual capacities. In the typical case, as argued above, the process of learning a poem will go in cycles.[13]

We have arrived at an explanation for the difference between (43) and (44). Let me summarise it. If we look at the verbal predicate alone, we will find the predicate being associated with a non-directed scale, which leads us to expect the imperfective. If we take further linguistic material into account, however, the situation can change. In (43), the presence of the expression *rajfn wekslwe* 'tyre change' evokes a script that "associates" the predicate with another gradable property, namely the property of how far one has gone through the instructions of a wheel change. This property suggests a learning path that leads through the steps 1 to 8 of the linear script in (45).[14]

(45) 1. Apply the handbrake.
      2. Position the wheel chocks.
      3. Loosen the wheel nuts.
      4. Jack the car up.
      5. Remove the flat tyre.
      6. Mount the spare wheel.
      7. Lower the car and tighten the bolts.
      8. Fully lower the car.

Since the path from 1 to 8 is directed, we now expect a perfective verb, which is what we find.

Let us now look at a second example in which the otherwise unexpected perfective aspect is licensed pragmatically:

(46) *Ta jo   tón basejn nawukła   ha jo so   hrajkać šla.*
      she AUX the poem learn.PST.PFV.F then AUX REFL play   go.PST.F
      'She learned the poem and went playing.'

The form *nawukła* '(she) learned' in (46) is perfective. According to the determinateness approach argued for in this paper, the denoted event will have to be understood as proceeding along a single, directed and upper-bounded path. It

---

[13]The android Data from Star Trek will learn a poem quickly by linearly scanning the text words once.

[14]https://www.rac.co.uk/drive/advice/car-maintenance/how-to-change-a-tyre

seems obvious that the learning of the poem in (46) is such that it traverses a single and upper-bounded path. But how about directedness?

With respect to (46), there is no reason not to assume that 'she' has learned the poem "in the usual way", that is, by revisiting the same lines, verses etc. again and again. Given the cyclic nature of the event, we should expect the imperfective aspect to be usable, and it is indeed possible to replace the perfective in (46) with its imperfective counterpart *wukła* (Scholze 2008: 245). Werner (2013) presents the following example:[15]

(47) *Wón je   list čitał   a   je šeł   preč.*
      he   AUX letter read.PST.IPFV and AUX go.PST.PFV away
      'He read a letter and went away.'

Returning to (46), how can we make sense of the possibility of perfective *nawukła* given that learning a poem typically proceeds in cycles? I would propose that the use of the perfective in (46) triggers the implicature that the use of the imperfective, although possible as well, is avoided. Avoiding the imperfective is motivated by the fact that the expression of non-directedness is not in the interest of the speaker's message. The interpreter is thus invited to "associate" the verbal concept with another gradable property, one that implies a directed path/scale. Such a property is plausibly available in the context at hand, namely the property of how far one is in the completion of a tedious task. In the given case the task is the learning of a poem (presumably as homework). Only after that job is done, will 'she' be free to do what she likes to do, playing. For the message that the tedious job is done, the question of *how* the poem was learned is irrelevant. Important is only *that* she went from the stage of not being able to recite the poem to the stage of being able to do so, (which is what she will probably have to demonstrate at school the next day). In (46), in other words, the predicate *learn the poem* is understood as abstracted from the details of the real learning process. Here the predicate describes events that traverse along a very simple path consisting of only two relevant points, $p(0)$ and $p(1)$. Figure 2 shows the path relevant for the interpretation of the first sentence of (46). I argue that this path *is* directed and that, therefore, the perfective is the appropriate form.

My discussion of the Upper Sorbian language data ends here. In the following section, I summarise the results once again and place them in a wider context.

---

[15]Note that I treat the motion verb *šeł* in CUS as a perfective (!) form, although we know its counterparts in other Slavic languages to be imperfectives. This is only consequent in view of the present analysis, within which perfectivity boils down to determinateness. See Scholze (2008: 282) for discussion.

not knowing   *knowing*
the poem   *the poem*

$p(0)$ $p(1)$
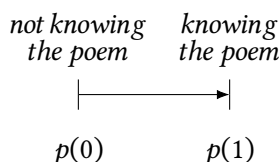
Figure 2: Path of learning the poem in (46)

10 Conclusions

Although Dickey's (2000) programmatic study is already more than 20 years old, the enterprise of carefully investigating inner-Slavic variation in aspect selection has only just begun. Until now, most studies have focused on comparing Czech and Russian, because these two well-studied languages are supposed to be good examples of Dickey's Western-type and Eastern-type languages. In the present paper, I have drawn attention to Colloquial Upper Sorbian (CUS). I have reviewed the relevant linguistic literature on that language, and I have discussed its aspectual grammar against the background of differences between Czech and Russian. My overall results may be summarised as follows.

I started from the idea that every verbal description of a dynamic event involves a path structure. The use of an imperfective verb does not impose any restrictions on the path structure of the event. The use of a perfective, however, requires the path of the denoted event to meet certain conditions. These conditions are written in the aspectual operator PFV, while its counterpart IPFV remains underspecified.

More specifically, I concluded that a perfective verb requires the path of the denoted event to be unidimensional, directed, and bounded. The use of a CUS perfective does not impose any more restrictions on interpretation than that (in Czech and Russian, by contrast, perfectives come with additional constraints). It follows that the use of a perfective in CUS will be dispreferred if the context suggests that the event path is either not unidimensional, not directed, or not bounded. The perfective form will be the preferred choice, on the other hand, if the context suggests that all three conditions are met.

If the verbal predicate describes an event that develops along a cyclic (i.e. non-directed) path, as with meanings like *learn a poem*, *iron a shirt*, *painting a wall with a roll*, or *blow-dry one's hair*, the expected aspectual form will be the imperfective. This default may be overridden, however, if the context makes salient an alternative directed path that the verbal description can adapt to. In this paper, we came across three such cases. The first was represented by (46). In this case,

the context draws attention to the finishing of the event (here: the finishing of a tedious homework), suggesting the relevant path to be a (directed) two-point scale. In the second case, represented by (43), the directed path was introduced by a linear script evoked by overt contextual material (here: the wheel change instructions). The third case was represented by (17), where the only interpretations pragmatically available involve directed paths due to the impact of a focus-bearing expression (here: the adverb *chĕtř* 'quickly').

I have drawn attention to several examples which show that CUS does not subsume to the aspectologist's common sense understanding according to which perfectives would express completed events. Does that mean that CUS is, in some sense, "unslavic", perhaps due to influence from German? Not necessarily. As Comrie (1976) points out, "completed" is the wrong feature for grasping the content of perfectivity, what is suitable instead is "complete":

> A very frequent characterisation of perfectivity is that it indicates a completed action. One should note that the word at issue in this definition is "completed", not "complete": despite the formal similarity between the two words, there is an important semantic distinction which turns out to be crucial in discussing aspect. The perfective does indeed denote a complete situation, with beginning, middle, and end. The use of "completed", however, puts too much emphasis on the termination of the situation (Comrie 1976: 18)

The notion of a complete (not: completed) event is indeed the appropriate umbrella term to subsume the perfectives in all the three languages discussed in this paper. Differences result from how this general notion manifests itself in each case. We have found three different strengths of what it means to be complete.

The strongest perfective condition is found in Russian. Here, the event referred to by a perfective has to be complete and its consequences have to be occasions for subsequent events. Czech features a weaker perfective condition. In that language, the event referred to by a perfective has to be complete, and that's all. Since there is no requirement as to "connecting" the event to neighbouring event tokens, the Czech perfective may be used to refer to a plurality of events, if only the elements of the plurality are to be understood as complete events. The weakest way of being complete is instantiated by perfectives in CUS. Here, it is not the event/situation as such which has to be complete, contra to what is said in the quote above, but the path along which the denoted event evolves.

Abbreviations

3	third person	IPFV	imperfective
AUX	auxiliary	PFV	perfective
DU	dual	PST	past
F	feminine	PRS	present
INF	infinitive	REFL	reflexive

Acknowledgements

References

Alvestad, Silje Susanne. 2013. *Beware of Fakes! Fake imperfectives in the Slavic imperative*. Oslo: University of Oslo.

Arregui, Ana, Maria Luisa Rivero & Andrés Salanova. 2014. Cross-linguistic variation in imperfectivity. *Natural Language & Linguistic Theory* 32. 307–362. DOI: https://doi.org/10.1007/s11049-013-9226-4.

Barentsen, Adrian. 1995. Trechstupenčataja model' invarianta soveršennogo vida russkom jazyke. In Stanisław Karolak (ed.), *Semantika i struktura slavjanskogo vida*, 1–28. Kraków: Wydawnictwo Naukowe WSP.

Barentsen, Adrian. 1998. Priznak sekventnaja svjas' i vidovoe protivopostavlenie v russkom jazyke. In Marina Ju. Čertkova (ed.), *Tipologija vida: Problemy, poiski, rešenija*, 43–58. Moskva: Jazyki russkoj kul'tury.

Berger, Tilman. 2013. Ungewöhnliche Verwendungen des Aspekts im Tschechischen: Der imperfektive Aspekt in Handlungssequenzen. *Zeitschrift für Slawistik* 58(1). 31–42. DOI: 10.1524/slaw.2013.0002.

Berger, Tilman. 2016. Noch einmal Imperfektiva in Handlungsfolgen. *Wiener Slavistischer Almanach* 77. 37–54.

Breu, Walter. 2000a. Der Verbalaspekt in der obersorbischen Umgangssprache im Rahmen des ILA-Modells. In Walter Breu (ed.), *Slavistische Linguistik 1999*, 37–76. München: Sagner.

Breu, Walter. 2000b. Zur Position des Slavischen in einer Typologie des Verbalaspekts. In Walter Breu (ed.), *Probleme der Interaktion von Lexik und Aspekt (ILA)*, 21–54. Tübingen: Niemeyer. DOI: 10.1515/9783110924800.21.

Breu, Walter. 2012. Aspect forms and functions in Sorbian varieties. *Language Typology and Universals* 65(3). 246–266. DOI: 10.1524/stuf.2012.0016.

Carlson, Gregory N. 1995. Truth conditions of generic sentences: Two contrasting views. In Gregory Carlson & Francis Pelletier (eds.), *The generic book*, 224–237. Chicago: University of Chicago Press.

Cohen, Ariel. 2022. Genericity. In Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics*, 1–35. Oxford: Oxford University Press. DOI: 10.1093/acrefore/9780199384655.013.326.

Comrie, Berhard. 1976. *Aspect*. Cambridge: Cambridge University Press.

Dickey, Stephen M. 2000. *Parameters of Slavic aspect*. Stanford: CSLI.

Dickey, Stephen M. 2006. Aspectual pairs, goal orientation and *po*-delimitatives in Russian. *Glossos* 7. 1–37.

Dickey, Stephen N. 2015. Parameters of Slavic aspect reconsidered: The east-west division from a diachronic perspective. In Miriam Shrager, Edna Andrews, George Fowler & Steven Franks (eds.), *Studies in accentology and Slavic linguistics in honor of Ronald F. Feldstein*, 29–45. Bloomington: Slavica.

Fasske, Helmut. 1981. *Grammatik der obersorbischen Schriftsprache der Gegenwart. Morphologie*. Bautzen: Domowina-Verlag.

Filip, Hana. 2000. The quantization puzzle. In James Pustejovsky & Carol Tenny (eds.), *Events as grammatical objects: The converging perspectives of lexical semantics and syntax*, 3–60. Stanford: CSLI.

Filip, Hana. 2008. Events and maximalization: The case of telicity and perfectivity. In Susan Rothstein (ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect*, 217–256. Amsterdam: John Benjamins. DOI: 10.1075/la.110.10fil.

Filip, Hana. 2017. The semantics of perfectivity. *Italian Journal of Linguistics* 29(1). 167–200. DOI: 10.26346/1120-2726-107.

Gehrke, Berit. 2008. *Ps in motion: On the semantics and syntax of P elements and motion events*, vol. 184 (LOT Dissertation Series). Utrecht: LOT.

Gehrke, Berit. 2022. Differences between Russian and Czech in the use of aspect in narrative discourse and factual contexts. *Languages* 7(2). 155. DOI: https://doi.org/10.3390/languages7020155.

Greenberg, Yael. 2003. *Manifestations of genericity*. New York: Routledge.

Grønn, Atle. 2004. *The semantics and pragmatics of the Russian factual imperfective*. Oslo: Acta Humaniora.

Heck, Stefan. 2018. Verbal aspect in the Czech and Russian imperative. In Erik Fuß, Marek Konopka, Beata Trawínski & Ulrich Waßner (eds.), *Grammar and corpora 2016*, 249–258. Heidelberg: Heidelberg University Publishing. DOI: 10.17885/heiup.361.c4705.

Hesni, Samia. 2022. Normative generics and social kind terms. *Inquiry*. 1–24. https://doi.org/10.1080/0020174X.2022.2032323.

Kagan, Olga. 2015. *Scalarity in the verbal domain: The case of verbal prefixation in Russian*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781316136195.

Karlík, Petr, Marek Nekula & Zdenka Rusínová. 1995. *Pžíruční mluvnice češtiny*. Praha: Nakladatelství Lidové noviny.

Klein, Wolfgang. 1994. *Time in language*. London & New York: Routledge.

Klimek-Jankowska, Dorota. 2008. On an analogy between English nominal and Polish aspectual manifestations of genericity. In Atle Grønn (ed.), *Proceedings of Sinn und Bedeutung 12*, 318–336. Oslo: ILOS.

Klimek-Jankowska, Dorota. 2012. Imperfective and perfective habituals in Polish: A bi-directional OT account of variation and ambiguity. *Journal of Logic, Language and Information* 21(1). 31–52. DOI: https://doi.org/10.1007/s10849-011-9154-9.

Klimek-Jankowska, Dorota. 2022. Variation in aspect usage in general-factual contexts: New quantitative data from Polish, Czech, and Russian. *Languages* 7(2). 146. DOI: https://doi.org/10.3390/languages7020146.

Krifka, Manfred. 1998. The origins of telicity. In Susan Rothstein (ed.), *Events and grammar*, 197–235. Dordrecht: Kluwer.

Krifka, Manfred. 2013. Definitional generics. In Alda Mari, Claire Beyssade & Fabio Del Prete (eds.), *Genericity*, 372–389. Oxford: Oxford University Press.

Krifka, Manfred, Francis Pelletier, Gregory N. Carlson, Alice ter Meulen, Gennaro Chierchia & Godehard Link. 1995. Genericity: An introduction. In Gregory N. Carlson & Francis Pelletier (eds.), *The generic book*, 1–124. Chicago: University of Chicago Press.

Leslie, Sarah-Jane & Adam Lerner. 2022. Generic generalizations. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Fall 2022. Stanford University: Metaphysics Research Lab. https://plato.stanford.edu/archives/fall2022/entries/generics/.

Lewaszkiewicz, Tadeusz. 2002. Obersorbisch. In Miloš Okuka (ed.), *Lexikon der Sprachen des europäischen Ostens*, 343–354. Klagenfurt: Wieser.

Mueller-Reichau, Olav. 2018. General-factual perfectives: On an asymmetry in aspect choice between western and eastern Slavic languages. In Radek Šimík, Denisa Lenertová, Roland Meyer & Luka Szucsich (eds.), *Advances in formal*

Slavic linguistics 2016, 289–311. Berlin: Language Science Press. DOI: DOI:10. 5281/zenodo.2545531.

Mueller-Reichau, Olav. 2020. On concealed properties in Polish perfective generics. In Teodora Radeva-Bork & Peter Kosta (eds.), *Current developments in Slavic linguistics: Twenty years after (based on selected papers from FDSL 11)*, 273–290. Berlin: Language Science Press. DOI: 10.3726/978-3-653-07147-4.

Mueller-Reichau, Olav. Forthcoming. The morphosemantics of Russian aspect. In Berit Gehrke & Radek Šimík (eds.), *Topics in the semantics of Slavic languages.* Berlin: Language Science Press.

Müller, Ana & Luciana Sanchez-Mendes. 2020. Pluractionality: The phenomenon, the issues, and a case study. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann & Thomas Ede Zimmermann (eds.), *The Wiley Blackwell companion to semantics*, 1–34. Hoboken: John Wiley. DOI: https://doi.org/ 10.1002/9781118788516.sem134.

Petruchina, Elena V. 2000. *Aspektual'nye kategorii glagola v russkom jazyke v sopostavlenii s češskim, slovackim, pol'skim i bolgarskim jazykami.* Moskva: MGU.

Prasada, Sundeep & Elaine Dillingham. 2006. Principled and statistical connections in common sense conception. *Cognition* 99(1). 73–112. DOI: 10.1016/j. cognition.2005.01.003.

Rachilina, Ekaterina V. 2000. *Kognitivnyj analiz predmetnych imen: Semantika i sočetaemost'.* Moskva: Russkie slovari.

Rivero, Maria Luisa & Ana Arregui. 2010. Aspectual microvariation: The case of Slavic imperfectives. In Melinda Heijl (ed.), *Proceedings of the 2010 annual conference of the Canadian Linguistic Association*, 1–15. Toronto: Toronto University.

Ščerba, Lev V. 1973. *Vostočnolužickoe narečie (reprint from 1915).* Bautzen: Domowina.

Schank, Roger & Robert Abelson. 1977. *Scripts, plans, goals, and understanding.* Hillsdale: Erlbaum.

Scholze, Lenka. 2008. *Das grammatische System der obersorbischen Umgangssprache im Sprachkontakt.* Bautzen: Domowina-Verlag.

Scholze, Lenka. 2023. Der Verbalaspekt in der obersorbischen Umgangssprache im Sprachkontakt. In Walter Breu & Malinka Pila (eds.), *L'aspettualità nel contatto linguistico: Lingue slave e oltre*, 161–182. Firenze: Firenze University Press. DOI: 10.36253/979-12-215-0184-1.11.

Seres, Daria & M. Teresa Espinal. 2019. Russian definitional generic sentences. *Glossa: a journal of general linguistics* 4(1). 1–30. DOI: https://doi.org/10.5334/ gjgl.760.

Stone, Gerald. 1993. Sorbian. In Bernard Comrie & Greville Corbett (eds.), *The Slavonic languages*, 593–685. London & New York: Routledge.

Stunová, Anna. 1991. In defence of language-specific invariant meanings of aspect in Russian and Czech. In Adrian Barentsen, Ben Groen & Rob Sprenger (eds.), *Studies in West Slavic and Baltic linguistics*, 259–319. Amsterdam: Rodopi.

Stunová, Anna. 1993. *A contrastive analysis of Russian and Czech aspect: Invariance vs. discourse.* Amsterdam: University of Amsterdam. (Doctoral dissertation).

Švedova, Natalija J., Nina D. Arutjunova, Aleksandr V. Bondarko, Valerij V. Ivanov, Vladimir V. Lopatin, Igor' S. Uluchanov & Fedot P. Filin. 1980. *Russkaja grammatika.* Moskva: Nauka.

Toops, Gary. 1998. The scope of secondary imperfectivization in Bulgarian, Russian and Upper Sorbian. In Robert A. Maguire & Alan Timberlake (eds.), *American contributions to the twelfth international congress of Slavists: Cracow, August–September 1998*, 515–529. Bloomington: Slavica.

Toops, Gary. 2001. Aspectual competition and iterative contexts in contemporary Upper Sorbian. *Journal of Slavic Linguistics* 9. 127–154.

Werner, Eduard. 2003. *Die Verbalaffigierung im Obersorbischen.* Bautzen: Domowina-Verlag.

Werner, Eduard. 2013. Upper Sorbian and verbal aspect. In Sorin Paliga (ed.), *The verbal aspect in the Slavic languages. Special issue of Romano-Bohemica*, 165–182. Bucharest: Editura.

Werner, Eduard. 2018. Aspect and performativity in Sorbian. *Slavia Occidentalis* 74(1). 143–149. DOI: https://doi.org/10.14746/so.2018.75.10.

Wiemer, Björn. 2008. Zur innerslavischen Variation bei der Aspektwahl und der Gewichtung ihrer Faktoren. In Karl Gutschmidt, Witold Kosny & Peter Kosta (eds.), *Deutsche Beiträge zum 14. internationalen Slavistenkongress, Ohrid 2008*, 383–409. München: Sagner. DOI: 10.20378/irbo-3763.

Zwarts, Joost. 2005. Prepositional aspect and the algebra of paths. *Linguistics and Philosophy* 28(6). 739–779. DOI: 10.1007/s10988-005-2466-y.

Chapter 14

Polar questions in Czech and Russian: An exploratory corpus investigation

Maria Onoeva & Anna Staňková
Charles University

This study aims to bring new insights into the topic of polar questions in Czech and Russian based on corpus data. What is of particular interest are the complex differences in meaning among the Czech and Russian counterparts of English structures such as *Is Jane coming?*, *Isn't Jane coming?* and *Is Jane not coming?*. We examine the formal and semantic/pragmatic features of polar questions in these two Slavic languages, namely word order, presence and position of negation, presence of question tags, presence of question particles, and their relation to the question's meaning and its bias towards a possible answer. Using authentic data from spoken corpora allowed us to observe some prominent tendencies of polar questions usage.

1 Introduction

Polar questions (PQs) have been widely studied from different points of view in recent years.[1] A number of researchers brought important insights about how the meaning of PQs is affected by negation (e.g. Büring & Gunlogson 2000, Romero & Han 2004, Repp 2013, AnderBois 2019), word order (e.g. Gunlogson 2002), particles (e.g. Sudo 2013, Frana & Rawlins 2019, Gärtner & Gyuris 2022, Gonzalez 2023), intonation or focus (e.g. Gyuris 2019, Rudin 2022, Goodhue 2022), and other phenomena. The goal of this paper is to contribute to the topic from the perspective of Slavic languages, namely Czech and Russian.

---

[1]The term "question" is related rather to a speech act, whilst "interrogative" is used for syntactic and semantic descriptions. In this paper, however, we stick to the term "polar question" as it is more frequent in the literature.

There are two main strategies how to ask a PQ in these languages – overt and intonational (Šimík forthcoming). In Czech, a PQ can be constructed by interrogative word order (Štícha 1995), which involves the finite verb preceding an overt subject, as in (1a). The second strategy is using intonation, either the rise or fall-rise pattern (Daneš et al. 1987, Palková 1994). Thanks to this, declarative sentences can be interpreted as PQs, as shown in (1b). The results of the experiment run by Staňková (2023) showed that the choice between these two strategies (interrogative vs. declarative) can be motivated by the presence of evidential bias (more on bias in Section 2.2).

(1) a. Koupil si Petr auto?
bought REFL Petr car
'Did Petr buy a car?'

b. Petr si koupil auto?
Petr REFL bought car
'Petr bought a car?' (Cz)

In Russian, the overt strategy is to place the particle *li* after the first phonological word as in (2a). Any word can appear with it and then be in the question focus (King 1994). The intonational strategy is shown in (2b). For out-of-the-blue PQs, word order is declarative but the pitch locus (a steep rise and an immediate fall; Q-PEAK by Esipova 2024) is placed at the verb, here it is *vyigrala* 'won', whereas in statements it is usually placed at the most deeply embedded argument (*priz* 'prize' in this case) (Meyer & Mleinek 2006, Rathcke 2006). Schwabe (2004) and Brown & Franks (1995) mention the markedness of *li* in main clauses and its ongoing loss in colloquial Russian. Nevertheless, *li* must be still present in embedded PQs, (2c). Esipova & Korotkova (2024) argue that *li*-PQs simply present two alternatives and thus are true neutral questions, whereas intonation PQs convey pressure to respond.

(2) a. Vyigrala li Daša priz?
won LI Daša prize
'Did Daša win a prize?'

b. Daša vyigrala priz?
Daša won prize
'Did Daša win a prize?'

c. Ja ne znaju, vyigrala li Daša priz.
I not know won LI Daša prize
'I don't know whether Daša won the prize.' (Ru)

In the present study, we looked at PQs in general through the lens of corpus data. Besides the above-described features, Czech and Russian questions could contain various elements which directly influence their meaning, such as indefinites, different particles, tags, negation etc. Due to the limitations of the Russian corpus, prosody was not taken into account. Answers to PQs were also laid aside. It was an exploratory study, in which we addressed the following research questions:

1. What are the formal properties of PQs in real communication?

2. Besides the core interrogative semantics, what semantic/pragmatic implications do PQs have?

3. Are there any correlations between the formal and semantic/pragmatic aspect?

To answer the first research question, we annotated each PQ with respect to its structure. For the second research question, we explored question biases (Büring & Gunlogson 2000, Sudo 2013, Gärtner & Gyuris 2017) and their distribution among Czech and Russian PQs. The third research question was to check if there is any relation between their form and meaning.

The paper is organized as follows. In Section 2, we describe the method of annotation. Section 3 reports on the absolute values of the annotated features and the results of the inter-annotator agreement. In Section 4, we discuss the results. Section 5 concludes the paper.

2 Method

In this section, we describe the method and procedure of the annotation, which were the same for both languages. We used the spoken corpus of the Russian National Corpus (Grišina 2005, Grišina & Savčuk 2009) and the ORTOFON v2 corpus (Kopřivová et al. 2020) of the Czech National Corpus, the latter accessed via the KonText interface (Machálek 2014). Both corpora contain informal everyday conversations with the option to display a limited context around the question. Audio is not available for the spoken part of the Russian corpus, thus intonation was not taken into account.

For each language, a random sample of 500 instances was manually collected. We queried for the question mark and filtered out *wh*-interrogatives. In order to address the first and second research questions, the annotated features were

divided into two groups – formal and semantic/pragmatic. They are described below.[2]

2.1 Formal features

We have already mentioned some of the formal features of PQs, such as the specific word order or usage of question particles. In our sample, we annotated word order with respect to the position of the verb – it was either initial, medial or final. As for particles, for each one we marked their presence ('1' = present, '0' = absent).

Previous research paid attention to negation in PQs because of non-trivial implications it involves (e.g. Ladd 1981, Repp 2013). It was claimed that there are two types of negation – inner (≈ semantic) and outer (≈ pragmatic) – and that they differ in their syntactic and semantic/pragmatic features. Inner negation is interpreted and licenses Negative Polarity Items [NPIs] (Negative Concord Items [NCIs] in Czech and Russian), whereas outer negation does not trigger the negative operator per se and licenses Positive Polarity Items [PPIs] (Romero & Han 2004; cf. Goodhue 2022). Based on these observations, we annotated our data set for the presence of negation as well as certain indefinites. For Czech, these were *ně*-indefinites (considered as PPIs) and *ni-/žád*-indefinites (considered as NCIs). In Russian, they were *-nibud', -to, koe-* indefinites and *ni-* NCIs.

The last annotated formal feature was the question tag. Tag questions consist of an anchor (the PQ) and a tag. There are different types of tags based on their polarity (e.g. Krifka 2015). The first type agrees in polarity with the PQ (matching tags), the second type is of the opposite polarity than the PQ (reverse tags). Moreover, tags can differ in their intonation patterns (Ladd 1981). In our annotation, we marked their presence, but did not distinguish them any further.

In (3), we provide an example of a Czech PQ annotated from a formal point of view for all the features just mentioned.

(3) Snad ho   teďka nebudeš stavět ne?
     SNAD him now   NEG-will build no
     'You're not going to build it now, are you?'   (Cz)
     Formal annotation: WO: xVx; PRT: 1 *snad*; NEG: 1; INDEF: 0; TAG: 1 *ne*

2.2 Semantic/pragmatic features

After Hamblin (1973), Karttunen (1977), and Groenendijk & Stokhof (1984), the semantic interpretation of questions is represented as a set of their (true or pos-

---

[2]The complete annotation is available here: https://bit.ly/3xKM9XX

sible) answers. In case of PQs, it can be simplified to $\{p, \neg p\}$ where p is a question radical, e.g. for a question *Is it raining?*, $p = it\ is\ raining$.

Aside from that the structure of PQs may indicate a certain favor, or BIAS, towards a particular reply, which is not captured by the set of their possible answers. So far SPEAKER (or EPISTEMIC) and EVIDENTIAL biases are recognized. They usually either support or oppose p. Speaker bias is based on prior and private speaker's beliefs,[3] while evidential bias comes from contextual information available to all interlocutors. Not all PQs are equally biased, it is possible that one of the biases or both are absent. If no bias is present, the question is considered to be neutral. Different combinations of biases and their absence represent BIAS PROFILES of PQs and could be universal to specific question forms or particles (Sudo 2013, Gärtner & Gyuris 2017).

In our annotation, we also distinguished a third type of bias which was related to the speaker's awareness of the answer, and we refer to it as KNOWLEDGE bias. If the speaker knows the answer for sure, the PQ is biased. This type of bias is sort of in between speaker and evidential biases. It typically occurs in exam (Krifka 2011) or surprise echo questions.

To be able to investigate the bias profiles of Czech and Russian PQs, we manually constructed an affirmative prejacent ϕ for each question from their radicals. We performed the following steps to produce it: (i) remove negation if it is present, (4a); (ii) remove particles, question tags and other elements that do not appear in statements as in (4b) and (4c); (iii) if the first or second person pronouns appear, replace them with 'speaker' or 'addressee' as in (4d).

(4)   a.   Není to kočka?
         NEG.is it cat
         'Isn't it a cat?'   (Cz)
         ϕ = It's a cat.

     b.   A Daník tady bude ne?
         and Daník here will.be no
         'And Daník is going to be here, isn't he?'   (Cz)
         ϕ = Daník is going to be here.

     c.   Neuželi oni tože slyšat kak my rugaemsja?
         NEUŽELI they also hear how we argue
         'Do they also hear how we argue?'   (Ru)
         ϕ = They also hear how we argue.

---

[3]Sudo (2013) suggests that epistemic/speaker bias can also include deontic or bouletic states. Since these are too complex to judge based on written text, we only work with speaker's beliefs.

d. A u tebja pomimo sobački est' eščë kto-nibud'?
    and at you besides doggie is else anyone
    'Do you have anyone else besides a doggie?'   (Ru)
    ϕ = The addressee has someone else besides a doggie.

We used affirmative prejacents and not question radicals to decide whether or not speakers had any prior belief. The same applies for evidential bias. With the aid of prejacents, it was easier to judge the type of bias and its value in some controversial cases such as PQs with certain particles or outer negation cases, where it was not clear if the radical was affirmative or not.

Judgments about the biases were based on our intuition as native speakers, which were later compared with judgments from three additional annotators (see Section 3.4). We always annotated the questions in some amount of context in order to detect evidential bias. Due to the limitations of the corpora, we were able to capture only linguistic cues of evidence.

To construct the bias profiles, we assigned each bias one of the three values: '1', '0' and '−1'. The value '1' was assigned if the bias supported the affirmative prejacent. For instance, if the speaker believed that ϕ before posing the question, '1' was assigned to speaker bias. On the contrary, the value '−1' was assigned when the bias went against ϕ or, in the other words, supported that $\neg\phi$. E.g. if in the context there was a cue suggesting that $\neg\phi$, evidential bias for such cases was '−1'. The value '0' was assigned if no bias was detected.

The following examples clarify the annotation of the bias profile. In the context of (5), it is mentioned who is the oldest brother and the youngest, Leonid and Aleksandr Aleksandrovič, respectively. The context supports the prejacent ϕ, hence, the value assigned to evidential bias is '1'. The particle *razve* indicates that the speaker's prior belief was that $\neg\phi$ (Geist & Repp 2023, Korotkova 2023), so Viktor believed that Aleksandr Aleksandrovič was not the youngest. The value assigned to speaker bias is '−1'. Since it is clear from the context in (5) that the speaker now knows that ϕ (the speaker mentions the brothers' age difference explicitly), the value assigned to knowledge bias is '1'.

(5) Context: The addressee says her husband, aged 28 at the time, had two brothers: Leonid, 30, and Aleksandr Aleksandrovič, 27. The speaker asks:

Sp: {Kak / razve} Aleksandr Aleksandrovič mladšij?
       how RAZVE Aleksandr Aleksandrovič youngest
     'Wait a second, is Aleksandr Aleksandrovič the youngest?'   (Ru)
     ϕ = Aleksandr Aleksandrovič is the youngest.

Semantic/pragmatic annotation: SPEAKER −1, EVIDENTIAL 1, KNOWLEDGE 1

In (6), the speaker has some prior belief when the addressee leaves because they explicitly say the time of leaving, so the epistemic bias value is '1'. There were no contextual cues, so the evidential and knowledge bias values are '0'.

(6) Context: The speaker promised some sausage to the addressee but did not manage to bring it. They want to do it later and check when the addressee is available.

Sp: Ty   pojedeš kolem vosmý nějak?
     you leave   around eight   somehow
'Are you leaving around eight?'   (Cz)
$\phi = $ The addressee is leaving around eight.

Semantic/pragmatic annotation: SPEAKER 1, EVIDENTIAL 0, KNOWLEDGE 0

In (7), there is an example of evidential bias only. The speaker bias value is '0' because the speaker had no prior belief about the prejacent and guesses the number from the context. Since they do not know for sure how many cars were there, the value assigned to knowledge bias is '0'.

(7) Context: The addressee lists how many people were with them on a trip. The speaker assumes the following from the provided number of people:

Sp: Vy jste jeli   třema autama?
     you AUX went tree   cars
'Did you guys travel in three cars?'   (Cz)
$\phi = $ The addressee and the group travelled in 3 cars.

Semantic/pragmatic annotation: SPEAKER 0, EVIDENTIAL 1, KNOWLEDGE 0

The example in (8) was annotated as carrying knowledge bias only. There is no prior belief about the cat being in the speaker's spot and no linguistic evidence of that.

(8) Context: The speaker sees their cat Ryžik sitting in the speaker's spot.

Sp: Ryžik ty   čto   moë mesto zanjal?
     Ryžik you what my   spot   taken
'Ryžik, have you taken my spot?'   (Ru)
$\phi = $ Ryžik has taken the speaker's spot.

Semantic/pragmatic annotation: SPEAKER 0, EVIDENTIAL 0, KNOWLEDGE 1

3 Results

In this section, we outline the results of the study. We begin with the absolute values for both the formal and semantic/pragmatic features, then we report the results for Czech and Russian separately. The inter-annotator agreement report concludes the section.

3.1 Overall values

Table 1 summarizes the overall frequencies of the formal features we annotated. From the NEG column follows that, out of 500 PQs, only 89 Czech and 79 Russian were negated. Tag questions were much more frequent in Czech (154) than in Russian (46).

The occurrence numbers of the possible verb positions are in the third column. For Czech, the medial position was the most frequent one (220), followed by initial (150) and final (97). For Russian, we have found fewer PQs with the verb placed at the initial position (53), medial was the second frequent option (148), the most popular was final (237). Some cases were excluded from the final analysis because they were not full sentences (e.g. only subject or object with no predicate).

Table 1: Formal features

	NEG	TAG	VERB POSITION		
	1	1	INI	MED	FIN
Czech	89	154	150	220	97
Russian	79	46	53	148	237

The distribution of the biases in the samples is reported in Table 2. The most striking difference between the languages is present in the column SPEAKER with the value '1' representing speaker bias that ϕ (104 instances for Czech and only 35 for Russian). Aside from that the other bias values were distributed equally in both languages.

In the next subsections, we focus on particular form-meaning correlations in Czech and Russian PQs, respectively. For purposes of the following analyses, we pooled non-null speaker bias and evidential bias and excluded the few exceptional cases where the polarity of speaker/evidential bias was not in accord with the question's polarity.

Table 2: The distribution of the biases in the samples

	SPEAKER			EVIDENTIAL			KNOWLEDGE		
	0	1	-1	0	1	-1	0	1	-1
Czech	379	104	17	353	113	34	446	45	9
Russian	453	35	12	364	109	27	465	31	4

3.2 Czech

The two form-meaning correlations we zoomed in on were: (i) the presence of a TAG and the value of speaker bias (SB), and (ii) verb position and the value of evidential bias (EB). According to the null hypothesis, the variables in the pairs (i) and (ii) are independent of each other. According to the alternative hypothesis, the variables in those two pairs correlate. To test the alternative hypothesis for each pair, we ran two chi-square tests on the Czech data. Since there were two tests run on the data, we used Bonferroni correction (α divided by n, where n is the number of tests) to adjust the alpha level. The adjusted alpha level was 0.025 (0.05 divided by 2).[4] We report the absolute values in the contingency tables 3 and 4. Values expected in the case the null hypothesis is true are in brackets. Now we comment on the two pairs of variables individually.

As mentioned above, tag questions occurred frequently in the Czech data set, and mostly they (the anchor) exhibited declarative word order. Table 3 shows the correlation between tag PQs and speaker bias, i.e. previous beliefs of the speaker; which was statistically significant ($\chi^2(1) = 120.9$; $p < .001$). For non-null speaker bias (= biased PQs), tags were present in 77 cases, even though the expected value by the null hypothesis was 31.7. For null speaker bias, the expected value was 120.3, but tags occurred only 75 times.

Table 3: Tag–belief correlation ($p < .001$)

	NO TAG		TAG	
SB ±1	21	(66.3)	77	(31.7)
SB 0	297	(251.7)	75	(120.3)

---

[4]We thank an anonymous reviewer for pointing out the inconsistencies in our reporting of the results.

In our sample, initial verb position (V ɪɴɪ) negatively correlates with non-null evidential bias, and this correlation was significant ($\chi^2(1) = 11.8$; $p < .001$). This is apparent from Table 4, where the expected value of v ɪɴɪ in non-null ᴇʙ PQs is higher (42.6) than the actual value (27). What follows from this is that the interrogative word order is preferred in PQs occurring in neutral context, i.e. those, where evidential bias equals '0'. When it was non-null, the PQ exhibited declarative word order (V ɴᴏɴ-ɪɴɪ).

Table 4: V-position–evidence correlation ($p < .001$)

	V ɴᴏɴ-ɪɴɪ		V ɪɴɪ	
ᴇʙ ±1	103	(87.4)	27	(42.6)
ᴇʙ 0	211	(226.6)	126	(110.4)

As for indefinites, they were present in 59 Czech PQs. Fifty-four of them were of the *ně*-type (PPIs) and the rest were of the *ni-* or *žád*-type (NCIs).

Question particles were very sparse in the Czech data set. There was only one occurrence of *náhodou* and one of *snad.*

3.3 Russian

In the Russian sample, various particles were found. The most frequent was the particle *čto li* and its variations (26). The initial *čto* (occasionally also medial) was also quite frequent in PQs, we have detected 14 cases. The presence of the particle *razve* suggestively correlates with the speaker bias: all 10 cases displayed it, 6 of them carried evidential bias. Out of 143 cases with particles, *li* was found in 6 of them. Only one question had the particle *neuželi.* Thirty-eight PQs with indefinites were discovered, 18 of them contained the *nibud'* series indefinites, 14 *to*, 6 *ni.*

Unlike for Czech, verbs (and non-verbal predicates) were not very often at the initial position. The most frequent was the final position but no significant correlation was found between the verb position and any of the biases.

Due to the low number of speaker bias in the Russian sample, we were unable to draw any conclusive results about the correlations between the form and this type of bias. The same applies to tags. We address it in the discussion.

3.4 Inter-annotator agreement

To test the reliability of our annotation, we recruited three students in order to later conduct the inter-annotator agreement for the semantic/pragmatic features. One Czech and two Russian speakers were paid to perform the same annotation of the bias profiles for 100 random instances from the samples. They were given instructions how to judge the biases and the affirmative prejacents for each PQ.

We compared the annotation of the recruited students with our own. Table 5 summarizes the results, the annotators are in complete agreement if $\kappa = 1$ (Cohen 1960). The agreement for Czech was moderate to substantial. For Russian, it was poor in the first case, slightly better in the second, moderate for evidential bias but still poor for the other two. Potential reasons for it will be discussed in the next section.

Table 5: The agreement for the bias annotation, κ

	Czech	Russian 1	Russian 2
SPEAKER	0.54	0.05	0.24
EVIDENTIAL	0.60	0.16	0.43
KNOWLEDGE	0.62	0.33	0.06

4 Discussion

In this section we discuss the results and further comment on the process of annotation. Since both Czech and Russian are Slavic languages, we expected them to behave similarly. This assumption was met to some extent. For example, negative PQs were much less frequent than the positive ones in both languages – out of 500 PQs in each language, there were 89 negative PQs in Czech and 79 in Russian. This finding is consistent with previous research which claims that positive PQs are the unmarked way of requesting information. However, the two languages differed substantially in the frequency of tag PQs. There were also differences in the verb position, which was probably connected to the languages' preferred syntactic mechanisms. As for indefinites, their occurrence in our sample was too sparse to draw any generalizations based on them, although it is an issue that we would definitely like to address in the future.

Czech and Russian showed comparable distribution of values of evidential bias and knowledge bias, but they differed in speaker bias. We suppose that speaker

bias was a category too difficult to be objectively judged based on written material, which might have led to this discrepancy between the languages. The overall number of instances with knowledge bias was small for both languages (around 10%). It is perhaps uncommon for speakers to know for sure the answer to the question they ask because it violates Interrogativity principle (Goodhue 2018).[5]

We continue with the discussion for each of the languages separately.

4.1 Czech

In Czech tag PQs, declarative word order correlated with non-null speaker bias, i.e. tag PQs showed a strong tendency to be biased with respect to what the speaker believed. By uttering such a question, the speaker expresses their prior belief, but as it is not definitive, they shift the commitment onto the addressee at the same time. Since our annotation did not go into so much detail, we could not draw any conclusions about the different types of tags, although it would be an interesting follow-up. Our findings agree with previous research, which claims that tag PQs are mostly biased (Bill & Koev 2023).

Additionally, we observed a negative correlation between interrogative word order and null evidential bias. In these PQs, the speaker has no expectation about the possible answer based on public information shared by the participants. If there is, however, a piece of compelling contextual evidence, the declarative word order would be favored. Again we see that declarative PQs tend to carry a bias. Since this is the case, we have a reason to believe that the interrogative word order is the default strategy of forming an unbiased PQ in Czech, supporting previous claims e.g. by Štícha (1995).

The situation is different when it comes to interrogative PQs with negation, as they do come with a bias. By uttering such a question the speaker expresses that they think that one of the alternatives (p or $\neg p$) is possibly true. It seems that this bias is weaker in its meaning than that of English preposed negation, which is claimed to convey that the speaker *believes* that p or $\neg p$. Czech preposed negation is another issue that would deserve a closer look.

4.2 Russian

We had to analyze the results of the semantic/pragmatic features for Russian cautiously, since the inter-annotator agreement was poor for both annotators. The potential explanation for this is the nature of the Russian corpus. Compared

---

[5] INTERROGATIVITY PRINCIPLE: Ask a question $?p$ only if the context set c does not entail a complete answer to $?p$.

to the Czech corpus, it displays less context around the question; usually there were two additional lines of text. This could drastically influence the judgments of the bias profiles because evidential bias requires as much context as possible. We suppose that the number of PQs biased with respect to what the speaker believes is different from Czech in Table 2 for the same reason, since it is tricky to judge it considering formal features only. However, we have checked the cases where there was at lease some agreement between the annotators.[6] Such PQs usually contained particles or were annotated as carrying no bias. For the cases with particles, agreement is justified since particles are reliable markers of various biases. In no bias agreement cases, we cannot be certain that the PQs are completely neutral due to little context availability and no audio.

Based on our corpus data we cannot conclude that the low number of tags for Russian in Table 1 signals their lack in colloquial speech compared to Czech. The spoken Russian corpus contains various texts from the sixties until the present days (Grišina & Savčuk 2009) and it seems they were annotated differently in the corpus itself. For instance, tags were often separate one-word questions in the older texts, while in the modern ones they were divided by the pause marking slash '/'. Therefore, more investigation is required, preferably with audio.

When it comes to particles, it is not surprising that *li* was not very frequent in the spoken corpus. As mentioned in the introduction, it is quite marked in colloquial Russian or used in truly neutral contexts but it was not completely absent.

The particles *neuželi* and *razve* occurred in our sample and were recently investigated in a series of experiments by Geist & Repp (2023). Their claim is that *neuželi* denotes VERUM, an epistemic operator indicating the speaker's intention to add the proposition in question to the common ground. *Neuželi* is also incompatible with another illocutionary operator FALSUM which is responsible for outer negation interpretation. *Razve* is compatible with both VERUM and FALSUM. Our findings neither support nor dismiss that since the number of the particles (1 and 10, respectfully) was not sufficient to make any constructive judgments; however, all the cases were biased in one way or another.

Čto li (literally translated as 'what whether') was the most frequent particle. Restan (1969) and Dobrovol'skij & Levontina (2014) mention its presumptive and emotional nature, in other words saying it introduces some bias and is infelicitous in out-of-the-blue PQs. Generally, questions with this particle, e.g. *Na ulice dožd' čto li idët?* 'It is raining outside?', are used in contexts when there is an evidence for *p* and the speaker wants to confirm that *p* (out of 26 cases, 10 questions had

---

[6]We thank an anonymous reviewer who recommended to check it.

the value '1' for evidential bias). A private speaker belief that $\neg p$ is possible but not necessary for affirmative PQs but obligatory for negated PQs with the particle. Unlike the mentioned *neuželi* and *razve*, this particle is also available in declaratives and imperatives contributing epistemic modality flavor as non-at-issue meaning (cf. Bernasconi 2023).

5 Conclusion

The goal of our study was to contribute to the empirical investigation of Slavic PQs. By exploring the properties of PQs through corpora, we addressed three research questions concerning their form and meaning. Corpus proved to be a convenient means of investigation, which allowed us to quickly collect authentic language data. We collected a sample of 500 PQs for each language, in which we were able to observe some tendencies, although it was probably not the best way to evaluate meaning shades, such as the biases.

We have run the inter-annotator agreement for semantic/pragmatic features for both languages. The agreement was moderate to substantial for Czech but poor for Russian. We hypothesize that variations in the corpora may account for this discrepancy. To avoid it in future, we suggest to ensure that a corpus shows at least 10 lines of text prior to a query, corpus texts are annotated in a unified fashion and audio is available.

The contribution of our research is mainly empirical. We observed some interesting form-meaning correlations for Czech, for instance, tag PQs tend to express speaker's belief and initial verb PQs mostly do not carry evidential bias. For Russian, we have seen that the intonational strategy is used predominately in spoken language which supports the previous observations. Moreover, we found some different particles that exhibit certain biases, e.g. *neuželi*, *razve* and less studied *čto li*. In future research, we plan to concentrate on particular phenomena, e.g. negation, particles or intonation in Czech and Russian PQs, since we barely touched upon these or did not even consider them in the corpus investigation.

Abbreviations

AUX	auxiliary	PRT	particle
INDEF	indefinite	REFL	reflexive
NEG	negation	WO	word order

Acknowledgments

References

AnderBois, Scott. 2019. Negation, alternatives, and negative polar questions in American English. In Klaus von Heusinger, Edgar Onea & Malte Zimmerman (eds.), *Questions in discourse – Volume 1: Semantics*, 118–171. Leiden: Brill. DOI: 10.1163/9789004378308_004.

Bernasconi, Beatrice. 2023. Polyfunctional particles in spoken Russian: The case of *čto li. Journal of Pragmatics* 208. 77–90. DOI: 10.1016/j.pragma.2023.02.008.

Bill, Cory & Todor Koev. 2023. *Question bias from polarity focus.* Ms. ZAS/U. Konstanz. https://www.corybill.com/publications/bill2023question0.pdf.

Brown, Sue & Steven Franks. 1995. Asymmetries in the scope of Russian negation. *Journal of Slavic Linguistics* 3(2). 239–287. https://www.jstor.org/stable/24598950.

Büring, Daniel & Christine Gunlogson. 2000. *Aren't positive and negative polar questions the same?* Ms. USCS/UCLA. https://semanticsarchive.net/Archive/mYwOGNhO/polar_questions.pdf.

Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1). 37–46. https://doi.org/dghsrr.

Daneš, František, Zdeněk Hlavsa & Miroslav Grepl. 1987. *Mluvnice češtiny*, vol. 3. Prague: Academia.

Dobrovol'skij, Dmitrij & Irina Levontina. 2014. Diskursivnye slova v obščevoprositel'nyx predloženijax: Russko-nemeckie sootvetstvija. In Vladimir P. Selegej, Aleksej V. Bajtin, Vladimir I. Belikov, Igor' M. Boguslavskij, Boris V. Dobrov, Dmitrij O. Dobrovol'skij, Leonid M. Zaxarov, Leonid L. Iomdin, Irina M. Kobozeva, Elena B. Kozerenko, Maksim A. Krongauz, Natalija I. Laufer, Natal'ja V. Lukaševič, Dijana Makkarti, Joakim Nivre, Gennadij S. Osipov, Viktor Raskin, Ėduard Xovi & Sergej A. Šarov (eds.), *Komp'juternaja lingvistika i intellektual'nye texnologii: Po materialam ežegodnoj meždunarodnoj konferencii*

"Dialog" (2014), 138–149. Moskva: Komp'juternaja lingvistika i intellektual'nye texnologii.

Esipova, Masha. 2024. *Prosody across sentence types*. Submitted to "Proceedings of SALT 34". https://lingbuzz.net/lingbuzz/008389.

Esipova, Masha & Natasha Korotkova. 2024. *To li or not to li*. Talk at *Polar Question Meaning[s] Across Languages*. Amsterdam. https://natasha-korotkova.github.io/files/esipova&korotkova2024-poqal-li-not-li.pdf.

Frana, Ilaria & Kyle Rawlins. 2019. Attitudes in discourse: Italian polar questions and the particle *mica. Semantics and Pragmatics* 12(16). 1–48. DOI: 10.3765/sp.12.16.

Gärtner, Hans-Martin & Beáta Gyuris. 2017. On delimiting the space of bias profiles for polar interrogatives. *Linguistische Berichte* 251. 293–316.

Gärtner, Hans-Martin & Beáta Gyuris. 2022. On the absence of propositional negation from Hungarian polar e-interrogatives. *Studia Linguistica* 76(3). 661–683. DOI: 10.1111/stul.12191.

Geist, Ljudmila & Sophie Repp. 2023. Responding to negative biased questions in Russian. In Petr Biskup, Marcel Börner (né Guhl), Olav Mueller-Reichau & Iuliia Shcherbina (eds.), *Advances in formal Slavic linguistics 2021* (Open Slavic Linguistics). Berlin: Language Science Press. DOI: 10.5281/zenodo.10123641.

Gonzalez, Aurore. 2023. Interrogative particles in polar questions: The view from Finnish and Turkish. *Glossa: a journal of general linguistics* 8(1). DOI: 10.16995/glossa.6487.

Goodhue, Daniel. 2018. *On asking and answering biased polar questions*. Montreal: McGill University. (Doctoral dissertation). http://id.loc.gov/vocabulary/iso639-2/eng.

Goodhue, Daniel. 2022. Isn't there more than one way to bias a polar question? *Natural Language Semantics* 30. 379–413. https://link.springer.com/article/10.1007/s11050-022-09198-2.

Grišina, Elena. 2005. Ustnaja reč v Nacionaľnom korpuse russkogo jazyka. *Nacionaľnyj korpus russkogo jazyka 2003—2005*. 94–100. https://ruscorpora.ru/new/sbornik2005/07grishina.pdf.

Grišina, Elena & Svetlana Savčuk. 2009. Korpus ustnyx tekstov v NKRJ (RNC): Sostav i struktura. *Nacionaľnyj korpus russkogo jazyka 2006—2008. Novyje rezuľtaty i perspektivy*. 129–149. https://ruscorpora.ru/new/sbornik2008/07.pdf.

Groenendijk, Jeroen Antonius Gerardus & Martin Johan Bastiaan Stokhof. 1984. *Studies on the semantics of questions and the pragmatics of answers*. Amsterdam: UvA. (Doctoral dissertation).

Gunlogson, Christine. 2002. Declarative questions. *Semantics and Linguistic Theory* 12. 124–143. DOI: 10.3765/salt.v12i0.2860.

Gyuris, Beáta. 2019. Thoughts on the semantics and pragmatics of rising declaratives in English and of their Hungarian counterparts. In Gyuris Beáta, Mády Katalin & Recski Gábor (eds.), *K + K = 120: Papers dedicated to L. Kálmán & A. Kornai on the occasion of their 60th birthdays*, 247–280. Budapest: MTA Nyelvtudományi Intézet. DOI: 10.18135/kk120.2019.

Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of Language* 10(1). 41–53. https://www.jstor.org/stable/25000703.

Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1(1). 3–44. https://www.jstor.org/stable/25000027.

King, Tracy Holloway. 1994. Focus in Russian yes-no questions. *Journal of Slavic Linguistics* 2(1). 92–120. https://www.jstor.org/stable/24599026.

Kopřivová, Marie, Zuzana Laubeová, David Lukeš, Petra Poukarová & Marie Škarpová. 2020. *ORTOFON v2: Korpus neformální mluvené češtiny s víceúrovňovým přepisem*. Praha: Ústav Českého národního korpusu FF UK. https://wiki.korpus.cz/doku.php/en:cnk:ortofon.

Korotkova, Natasha. 2023. Conversational dynamics of Russian questions with *razve*. In Maria Onoeva, Anna Staňková & Radek Šimík (eds.), *Proceedings of Sinn und Bedeutung 27*. Prague: Institute of Czech Language & Linguistic Theory, Faculty of Arts, Charles University.

Krifka, Manfred. 2011. Questions. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 2, chap. 66, 1742–1785. Berlin: Walter de Gruyter.

Krifka, Manfred. 2015. Bias in commitment space semantics: Declarative questions, negated questions, and question tags. *Semantics and Linguistic Theory* 25. 328–345.

Ladd, D. Robert. 1981. A first look at the semantics and pragmatics of negative questions and tag questions. In Robert A. Hendrick, Carrie S. Masek & Mary Frances Miller (eds.), *Papers from the seventeenth regional meeting of the Chicago Linguistic Society*, 164–171. http://www.lel.ed.ac.uk/~bob/PAPERS/CLS1981.pdf.

Machálek, Tomáš. 2014. *KonText: Application for working with language corpora*. Praha: Faculty of Arts, Charles University. https://kontext.korpus.cz/.

Meyer, Roland & Ina Mleinek. 2006. How prosody signals force and focus: A study of pitch accents in Russian yes-no questions. *Journal of Pragmatics* 38(10). 1615–1635. DOI: 10.1016/j.pragma.2005.05.011.

Palková, Zdena. 1994. *Fonetika a fonologie češtiny: S obecným úvodem do problematiky oboru*. cs. Google-Books-ID: 1V9iAAAAMAAJ. Karolinum.

Rathcke, Tamara. 2006. A perceptual study on Russian questions and statements. *AIPUK* 37. 51–62. https://www.ipds.uni-kiel.de/pub_exx/aipuk/aipuk_37/37_5_Rathcke.pdf.

Repp, Sophie. 2013. Common ground management: Modal particles, illocutionary negation and verum. In Daniel Gutzmann & Hans-Martin Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning*, 231–274. Leiden: Brill. DOI: 10.1163/9789004183988_008.

Restan, Per. 1969. *Sintaksis voprositel'nogo predloženija: Obščij vopros.* Oslo: Universitetsforlaget.

Romero, Maribel & Chung-Hye Han. 2004. On negative *yes/no* questions. *Linguistics and Philosophy* 27. 609–658. DOI: 10.1023/B:LING.0000033850.15705.94.

Rudin, Deniz. 2022. Intonational commitments. *Journal of Semantics* 39(2). 339–383. DOI: 10.1093/jos/ffac002.

Schwabe, Kerstin. 2004. The particle *li* and the left periphery of Slavic yes/no interrogatives. In Horst Lohnstein & Susanne Trissler (eds.), *The syntax and semantics of the left periphery*, 385–430. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110912111.385.

Šimík, Radek. Forthcoming. Polar question semantics and bias: Lessons from Slavic/ Czech. In Berit Gehrke & Radek Šimík (eds.), *Topics in the semantics of slavic languages* (Open Slavic Linguistics). Berlin: Language Science Press. https://www.researchgate.net/publication/373494477_Polar_question_semantics_and_bias_Lessons_from_SlavicCzech.

Staňková, Anna. 2023. *The expression of speaker's bias in Czech polar questions.* Prague: Charles University. (MA thesis). https://hdl.handle.net/20.500.11956/182788.

Štícha, František. 1995. Otázky predikátové: Inference, implicitnost a explicitní výrazy ilokučních funkcí I. *Slovo a slovesnost* 56(2). 98–109. http://sas.ujc.cas.cz/archiv.php?lang=en&art=3625.

Sudo, Yasutada. 2013. Biased polar questions in English and Japanese. *Current Research in the Semantics/Pragmatics Interface* 28. 275–295. DOI: 10.1163/9789004183988_009.

Chapter 15

Dual preservation in Slovenian: The verb supports the noun in semi-spontaneous production

Matic Pavlič[a], Artur Stepanov[b] & Penka Stateva[b]

[a]Faculty of Education – University of Ljubljana [b]Center for Cognitive Science of Language – University of Nova Gorica

We investigated the use of dual morphology on nouns in six Slovenian dialects which either completely or only partially preserved dual marking on the verb in a language-production experiment using a picture description task with a pre-set vocabulary. We compared our results against the typology of Slovenian dialects with respect to preserving the dual feature, as presented in Jakop (2008), and based on data from the Slovenian Linguistic Atlas (collected 1946–1999). We found that the use of the targeted dual of a noun in Slovenian dialects is influenced by the dual of a verb via agreement. More specifically, the higher rate of preserved verb dual forms are associated with higher use of the dual and lower use of the plural in the subject – but not in the object. Greater use of the dual in the subject does not affect greater use of the dual in the object – although the nominative and accusative forms are identical in the masculine. These findings, for the first time, experimentally confirmed the previously suspected (Tesnière 1925) supporting role of verb agreement on preservation of the dual in Slovenian.

1 Introduction

This paper presents an experiment that seeks to identify and document the diachronic process of the loss of dual morphology in selected Slovenian dialects against Jakop's (2008) typological observations that are based on the Slovenian Linguistic Atlas (collected 1946–1999). To this end, we compared the proportion

of false plural forms instead of target dual ones, on the nouns as well as the verb in unmarked SVO sentences. We further ask whether the loss of dual morphology in nouns correlates with their grammatical function or, in other words, with the structural position of subject vs. object. To this end, we also compared the proportion of false plural forms instead of target dual as a function of the structural position of the noun in question. Our experimental results confirmed the earlier assumption that the dialects in which the dual in verbs is more preserved also preserve more dual morphology in (subject) nouns than the dialects in which the dual in verbs is less preserved. The subject-object asymmetry in dual preservation is attributed to the fact that subject-verb agreement facilitates the preservation of dual morphology in subject nouns.

1.1 Grammatical number

The vast majority of languages that encode grammatical number on a noun distinguish between singular and plural (e.g. English in Table 1), but there are also languages with a three-part distinction (e.g. Slovenian and some other Indo-European, Semitic, Austronesian, and South American languages). On the other hand, languages with a four- or five-part distinction are extremely rare.

Table 1: Number on English, Arabic and Slovenian noun for 'student'.

	English	Arabic	Slovenian
Singular	student-Ø	taalib-Ø	študent-Ø
Dual	/	taalib-een	študent-a
Plural	student-s	taalib-iin	študent-i

The dual is marked compared to the plural with respect to several criteria: children acquire the dual later than the plural (Ravid & Hayek 2003), having the dual in a language entails having the plural (Greenberg 1966),[1] and the use of the dual tends to decrease in diachronic change (Corbett 2000). We can also add the morphological criterion of markedness, as mentioned in Greenberg's Universal number 35 (1966: 94):

> There is no language in which the plural does not have some nonzero allomorphs, whereas there are languages in which the singular is expressed only by zero. The dual and the trial are almost never expressed only by zero.

---

[1]"No language has a trial number unless it has a dual. No language has a dual unless it has a plural." (Greenberg 1966: 94).

The markedness of a number can trigger changes that lead to neutralization, i.e. unification of the marked grammatical number with the less marked one (see, e.g. the *Markedness-triggered impoverishment hypothesis*, Nevins 2011; the *Morphosyntactic Feature Economy hypothesis*, Slobodchikoff 2019; or the *Diachronic model of the loss of dual in the context of minimalist syntax*, Stepanov & Stateva 2018). Moreover, we know from research on language acquisition under suboptimal conditions (i.e., through quantitatively and/or qualitatively limited contact with the first language) that grammatical number (and agreement processes in general) is particularly volatile or susceptible to change (Polinsky 2018), e.g. paradigm simplification (Berdicevskis & Semenuks 2022). However, a factor in the gradual loss of dual forms could also be the interaction or loss of other grammatical categories, e.g. the neutralization of certain inflectional forms within a single grammatical number, which would intuitively lead speakers to retain only the grammatical number that differentiates between multiple inflectional forms (Ivanov 1983). Tesnière (1925), on the basis of diachronic data from Indo-European languages, made generalizations regarding the order of pluralization of dual forms, according to: **inflection** (locative > genitive > dative > nominative/ accusative), **gender** (feminine > neuter > masculine), **part of speech** (adjective > demonstrative > noun > numeral > personal pronoun), **grammatical function** (object > subject).[2]

1.2 Dual in Slovenian

In languages with overtly expressed grammatical number, the latter may be encoded by a free or bound morpheme, a modification of the root, and/or by a substitute root. Finally, it may also be phonologically unexpressed – and in the latter case identifiable only via secondary marking, namely, via matched features on the dependent items because of agreement. In inflectional languages such as Slovenian, grammatical number is part of a formal system of agreement involving an agreement *target* and an agreement *controller* (Corbett 2000) either within a noun phrase (Toporišič 2000: 109) or within a tense/inflectional phrase (Toporišič 2000: 608). Thus, speakers of Slovenian determine the form of the demonstrative ('these'), numeral ('two') and adjective ('old') in relation to the form of the noun ('lorries') in a noun phrase like (1). Similarly, they determine the verbal form ('overtake') in relation to the form of the noun phrase ('lorry' or 'campers'), which serves as the subject in a sentence like (2a) or (2b).

---

[2]">" stands for "is followed by" on the diachronic trajectory.

(1) Ta   dva   stara tovornjaka
     these.DU two.DU old.DU lorries.DU
     'these two old lorries'

(2) a. Prikolice   prehiteva   TOVORNJAK.
         campers.F.PL.ACC overtake.SG lorry.M.SG.NOM
         'It is the campers that the lorry overtakes.'

     b. Prikolice   prehitevajo tovornjak.
         campers.F.PL.NOM overtake.PL lorry.M.SG.ACC
         'Campers overtake the lorry.'

Here, agreement serves as a crucial clue to determine the sentence's meaning, since both nouns are potential candidates for the agent thematic role. The ambiguity is rooted in the flexible word order in Slovenian and in the homophonous forms of nominative and accusative both in masculine singular and feminine plural. Thus, in example (2a), the addressee recognizes the singular noun 'lorry' as the subject on the basis of singular form of a verb, while in example (2b), the plural noun 'camper' is recognized as the subject – again based on the verbal features. As can be seen from examples (2a) and (2b), in Slovenian the grammatical categories of gender and number are encoded in a single ending, but there is both theoretical and experimental evidence that these are distinct features. In Slovenian, grammatical number must be expressed on the noun phrase, be it a noun (3a) or a personal pronoun (3b). This being said, the grammatical number is not morphologically expressed on the coordinated noun phrases involving proper names (3c) and, trivially, on the silent *pro* (3d); note that Slovenian is a *pro*-drop language. In the latter two examples, the number is reflected in the verbal features as an instance of agreement.

(3) a. Otroka špricata teto.
         kids.DU spray.DU aunt.SG
         'The kids spray an/the aunt.'

     b. Onadva špricata teto.
         they.DU spray.DU aunt.SG
         'They sprey an/the aunt.'

     c. Jan in   Rok špricata teto.
         Jan and Rok spray.DU aunt.SG
         'Jan and Rok spray an/the aunt.'

     d. *pro* špricata teto.
            spray.DU aunt.SG
         'They spray an/the aunt.'

In the Slovenian dialects as well as in the standard variety, agreement between the verb and the subject is obligatory (Toporišič 2000: 271): if the subject is marked for plural, the dual form of the verb is ungrammatical (4a); if the subject is marked for dual, the plural form of the verb is ungrammatical (4b).

(4) a. {Otroci / oni / Jan, Rok in Bor} {špricajo / *špricata} teto.
         kid.PL they.PL Jan Rok and Bor spray.PL spray.DU aunt.SG
         '{The kids / they / Jan, Rok and Bor} sprey a/the aunt.'

   b. {Otroka / onadva / Rok in Bor} {špricata / *špricajo} teto.
         kid.DU they.DU Rok and Bor spray.DU spray.PL aunt.SG
         '{The two kids / they two / Rok and Bor} sprey a/the aunt.'

1.3 Preservation of dual in Slovenian dialects

In examining the preservation of dual forms in Slovenian dialects, we follow the results of Jakop (2008), based on the material of the Slovenian Linguistic Atlas (SLA, Benedik 1999: 15). The questionnaire of SLA consisted of 870 numbered questions with sub-questions eliciting up to 2000 linguistic expressions from each informant. Mostly, the questions were expressions (nouns and verbs) in standard Slovenian, and the informants' task was to translate them into their dialect and, in some cases, to give the full paradigm of the expression (Benedik 1999: 15). Some expressions were elicited by picture naming or by asking specific questions about the target ("What is it …?").

The material was collected between 1946 and 1999 in 413 locations within the borders of present-day Slovenia and in neighboring countries with Slovenian minorities. Each location was represented by three informants (a man, a woman, and a child under 14). It was reported that speakers with the most authentic dialect language (children) and speakers with the most developed metalinguistic understanding (teachers) dominate, but more than a quarter of the total material contains no information about the informants (Kenda-Jež 2002: 154). The main shortcoming of the material is its non-homogeneity, which is due not only to the long years of collecting and the different collectors, but also to the different qualifications of the collectors, the imprecision of the questions, the methodology and the transcription. Therefore, from several hundred questions, Jakop (2008) selected only 10 questions useful for the study of the dual. Note, that answers to these questions are not available for all the 413 locations. Below, we present the results by giving the percentage of retained dual forms for three exemplar nouns (one per gender).

The dual noun forms in Slovenian dialects are most often preserved in the masculine gender (96%; *brat* 'brother', N_{loc}=275; Jakop 2008: 135).[3] Only half of the dialects preserve feminine dual forms (51%; *krava* 'cow', N_{loc}=324; Jakop 2008: 135–136). Note that neuter nouns are masculinized (and sometimes feminized) in 41% of the dialects (*okno* 'window', N_{loc}=237; Jakop 2008: 136), so that the neuter dual forms are retained only in 10% (*okno* 'window', N=237; Jakop 2008: 136). In addition, both masculine and feminine dual verb forms have been partially lost in Northern and Southern dialectal groups which leads Jakop (2008) to link dual loss to contacts with neighboring languages without dual (Italian and Croatian).[4] In these dialects speakers use plural forms instead of the dual to describe an event with a participant consisting of two entities (see Figure 1 left) – as in our hypothetical example (5a). According to our informal observations this is especially common with a coordinated noun phrase (5b), a third person plural pronoun (5c), or the silent personal pronoun *pro* (5d) as a subject.

(5) a. Otroci špricajo teto.
kid.3.PL sprey.3.PL aunt.SG
'The kids sprey a/the aunt.'

b. Jan in Rok špricajo   teto.
Jan and Rok sprey.3.PL aunt.SG
'Jan and Rok sprey a/the aunt.'

c. Oni   špricajo teto.
they.3.PL sprey.3.PL aunt.SG
'They sprey a/the aunt.'

d. *pro*$_i$ špricajo$_i$ teto.
sprey.3.PL aunt.SG
'They sprey a/the aunt.'

Finally, we should also note that the earliest evidence of Slovenian losing dual can be found in the first printed books (16th century), as reported by Derganc (2006), Jakop (2008) and Orel (2019), a.o. However, to date "*the geographical prevalence of the use of dual forms in Slovenian dialects has not decreased significantly – the dual is a productive and living category in Slovenian.*" (Jakop 2008: 145).

---

[3]The number given by N_{loc} represents the number of locations for which a response to a SLA question was available (out of 413 locations).

[4]It is not clear, though, why contacts with German on the North does not lead to the dual loss.

2 Experiment

Our goals were to (i) verify the pattern of usage of the dual marking on nouns relative to the (in)complete loss/retention of verbal dual marking in Slovenian dialects (as reported in Jakop's 2008 work) in a controlled production experiment using a larger and more diversified set of uniform materials and a unified procedure; (ii) document potential dynamics of any existing or ongoing loss of the dual morphology in these dialects or a subset thereof; and (iii) explore the distribution and potential loss of the dual morphology relative to the noun's grammatical function. The rationale for the third goal was that most of the previous studies of the loss of dual have focused on nouns and verbs largely independently. However, the existence of morphological manifestation of subject-verb agreement, but not object-verb agreement, in Slovenian suggests that if the loss of the dual is taking place, it may be affecting subjects and objects to a different extent because of the asymmetry regarding their respective association with the verb. If verbs retain the dual form this might reinforce the dual on the subject longer than on the object, because of this association, as has been previously noticed by Tesnière (1925). We were therefore interested in whether this asymmetry has a systematic character that could be detected in a production study.

We used a picture description task and further restricted the informants' utterances by furnishing them with three key words from which they had to form a transitive sentence. This way we were able to check whether their use of dual nouns was related, on the one hand, to the extent of preservation of the nominal morphology in Slovenian dialects according to the data of SLA and the analysis of Jakop (2008) and, on the other hand, to the noun's sentential function as subject or object. We decided to study only the masculine nouns as those are considered the most stable group of nouns in Slovenian dialects (see above) and in the Indo-European languages in general (Tesnière 1925). For the same reasons, we focused on the syntactic functions of the subject and object, which are in transitive sentences encoded by nominative and accusative case, respectively.

Since Slovenian dialects (with the possible exception of southwestern dialects and southern dialects) do not allow masculine noun phrase in subject position not to agree with a verb in number (Jakop 2008), we hypothesize, in line with previous research (Tesnière 1925), that subject-verb agreement supports the use of dual and possibly contributes to the preservation of dual morphology in nouns. In our experiment, a more preserved verbal dual morphology in a dialect would result in more dual forms in the subject compared to the object – while a less or non-preserved verbal dual morphology in a dialect would result in a balanced use of the dual in the subject and object.

2.1 Dialects and informants

Table 1 shows the dialects we selected for our research: according to Jakop (2008) all of them have preserved dual noun morphology in nominative/accusative masculine (A and B dialects) while the dual verb morphology is lost in Southern and Western dialects (B dialects) but retained in Pohorje mountains, Soča river and Upper Carniola dialects (A dialects). In selecting the dialects, we used Jakop's (2008) two maps, based on two lemmas from SLA: a regular noun (*brat* 'brother') and a regular verb (*delati* 'to work'). We selected only regular expressions, since irregular paradigms are often less affected in language change and consequently might give an inappropriate picture. We collected data from 140 adult self-reported native speakers of Slovenian (88 female, mean age=37,9, SD=11,4; median age=36) who participated in this experiment voluntarily (indicating online consent), anonymously, and for no material compensation. The participants all spoke the dialects under investigation as indicated in the pre-test demographic questionnaire, respective sample sizes per dialect are shown in the last column of Table 2. All informants had normal or corrected to normal vision and reported no history of neurological disorders. The informants that were not native speakers of selected dialects were excluded from the analysis. We also excluded informants that did not reach a 50% threshold of correctly producing 32 control trials. This led to exclusion of 47 participants. The data from the remaining 93 participants were subjected to analysis.

Table 2: Slovenian dialect with respect to dual preservation.

Dialects	Elicitation SLA	Verb *delati*	Noun *brat*	Number of informants
Pohorje mountains (A1)	1955–65	+	+	16
Soča river (A2)	1951	+	+	12
Upper Carniola (A3)	1959	+	+	28
Western (B1)	1954–82	−	+	20
Southern 1 & 2 (B2 & B3)	1957	−	+	17

2.2 Materials

We created 64 colour drawings with a resolution of 300 dpi and size 373x220 pixels (ratio 16:9). Each stimulus stood for a single transitive event involving two characters, the agent and the patient. Three boxes were vertically aligned along

the right margin of the picture; each contained a printed lexeme in standard Slovenian. Since we were not interested in word order, we arranged the lexemes from top to bottom so that they followed the unmarked word order in Slovenian, which is subject-verb-object (SVO): the top and bottom lexemes were nouns (in the nominative singular), and the middle lexeme was a verb (in the infinitive), as shown in Figure 1. Informants were instructed to record a sentence that would best describe the picture using these three lexemes: a verb to name the action, and nouns to name the characters in the event. Informants often replaced the lexemes with dialectal or dialectally pronounced expressions, which suggests that the task actually elicited the dialect rather than a superregional or even the standard language.
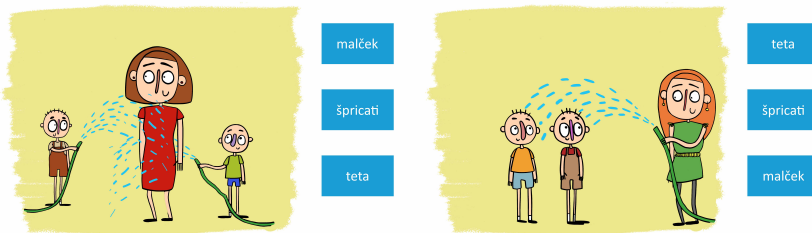
Figure 1: A target set of graphical stimuli.

We prepared 32 target stimuli with 16 different transitive verbs. Each of the 16 verbs from the target set was used twice, always with the same two characters, but with thematic roles reversed (so that each character served once as agent and once as patient). In the boxes, a noun was suggested for each of the characters in the sentence, one was feminine and one masculine. The masculine noun was intended to refer to a character consisting of two entities, i.e., it was intended to elicit a dual form. The feminine noun was intended to refer to a character consisting of either one or three entities, i.e., it was intended to produce a singular or plural form. A single informant had to produce only one sentence out of the set (i.e., half of the target sentences and only one version of each event), for the total of 16 target sentences out of 32. In addition, we prepared 32 control stimuli containing 32 different transitive verbs and 64 different nouns that were counterbalanced for gender (½ masculine and ½ feminine), number (½ singular and ½ plural), and sentence function (½ subject and ½ object). All nouns were repeated exactly once in the control stimuli. Thus, all informants saw 16 target trials and 32 control trials, based on which they recorded 48 SVO-sentences with 48 transitive verbs and 96 gender-balanced nouns.

2.3 Procedure

The experiment was conducted in the online environment Ibex Farm (Drummond 2021), enhanced with the PennController module (Zehr & Schwarz 2018). It generated quasi-random trials for each informant with at least one control trial between the two target trials. The informants conducted the experiment with their own equipment at a location of their choice, but were specifically instructed to do so in a quiet environment and to use their dialect and no formal or standardized language. Prior to the experiment, the informants gave informed consent and completed a brief demographic questionnaire. In the practice session, the informants were instructed to count to five and do a practice item to familiarize themselves with the instructions and stimuli and to learn how to turn the recording on and off. They were then able to play back the recording of the counting and practice item to check the function of the microphone, the volume, and the clarity of their own speech. In the experimental part the informants saw pictures one by one on the computer screen. Each picture appeared on the computer screen at the same time as the keywords. The informants had to put the words together as quickly as possible to form a sentence describing the picture, pronounce the sentence, and record it. The entire experiment lasted between 20 and 25 minutes.

2.4 Transcription, data cleaning and analysis

The recordings were manually transcribed into standard Slovenian, and the grammatical number was coded for each target noun. Prior to statistical analysis we excluded: incomplete recordings (due to premature termination of the recording); incomprehensible recordings or parts of recordings; recordings in which informants used a verb with a non-target sentence structure or kept the verb in the infinitive; recordings in which informants used a noun in a non-target gender, non-target case, non-target thematic role, or non-target sentence function; recordings in which informants retained the target noun in the singular rather than using it in the dual/plural form, consistent with the picture. The remaining 1247 target nouns (i.e., those denoting a character consisting of two entities) were statistically analyzed.

3 Results

3.1 The false plurals on nouns with respect to the dialect

First we checked for number mismatches but did not register a single case in which the subject is dual and the verb is plural – or the subject is plural and the verb is dual. Next, we counted the number of false plurals on nouns in each of the selected dialects to check for consistency between the two dialectal groups (with preserved dual on verbs and with less preserved dual on verbs). The total number of false plurals instead of the expected dual was 152 or 12% of the total data points. Table 3 summarises the results.

Table 3: The frequency of dual and false plural noun forms.

Dialect	more preserved dual on verb						less preserved dual on verb			
	Pohorje (A1)		Soča (A2)		Upper (A3)		Western (B1)		Southern (B2 & B3)	
Dual	216	(95.6%)	90	(75.6%)	375	(94.9%)	220	(83.3%)	194	(79.8%)
Plural	10	(04.4%)	29	(24.4%)	20	(05.1%)	44	(16.7%)	49	(20.2%)

The ratio of false plurals on the nouns in the Western (B1) and Southern (B2 and B3) dialects with less preserved dual on verbs did not differ significantly ($\chi^2(1) = 0.813$, $p = 0.367$). On the other hand, the three dialects with preserved dual on the verb were not homogeneous: the number of false plurals on the nouns in the Soča River dialect were significantly greater than in both the Pohorje ($\chi^2(1) = 28.971$; $p < 0.0001$) and Upper Carniola ($\chi^2(1) = 37.317$; $p < 0.0001$) dialects. The Pohorje and Upper Carniola dialects did not significantly differ in the ratio of false plurals ($\chi^2(1) = 0.026$; $p = 0.87$). These results confirm Jakop's (2008) findings regarding the preservation of noun dual morphology for all dialects except the Soča River dialect. The situation in the Soča River dialect is likely to have changed in 70 years since the data for the Slovenian Linguistic Atlas for the relevant dialects were collected. It should be noted that the Soča River dialect is spoken near the Slovenian-Italian border in a valley that extends into the area where the Western dialect is spoken. We hypothesise that in recent decades it has become easier for speakers of the Soča dialect to commute to neighboring regions and to establish and maintain contacts with speakers of the Italian language and the Western Slovenian dialect, which has led to increased contact and influence of these varieties (with less preserved dual) on their language.

3.2 The effect of the sentential function of a noun

We found a strong correlation between the preservation of verbal dual morphology and the use of dual/plural forms in the target noun with respect to its syntactic position, confirmed statistically by the χ^2-test and the Cramer coefficient V. For the subject position, Cramer's V was close to 1, implying a near perfect correlation between noun and verb forms. In contrast, Cramer's V for the object position was close to 0, indicating an almost complete lack of association between the noun and verb forms. This suggests that the choice of noun form does indeed depend on its syntactic function, as predicted, and that this dependence is strong. The results are summarized in Table 4.

Table 4: The frequency of dual and false plural by sentential function.

| Verbal morphology | Elicited noun forms | | | |
| | Subject | | Object | |
	dual	plural	dual	plural
dual preserved	1175	3	1091	87
dual less preserved	7	62	66	3
$\chi^2(1)$-test		1041.1		0.50
p-value		<0.0001		0.47
Cramer's V		0.9216		0.0268

To better understand the difference in sentential function, we evaluated this correlation separately for dialects with the preserved dual on the verb and dialects with the less preserved dual on the verb, as shown in Table 5.[5] In the dialect with the preserved verbal dual morphology, the difference between the number of plural forms instead of dual forms in the subject and in the object is statistically significant, while in the dialect with less preserved verbal dual morphology, the difference between the number of plural forms instead of dual forms in the subject and in the object is not statistically significant. As for the subject position, the difference between the number of plural forms instead of dual forms in the dialects with preserved dual and less preserved dual is statistically significant, while for the object position, the difference between the number of plural forms instead of dual forms in the dialects with preserved dual and less preserved dual is not statistically significant.

---

[5]The first number in the cell shows the number of occurrences of the false plural compared to the total number of target nouns (second number). The third number is the percentage.

Table 5: Analyzed target nouns by sentence function and dialect group.

			noun		
			subject	object	
			p<0.025		
verb	+dual	p<0.0001	15/353 (04%)	45/386 (12%)	p>0.05
	−dual		50/260 (19%)	42/248 (17%)	
			p>0.05		

Finally, we modelled the results. We added the target response (the dual of a noun referring to two entities) to the model as a reference value and used it to estimate the probability of the non-target response (the plural of a noun referring to two entities). Because the outcome was a categorical variable (either dual [1] or plural [0]), we used a mixed-effects logistic linear model (Jaeger 2008, Winter 2020) for binary outcomes via the *glmer* function in the *lme4* package in version 4.0.2 of the open-source R computing environment (R Core Team 2020) to test for dependence on a linear combination of independent predictor variables, while accounting for possible random noise. The independent variables in our model were syntactic position (noun in subject or object position) and dialect type (dialects with preserved dual on verb or dialects with fewer preserved dual on verb). Since we could not rule out the possibility that the informants' responses to the various conditions depended on the dialect they spoke, we also included in the model a test for the interaction between syntactic position and dialect type. The quality of the statistical model (i.e., the degree of agreement between the measured values and the values expected within the model) was validated using the Akaike information criterion: Compared with alternative models containing the same independent and dependent variables, the model with the selected function fitted our measured values best. Confidence intervals (CI) and p-values were calculated using the Wald test. The model results are in Table 6.

The statistically significant odds ratio (OR) for a non-target response in the subject position in the noun condition for dialect group with preserved dual of the verb is 0.01, which corresponds to a probability of 1%. The odds ratio for a non-target response in the subject position in the noun condition compared to the object position for both dialect groups combined is 6.31, which was found to be not statistically significant. At the same time, the effect of dialect group is statistically significant, as the odds ratio for a non-target response for dialects with a preserved dual verbal morphology (in both syntactic positions) is 0.11. This means that a non-target response is less likely for dialects with a preserved dual

Table 6: Statistical model.

Factors	OR	CI	p
(intercept)	0.01	0.00−00.07	<0.001
Function [object]	6.31	0.69−57.33	0.102
Group [+dual on verb]	0.11	0.01−00.99	0.049
Function [object] * Group [+dual on verb]	10.11	1.33−76.60	0.025
Number of informants		93	
Number of stimuli		32	
Data points (target nouns)		1247	

verbal morphology than for dialects with a less preserved dual verbal morphology (10% vs. 90% odds ratio for a non-target response). The interaction of dialect group and syntactic position is also statistically significant: a likelihood ratio of 10.11 shows that non-target responses are statistically significantly more likely to occur for object position compared to subject position in dialects with preserved verb dual morphology, but not in dialects with less preserved verb dual morphology. This difference is also evident in Figure 2.
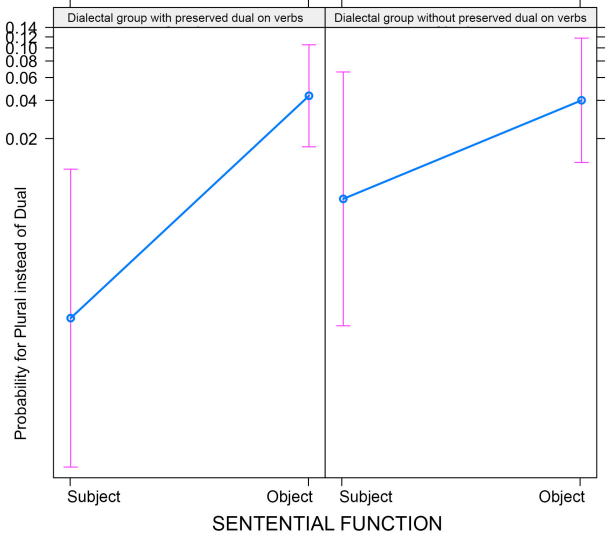
Figure 2: Comparison of the probability of false plural when syntactic position is crossed with dialect type.

To further investigate this interaction, we performed pairwise comparisons with the *emmeans* package in R, taking into account Tukey's adjustments for multiple comparisons. These comparisons reconfirmed that in dialects with preserved dual morphology of the verb, the probability of a target response is statistically significantly higher for a noun in subject position than for a noun in object position ($\beta = -4.15, \mathrm{SE} = 1.37, z = -3.03, p = 0.01$). In other words, informants were more likely to use a dual form for the target noun in subject position than for the target noun in object position. This was not the case in dialects with a less preserved dual morphology of the verb, where the likelihood of a dual form did not differ significantly with respect to the syntactic position of the noun ($\beta = 1.84, \mathrm{SE} = 1.12, z = 1.63, p = 0.36$). Thus, informants in this dialect group were equally likely to substitute the dual for plural in any syntactic position.

4 Discussion

In this sentence production experiment the realization of the dual in a noun was related to the sentential function of this noun. In dialects with preserved dual verb forms, the use of a noun in the dual to refer to two entities depends on its syntactic position (in subject position it is higher than in object position). This is not the case in dialects with less preserved dual verb forms. On this basis, we can both predict and explain why pluralization of the dual in Slovenian dialects occurs first in contexts where the noun is not morphologically marked for the dual (in the case of coordinated noun phrases or silent personal pronouns).

It should be noted that the pattern of subject-verb agreement is extremely strong in our sample, as we did not register a single case in which the subject is dual and the verb is plural – or the subject is plural and the verb is dual. Thus, when the dual is dropped in favor of the plural, it is dropped from both the subject and the verb. Second, we noted that in the dialects that, according to Jakop (2008) and based on the Slovenian Language Atlas, have lost the dual morphology in the verb, the dual is still used in the verbs, although less than in other dialects.

Next, considering that the dual forms of nominative and accusative masculine are homophonous, one might expect a "transitive" supporting effect of subject-verb agreement, in the sense that greater use of the dual in the nominative (due to agreement with the verb) would lead to greater use of the dual in the accusative (which is identical in form to the nominative). Our results show no such effect and suggest that the masculine nominative and accusative forms in the dual, although identical, are treated morphologically independently.

Finally, the results are interesting from a theoretical point of view. How can the diachronic changes (i.e., loss of the dual) in certain Slovenian dialects be explained in such a framework, especially when at the same time mismatch or split agreement (where the verb and the subject carry different features) are not supported? We tentatively hypothesise that homophony is an intermediate step in the gradual loss of the dual. First, the plural form is reinterpreted to refer not only to "more than two entities" but also to "more than one entity".[6] At this stage, dedicated dual forms may coexist with neutralized plural forms until the neutralized forms predominate and, as a result, the paradigm becomes partially homophonous until, finally, the dual forms are completely abandoned in favor of the plural forms. Under the assumption that diachronic change follows a principle of Economy (e.g. Martinet 1955), the "more than one entity" interpretation of the plural wins over the "more than two entities" interpretation since the former refers to a superset of possible referents in comparison to the latter and thus has a wider descriptive potential in comparison to the latter. According to this hypothesis, either the noun or the verb may be neutralized first, but the not-yet-neutralized form would support the dedicated interpretation of the homophonous form, even if the forms are no longer distinct.

5 Conclusion

Using a picture description task based on three given key words, we tested if the actual use of the dual form of a noun is related to the preservation of dual forms in Slovenian dialects and the sentential function of that noun. There was no split-matching (i.e., plural subject and dual verb or vice versa) in the sample studied. The results showed that speakers use the dual form of the noun in subject position more often than in object position. In the dialects with better preserved dual forms of the verb, the number of non-target plural nouns in subject position was lower than in object position because this dialect group preserved the dual forms of the verb and verb agreement seems to play a supporting role in preserving the dual morphology of the subject. In the dialect group where the dual forms of the verb were less preserved, the number of dual forms of the noun did not depend on its syntactic position because the dual morphology of the verb was lost.

---

[6]This step can conceivably be couched within the pragmatic theory of plural (cf. Sauerland 2008, Spector 2007), according to which the pragmatically enriched interpretation of plural results from enriching its lexical meaning associated with one or many objects with relevant pragmatic inferences. In languages featuring singular and plural number marking, this is an anti-singularity inference. In languages featuring singular, dual, and plural number marking, both an anti-singularity and anti-duality inferences strengthen the meaning of the plural.

Acknowledgments

Abbreviations

SG	singular	NOM	nominative
DU	dual	ACC	accusative
PL	plural	SVO	subject-verb-object
M	masculine	SLA	Slovenian Linguistic Atlas
F	feminine	CI	confidence interval
3	third person	OR	odds ratio

References

Benedik, Francka. 1999. *Vodnik po zbirki narečnega gradiva za slovenski lingvistični atlas (SLA)*. Ljubljana: Založba ZRC.

Berdicevskis, Aleksandrs & Arturs Semenuks. 2022. Imperfect language learning reduces morphological overspecification: Experimental evidence. *PLoS ONE* 17(1). e0262876. DOI: 10.1371/journal.pone.0262876.

Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.

Derganc, Aleksandra. 2006. Nekatere značilnosti dvojine v slovenščini. *Slavistična revija* 54(1). 415–434.

Drummond, Alex. 2021. *Ibex farm*. https://spellout.net/ibexfarm.

Greenberg, Joseph H. 1966. *Language universals*. Mouton The Hague.

Ivanov, V. 1983. *Istoričeskaja grammatika russkogo jazyka*. Moscow: Prosveščenie.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *J Mem Lang.* 59(4). 434–446. DOI: 10.1016/j.jml.2007.11.007.

Jakop, Tjaša. 2008. *Dvojina v slovenskih narečjih*. Ljubljana: Založba ZRC.

Kenda-Jež, Karmen. 2002. Model idealnega govorca v slovenskih dialektoloških raziskavah. In Marko Jesenšek, Bernard Rajh & Zinka Zorko (eds.), *Med dialektologijo in zgodovino slovenskega jezika*, 150–165. Maribor: Slavistično društvo. https://press.um.si/index.php/ump/catalog/book/41.

Martinet, André. 1955. *Économie des changements phonétiques: Traité de phonologie diachronique*. Bern: A. Francke.

Nevins, Andrew. 2011. Marked targets versus marked triggers and impoverishment of the dual. *Linguistic Inquiry* 42(3). 413–444. DOI: 10.1162/LING_a_00052.

Orel, Irena. 2019. Ženske dvojinske glagolske oblike v starejšem slovenskem knjižnem jeziku. *Slavistična Revija* 67(2). 273–280. https://srl.si/ojs/srl/article/view/2019-2-1-15.

Polinsky, Maria. 2018. *Heritage languages and their speakers*. Cambridge: University Press.

R Core Team. 2020. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Ravid, David & Lubna Hayek. 2003. Learning about different ways of expressing number in the development of Palestinian Arabic. *First Language* 23(1). 41–63. DOI: 10.1177/0142723703023001003.

Sauerland, Uli. 2008. Implicated presuppositions. In Anita Steube (ed.), *Sentence and context: Language, context, and cognition*, 581–600. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110209303.4.581.

Slobodchikoff, Tatyana G. 2019. *The evolution of the Slavic dual: A biolinguistic perspective*. Lanham: Rowman & Littlefield.

Spector, Benjamin. 2007. Aspects of the pragmatics of plural morphology: On higher-order implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and implicature in compositional semantics*, 243–281. London: Palgrave Macmillan. DOI: 10.1057/9780230210752_9.

Stepanov, Arthur & Penka Stateva. 2018. Countability, agreement and the loss of the dual in russian. *Journal of Linguistics* 54(4). 779–821. DOI: 10.1017/S0022226718000130.

Tesnière, Lucien. 1925. *Les formes du duel en slovène*. Paris: Ancienne Honoré.

Toporišič, Jože. 2000. *Slovenska slovnica*. Maribor: Obzorja.

Winter, Bodo. 2020. *Statistics for linguists: An introduction using R*. New York: Routledge.

Zehr, Jérémy & Florian Schwarz. 2018. *PennController for internet based experiments (IBEX)*. https://www.pcibex.net/.

Chapter 16

Word prosodic structure and vowel reduction in Moscow and Perm Russian

Margje Post

University of Bergen

Central Standard Russian is well-known for its vowel reduction in two degrees: the immediately pretonic vowel is much less reduced than vowels in other unstressed positions, both in quality and in quantity, at least when the allophone is a low vowel. This two-degree reduction is expressed clearly in speech from Central Russia, but earlier studies suggest a smaller difference between the degrees in non-central areas. We measured vowel duration and quality of unstressed /a, o/ in two modern urban Russian varieties: in read speech from 26 adolescents in Moscow (Central Russia) and Perm (Ural region). The Moscow speakers make a sharp distinction between the two degrees in both quantity (duration) and quality (F1), but we found only small, not statistically significant differences in Perm. Perm speech might lack phonological two-degree reduction altogether, in which case two-degree reduction is not a general feature of modern Russian urban speech.

1 Introduction: Unstressed vowels in East Slavic

1.1 Two-degree reduction in Central Standard Russian

Standard Russian is known for its typologically unusual word prosodic structure and vowel reduction pattern. Unstressed vowels are partly neutralized: most notably, /o/ merges with /a/. In most languages, vowel merger is combined with phonetic reduction, but in Standard Russian it is combined with an unusually prominent vowel in first pretonic position – i.e. the vowel immediately preceding the stressed syllable. This vowel can be very long and lack reduction in quality, especially in Central Standard Russian. The term vowel *reduction* can therefore

be misleading, as Dubina (2012) and Iosad (2012) remark, although the term is often used not in a purely phonetic, but in a more abstract, phonological sense. In the current paper, REDUCTION will mostly refer to phonetics, but Section 4.4 will discuss the phonological status of the two degrees of reduction in varieties of Russian. I refer to CONTEMPORARY CENTRAL STANDARD RUSSIAN (CSR), following Iosad (2012), because almost all studies on Russian unstressed vowels in Standard Russian are based on speakers from Moscow or Saint Petersburg, but I wish to point out that this pronunciation might differ from other, locally coloured varieties of Standard Russian.

In CSR, the immediately pretonic syllable forms a salient contrast, together with the stressed syllable, with unstressed syllables in other, weak positions, which are heavily reduced, both in quality and in quantity (Zlatoustova 1981, Kodzasov 1999), especially when the allophone is a low vowel. One could say that the first pretonic and the stressed syllable together form a nucleus in the word (Kodzasov 1999, *inter alia*), i.e. a strong centre that is opposed to a weak periphery of the word (Kasatkina 1996), or that word stress is realized over two syllables (cf. Dubina 2012: xiii).

The unusual prominence of the vowel in first pretonic position means that effectively, CSR has two degrees of vowel reduction: a moderate degree for the first pretonic vowel and a radical degree of reduction for unstressed vowels in other positions (e.g. Crosswhite 2000).[1] The term TWO-DEGREE REDUCTION usually refers to qualitative reduction, but the main distinction might be in duration. The phonological status of the difference in degrees of reduction is discussed in surprisingly few studies, as remarked by Mołczanow (2015: 133). Barnes (2006) and Iosad (2012), two of the exceptions, argue that only the difference in duration is phonological, the difference in quality being merely an effect of phonetic implementation rules; cf. Section 4.4 below. One should bear in mind that two-degree reduction accounts first and foremost for the vowels /a/ and /o/ in the position after non-palatalized consonants.[2] When unstressed, they merge, but they merge

---

[1]Moderate reduction is also found in other positions than the first pretonic if they allow long vowel durations, notably, in onsetless syllables and, optionally, in phrase-final open syllables (e.g. Barnes 2006, Iosad 2012). Based on a purely phonetic study, Kuznecov (1997) discerns an additional, third degree of reduction in duration of /a, o/ after non-palatalized consonants, which is found in posttonic vowels, but only one degree of qualitative reduction, since these vowels were not reduced in quality in first pretonic position.

[2]The durational differences between the syllables relative to stress are not found for all vowels, and they are highest for non-high vowels, i.e., for unstressed /a, o/ after non-palatalized consonants (Bondarko et al. 1966, Zlatoustova 1981, Padgett & Tabain 2005). A qualitative distinction into two different allophones according to position relative to stress is mentioned only for these vowels. In the position after palatalized consonants, all vowels except /u/ are produced as front vowels.

into two different allophones, depending on their position in the word. In CSR, the word /molo'ko/ 'milk' is pronounced as [məlɐˈko], with the first vowel, the /o/ in second pretonic (antepretonic) position, being substantially shorter than the prominent next vowel, in first pretonic position, which can have the same quality as a stressed /a/ (Kuznecov 1997 on speakers from Saint Petersburg; Knjazev 2006 on speakers from Moscow) and may be as long as or even longer than the stressed syllable (in non-focus position; Knjazev 2006).[3] In addition, it is often singled out by a local high tone, at least by Moscow speakers (on pitch accented words in declaratives, cf. Kasatkina 2005). Dubina (2012) and Mołczanow (2015) connect the heavy first pretonic syllable in Russian and Belarusian to an abstract phonological high tone that is associated with the first pretonic syllable, comparable to the anticipatory tone spreading in Bosnian/Croatian/Serbian (Dubina 2012: 175). In CSR, however, this high tone need not surface in the phonetic realization of the word (unlike, probably, in some conservative Russian dialects; see Section 1.2.2 below).

The distribution of syllable prominence in the word in CSR is captured by Potebnja's (1866) formula 112'311, where 1 means radical reduction, 2 means moderate reduction and '3 stands for the unreduced stressed syllable. Empirical studies have confirmed that Potebnja's 112'311 formula corresponds to a three-way distinction in duration in CSR (Bondarko et al. 1966, Zlatoustova 1981, Kuznecov 1997, Barnes 2006, Knjazev 2006).

The qualitative reduction of unstressed /a, o/ after non-palatalized consonants in two degrees to a low vowel or to *schwa* is part of the pronunciation standard (Avanesov 1984), which is taught to foreigners learning Russian.

Vowel reduction in Central Standard Russian has received due attention in the literature. Examples of phonological accounts of the two-degree reduction in CSR are Crosswhite (2000), Iosad (2012) and Mołczanow (2015). Empirical phonetic research on vowel reduction in CSR has been done by, among others, Bondarko et al. (1966), Zlatoustova (1981), Kuznecov (1997), Padgett & Tabain (2005), Barnes (2006), Kocharov et al. (2015). However, most existing acoustic studies are based on a very limited number of speakers, and many questions deserve more attention, among others, the relation between vowel quality, quantity and tone and

---

[3]Most older literature uses the symbol [ʌ] for the first degree reduction of /a, o/ after non-palatalized consonants, but for contemporary CSR, the symbol [a] or [ɐ] is more appropriate; cf. Kasatkina (2005) and Iosad (2012) for a discussion. This accounts for speakers from both Moscow and Saint Petersburg. Most differences in pronunciation between standard speakers from these cities have disappeared; cf. Verbickaja (1977) and Kuznecov (1997). Speakers from cities with a Northern Russian substrate use vowels further back (Kasatkina 2005).

the phonological status of two-degree reduction (cf. Barnes 2006, Bethin 2006, Mołczanow 2015, *inter alia*).

Much less is known about unstressed vowels and the word prosodic structure in other, non-central varieties of Russian (Section 1.2) and in the other East Slavic languages (Section 1.3). Contemporary Russian has little regional variation. Rural dialects show relatively small linguistic distance, compared to dialects of other European languages, and among today's urban population, geographically based differences are hardly present, but one can expect at least some variation in prosody between the regions (Grammatčikova et al. 2013, Post 2017). Therefore we decided to study the prosodic word in regional urban varieties of Russian.

1.2 Regional varieties of Russian

1.2.1 Quantitative reduction

Studies of traditional rural Russian dialects suggest that all varieties have two-degree reduction in duration, but to a different extent. Vysotskij (1973) measured vowel and consonant durations in words with the structure $CV_{-2}CV_{-1}'CV_0C$ with the vowels /a, o/ in a large number of Russian varieties. He discerned twelve different groups of rural dialects and three varieties of Russian spoken in Moscow, each with its own rhythmic structure. The first pretonic vowel – V_{-1} – was longer than the preceding vowel – V_{-2} – in all groups, but whereas many dialects in Central Russia combine unusually long first pretonic vowels with very short vowels in second pretonic position (Potebnja 1866, Vysotskij 1973, Al'muxamedova & Kul'šaripova 1980, *inter alia*), the difference between the two positions was much less pronounced in southern and northern parts of European Russia (Vysotskij 1973). Two-degree reduction in duration is possibly absent in some northern rural dialects, where both pretonic vowels had almost the same duration in Vysotskij's recordings (Vysotskij 1973).

Most Russians today do not speak a traditional rural dialect, however, but an urban variety with few local characteristics. An unusually long and low first pretonic [a] is typical for the old Moscow dialect and Moscow vernacular speech (Vysotskij 1973). To our knowledge, only two, preliminary, studies compare unstressed vowel durations in cities other than Moscow and Saint Petersburg (Erofeeva 2005, Grammatčikova et al. 2013). Both suggest that the geographical opposition between centre and periphery is retained in modern urban Russian, with a weaker difference in duration between the first and second pretonic vowels in non-central areas than in central European Russia. The number of speakers

and vowels measured was very low in these two studies.[4] More data are therefore needed to confirm their preliminary findings on non-Central modern urban Russian pronunciation.

1.2.2 Qualitative reduction, dissimilation and tone

Two-degree reduction in quality is not possible in traditional Northern dialects, for unstressed /o/ and /a/ do not merge in these dialects. Some Central-Russian dialects combine partial neutralization with strong durational two-degree reduction: The distinction between /a/ and /o/ is retained in the – prominent – first pretonic position, but they merge in other, radically reduced positions (cf. Avanesov & Bromlej 1986). This means that these dialects do distinguish two degrees in quality as well – with no reduction in first pretonic position and strong reduction (to *schwa*) in second pretonic position. A comparison of studies on qualitative reduction (e.g. Avanesov & Bromlej 1986) with those on quantity (such as Vysotskij 1973) suggests that the area with incomplete merger coincides with the area with very strong two-degree durational reduction, with uncommonly short second pretonic and extremely long first pretonic vowels.

In the remaining Central Russian and in all Southern Russian traditional dialects /o/ and /a/ are neutralized in all unstressed positions. In a large part of these neutralizing dialects (and in some neighbouring dialects in Belarus) the vowel reduction pattern is complicated by vowel dissimilation. Vowel dissimilation means that the quality and duration of the first pretonic vowel are dependent on the quality of the stressed vowel, for the two vowels must be different (Avanesov & Bromlej 1986, *inter alia*). Clear dissimilation tendencies in duration, with a length trade-off between the first pretonic and tonic vowel, are observed even in modern Moscow speech, which does not have vowel dissimilation in quality (Kasatkina 2005). For instance, the first pretonic low vowel [a] is longest before the high vowel [i], and shorter before a low vowel [a], and this difference is substantially larger than what can be accounted for by the intrinsic durational properties of high and low vowels (Zlatoustova 1981, Kasatkina 2005, Iosad 2012). The strong pretonic vowels were almost invariably marked by a local high tone in

---

[4]Erofeeva's (2005) study of vowel duration is based on 600 vowels from spontaneous speech by two male and two female speakers from Perm (city or region) with an audible local accent; Grammatčikova et al. (2013) used tokens from a read text that was read once by only 6 individual speakers, one from each city, who were assessed to speak Standard Russian.

some conservative central dialects of Russian and Belarusian (Broch 1916, Bethin 2006; *inter alia*).[5]

The unreduced, labialized production of unstressed /o/ – *okan'e* – has low social prestige (Andrews 1995) and is rare in contemporary urban Russian, especially in read speech (Verbickaja et al. 1984, Erofeeva 1993), but incomplete neutralization, with a distribution of vowel allophones for unstressed /a/ different from /o/, is more common (Erofeeva 1993 and, less so, Erofeeva 2005, on speech from Perm – a city in the Ural region with Northern Russian traits). Even more common is "non-normative" vowel reduction, i.e. a failure to distinguish two different allophones in the second and first pretonic position. This was reported to be frequent in all seven regional cities studied by Verbickaja et al. (1984). In a small empirical study, Erofeeva (2005) found substantially lower F1 values for unstressed /a/ and /o/ than for stressed /a/ in Perm, suggesting substantial qualitative reduction. Besides, the allophone in first pretonic position is reported to be further back in cities with a Northern dialect substrate than in CSR (Kasatkina 2005), suggesting low F2 values. No other empirical studies of vowel reduction patterns in regional urban Russian are known to us.

1.3 Belarusian and Ukrainian: quantitative, but no qualitative two-degree reduction

First pretonic prominence is a feature shared by all three East Slavic languages (Dubina 2012), but this accounts only for duration, not for vowel quality. Neither Standard Belarusian nor Ukrainian has two-degree reduction in quality, depending on the position of the vowel in relation to the stressed syllable. In Ukrainian, unstressed /o/ and /a/ do not merge at all, whereas in Standard Belarusian they merge into a low vowel in all unstressed positions (Černjavskij 2012, Dubina 2012). Besides, the durational distinction between the second and first pretonic vowels appears to be smaller in Standard Belarusian and Ukrainian than in Central Standard Russian; cf. Dubina (2012) on Belarusian, referring to an empirical study by Andreev (1984), and Łukaszewicz et al. (2022) on Ukrainian. The difference is very small in the latter study, but the word prosodic pattern in Ukrainian is complicated by iterative secondary stresses (Łukaszewicz & Mołczanow 2018).

---

[5]Note that Borise (2017) could not confirm a local high peak before a fall on the stressed syllable on the words with pretonic strengthening, which was earlier found in the Belarusian dialect she studied (the dialect of Aŭciuki; cf. Bethin 2006). Instead, she found a generally higher tone on both first pretonic and tonic syllable (compared to words without pretonic strengthening in this dialect with vowel dissimilation).

Besides, the study is based on speakers from Western Ukraine. The geographical distribution of strong first pretonic prominence in Russia suggests that the difference might be larger in Central and Eastern Ukraine.

The variation of vowel reduction patterns in the East Slavic languages and their dialects shows that first pretonic vowel prominence in duration in East Slavic need not cooccur with vowel neutralization and qualitative reduction in two degrees, in the way it does in Central Standard Russian. Vowel neutralization, first pretonic prominence and qualitative reduction are distinct processes (cf. Dubina 2012: 166), which interrelate in various ways.

1.4 Research question

We wanted to know how clearly the two-degree reduction in quality and quantity of Central Standard Russian is expressed in today's urban speech. More particularly, how different are the second and first pretonic vowels /a, o/ after non-palatalized consonants from each other, and from the tonic (stressed) vowel, in Russian regional speech, both in quantity (duration) and in quality (as expressed in vowel formants)?

The scarce existing literature on regional variation in two-degree reduction suggests that we would find a substantial difference between the second pretonic and the first pretonic vowel in Moscow speech, but a smaller difference in Russian speech from other regions. We chose to compare the speech of young urban Russians from Moscow, as a representative of Central European Russian, and Perm (Ural), representing non-central Russian.

2 Our study: Moscow vs. Perm

2.1 Participants

We recorded speech by young speakers in Moscow (central variety) and Perm (Ural region, non-central variety; cf. Erofeeva 2005). We chose Moscow because it is Russia's capital and because Moscow's speech, especially its vernacular speech, is known for its long and open realizations of /o/ and /a/ in first pretonic position. The city of Perm is situated in the Ural region on the border between European Russia and Siberia. Perm speech is known for its relatively strong local accent, with traits from Northern Russian (Erofeeva 2005). The quality and duration of the unstressed vowels are claimed to play an important role in its local colouring (Erofeeva et al. 2000). The phonetics of speech by Russians from Perm have been described extensively (e.g. Erofeeva 2005), but mostly auditorily. Acoustic studies measuring durations and formants are all but absent.

We chose to record both sexes, since gender differences are often found in other countries, with a tendency of young urban women speaking with less local colouring than men (Labov 2001). All participants (but one, whose data were left out) were raised in the city they were recorded (Perm or Moscow) and almost all had parents with a high educational level. We did not select our participants for speaking (perceived) Standard Russian or having a local accent. Therefore, the speech of our Moscow participants could differ from Central Standard Russian, and the speech of our adolescents from Perm from Erofeeva's (2005) results, which were based on speakers with perceived local colouring. We can expect some degree of local colouring for our mean values, but not a strong local vernacular accent, since we use a reading task read by pupils with highly educated parents made in a formal school setting (see Section 2.3).

2.2 Materials

The reading task consisted of 14 sentences with 4 target words under several prosodic conditions, read in the same order. The target words were *topotát'* 'to patter', *pokopát'* 'to dig (a little)' and *potakát'* 'to connive', three words that were also analysed in Vysotskij (1973), and *kopát'* 'to dig', with a $(CV_{-2})CV_{-1}'CV_0C$ structure with pretonic /o/ and /a/ after non-palatalized consonants.[6] We will call the vowels in second pretonic position a_{-2}, since they occur two syllables prior to the stressed syllable, the first pretonics a_{-1} and the tonic (stressed) vowels $á_0$. The symbols a_{-2} and a_{-1} stand for both /a/ and /o/, which merge in unstressed positions in Standard Russian. Like most previous studies, we confined ourselves to /a/ and /o/, since other vowels might not show reduction in two degrees (cf. footnote 2). The task was designed preliminary to measure durations rather than formants, which led to some limitations to what we can conclude from our formant data (caused by an uneven distribution of /a/ and /o/, which are surrounded by different plosive consonants leading to different consonant coarticulation effects; cf. Section 4.2 and Section 4.3 below).[7]

---

[6]Russian text, including Russian surnames, is transliterated using Comrie & Corbett's (1993) transliteration system.

[7]As remarked by one of the anonymous reviewers, the etymology of the pretonic vowels in *potakát'* and *topotát'* is not certain, let alone their phonemic analysis by the speakers, so the single unstressed /a/ in our target words might in fact represent not /a/, but /o/ for some of the speakers, in which case all pretonic vowels in the reading task are /o/. This is only of minor importance for our analysis, since unstressed /a/ and /o/ merge completely in Moscow speech and probably almost completely in read speech from Perm; cf. Section 4.2 below. Another limitation of our reading task is that *potakát'* and *topotát'* are infrequent words that may not have been recognized by all speakers and may have provoked unnatural pronunciations. We minimized this problem by leaving out the first occurrences of the target words.

For the present study we left out the 4 sentences with *kopat'*, for lacking vowels in second pretonic position, and the first occurrences of the target words, in citation form, and the sentence with *pokopat'* in initial position, to avoid hesitations and boundary phenomena. The remaining 6 sentences cover two prosodic conditions for each of the three target words:

(1)   utterance-medial position, carrying nuclear pitch accent (here, marked in small caps):
       Ja   POKOPAT' pošla.
       I.NOM dig.INF   go.F.SG.PST
       'I went to dig a little.'

(2)   utterance-medial position, not carrying nuclear pitch accent:
       Ja   topotat' uže   ne BUDU.
       I.NOM patter.INF already NEG 1SG.IPFV.FUT
       'I won't patter anymore.'

The position of the accents was not marked in the reading task, but all participants read sentence (1) with the target word carrying the last, nuclear accent, followed by an unaccented word. In sentence (2) the nuclear accent is carried by the last word of the sentence. The target word carries a prenuclear pitch accent in most of the renditions, but other accentuation patterns occur as well, including deaccentuation of *topotat'* following a contrastive accent on the preceding word *Ja.* The nuclear pitch accent can vary as well, since the speakers can choose list intonation, which has a falling-rising tune. This variation in accentual patterns ensures that our results are not restricted to one specific prosodic structure but have a broader validity. It can potentially have a small effect on the actual vowel durations.

2.3 Procedure

We recorded a total of 34 adolescents (born in 1996–2000, recorded in 2015), 10 girls and 9 boys in Perm and 7 girls and 8 boys in Moscow. The participants were recorded in quiet rooms at school using digital recorders and head-mounted microphones (Zoom H5 with Shure WH20 in Perm, Zoom H2 with Samson QV in Moscow, set at 44.1 kHz, 16-bit, .wav). The reading task was performed along with recordings of a range of other tasks for, or connected to, Benedikte Vardøy's PhD project on young Russians' perception of regional variation in Russian (cf. Vardøy 2021, Post & Andreeva 2023). The utterances were read from paper, only

once in Perm, but twice in Moscow. We used only the first iteration of the speakers in Moscow, in order to have comparable data, but their second reading gave us the possibility to replace unsuccessful first renditions.

Several speakers were excluded from the analysis. We left out a boy in Moscow who was raised elsewhere and a girl in Perm because of creaky voice. After the segmentation, six additional speakers were discarded, because they had gaps in the data (see next section).

Six utterances by 13 female and 13 male speakers were analysed statistically (Table 1).

Table 1: Number of speakers and tokens used for the statistical analyses

	speakers	female	male	word tokens	vowel tokens
Moscow	13	7	6	78	234
Perm	13	6	7	78	234
Total	26	13	13	156	468

2.4 Data analysis and statistics

The target vowels in the speech samples were segmented manually in Praat (Boersma & Weenink 2023) via visual inspection of the waveform and spectrogram. We used standard criteria, among others, vowel boundary labels were placed at zero crossings on the waveform close to the onset and offset of the vowel formants. A number of tokens was discarded, due to misreadings, long pauses or creaky voice. For the Moscow speakers, discarded productions were replaced by their second readings. The speakers with missing data were excluded from the analysis, which left us with 13 speakers from Moscow and 13 speakers from Perm (see Table 1).

The durations, F0, F1, F2 and F3 of the target vowels per speaker and location were extracted using Praat scripts. Vowel durations were log-transformed because of positive skewness. F1 and F2 were measured at the temporal midpoint in vocalic nuclei. Speaker-dependent standard normalization was applied to control for differences in formant values due to speaker identity and sex (Lobanov 1971). We used the software JMP 16.2.0 for statistical analysis. Linear mixed models (LMM) were fitted with the respective measure as dependent variable and DISTANCE-TO-STRESS with three factor levels (*0/1/2*), LOCATION with two factor levels (*Moscow/Perm*) and GENDER with two factor levels (*male/female*) as fixed

factors, as well as all their possible interactions. SPEAKER, WORD (*topotat'*/*pokopat'*/*pokatat'*) and POSITION-IN-UTTERANCE (*utterance-medial prenuclear*/*utterance-medial nuclear position*) were taken as random factors. Separate Tukey HSD post-hoc tests were carried out per variable, if appropriate. The confidence level was set at $\alpha = .05$.

3 Results

3.1 Vowel quantity: Duration

The statistical analysis on the durational data (log-transformed) shows a main effect of GENDER (F [1, 22] = 6.0601, $p < .02$) on the target vowel duration, with female speakers having somewhat longer overall vowel durations than their male peers (cf. Figure 1, light-coloured vs. dark-coloured bars), and, predictably, of DISTANCE-TO-STRESS (F [2, 431] = 815.4856, $p < .001$), with stressed vowels ($á_0$) being significantly longer than the vowels in the first pretonic syllable (a_{-1}), which in turn are longer than the vowels in the second pretonic syllable (a_{-2}).[8] Not significant proved LOCATION (F [1, 22] = 4.1169, $p = .055$). The analysis revealed two significant interactions, first, between LOCATION and GENDER (F [1, 22] = 8.4120, $p < .01$). Post-hoc tests show that the girls have significantly longer vowels than the boys only in Perm (Figure 1, right), for the boys and girls in Moscow (left) used similar average vowel durations (see also Post 2025). This shows that the earlier mentioned gender effect is due to the difference in Perm.

---

[8]Bar plots of the mean durations (log-transformed) according to gender and according to distance to stress, as well as sound files and details on the data can be provided by the author.
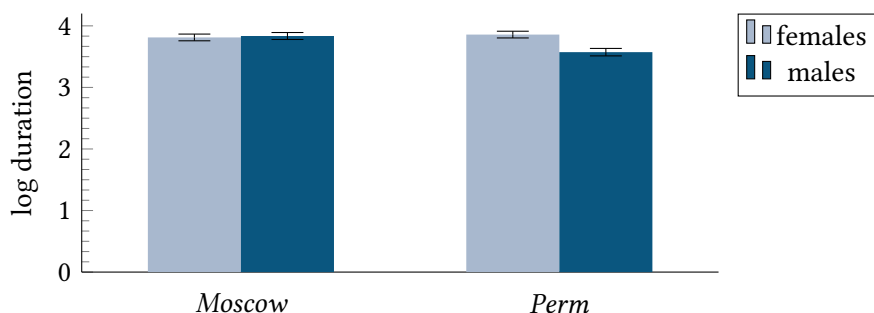
Figure 1: GENDER VS. LOCATION: Mean duration (log-transformed) of vowels (all three positions together), for female speakers (light-coloured) and Perm speakers (dark-coloured), with Moscow speech on the left and Perm speech on the right
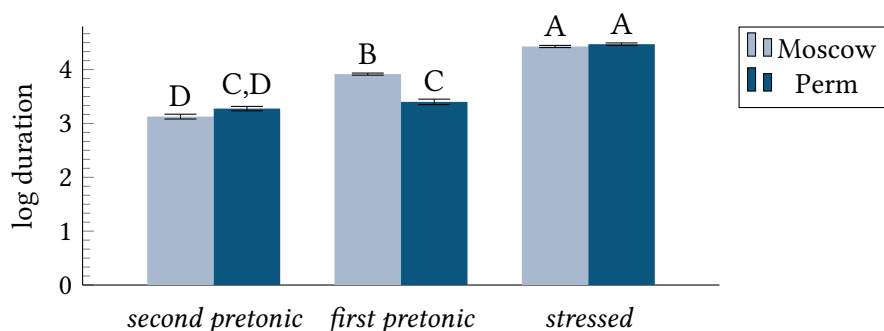
Figure 2: Mean duration (log-transformed) according to LOCATION and DISTANCE-TO-STRESS (light-coloured bars = Moscow; dark-coloured bars = Perm; both genders). Error bars represent standard errors, with Compact Letter Display of pairwise comparisons of the mean durations from the post-hoc pairwise Tukey HSD test. Bars not connected by the same letter are significantly different.

The second interaction is the highly significant interaction between LOCATION and DISTANCE-TO-STRESS (F [2, 431] = 65.0217, p < .001), which is the main result for our research question. In the realizations of the speakers from Moscow (Figure 2, light-coloured bars) the vowel duration becomes significantly shorter with increasing distance from the stressed syllable, whereas in the realizations of the speakers from Perm (Figure 2, dark-coloured bars) the only significant opposition we find is between the stressed vowels on the one hand and both unstressed vowels on the other. The small difference in average duration between the two prestressed vowels in Perm – cf. the minimal difference between second and first pretonics in Perm (dark-coloured bars) in Figure 2 – is not statistically significant. This is shown by the letter report from the post-hoc pairwise Tukey HSD test of mean values, the results of which are given above each bar in Figure 2. Bars not connected by the same letter are significantly different. The test gave a_{-2} in Perm the letters C and D, whereas a_{-1} received C, so the two positions share the same letter C, meaning that their small average difference in duration was not significant.

In Moscow, the actual mean durations (not log-transformed) of the three consecutive vowels (a_{-2} : a_{-1} : $á_0$) were 24.1 ms : 49.5 ms : 82.8 ms, which gives relative durations of almost 1 : 2 : 3. In Perm, the ratio is close to 1 : 1 : 3, with its mean durations of 28.3 ms : 33.1 ms : 89.3 ms.[9]

---

[9]For details on mean durations and an elaborate discussion of the durational measurements from a sociolinguistic perspective, see Post (2025).

There was no significant interaction in our data between GENDER and DISTANCE-TO-STRESS (F [2, 431] = 1.3204, $p >$ .05), nor between all three variables (F [2, 431] = 1.4552, $p >$ .05), for the boys used the same relative durations as the girls in both cities (see Post 2025).

These numbers show mean values, but the variation between individual tokens is large. Even with this high level of variability, the difference between Moscow and Perm in mean and relative durations of the pretonic vowels is highly significant.

3.2 Vowel quality: Formants

Parallel to the durational data, the statistical analyses of the normalized values of the first formants (zF1 Lobanov normalization) show a main effect of DISTANCE-TO-STRESS (F [2, 484] = 218.7679, $p <$ .001), with significantly different mean F1 values in each of the three positions (with letters C, B and A for a_{-2}, a_{-1} and $á_0$ in the letters report from the Tukey HSD test).

Unlike for the durational values, a significant interaction for zF1 was found between DISTANCE-TO-STRESS and GENDER (F [2, 478] = 6, 4192, $p <$ .01). This gender effect is mainly due to a difference in F1 in stressed position. Both the men and women differ significantly in F1 according to each position, but the Tukey HSD test showed that the female speakers produce relatively higher F1 values than the men in $á_0$.

The most relevant result for our research question, concerning the relative difference between the two pretonic positions in the two cities, is the correlation between LOCATION and DISTANCE-TO-STRESS, which again is highly significant (F [2, 484] = 89.7769, $p <$ .001; Figure 3), just like for duration (Figure 2). In Perm
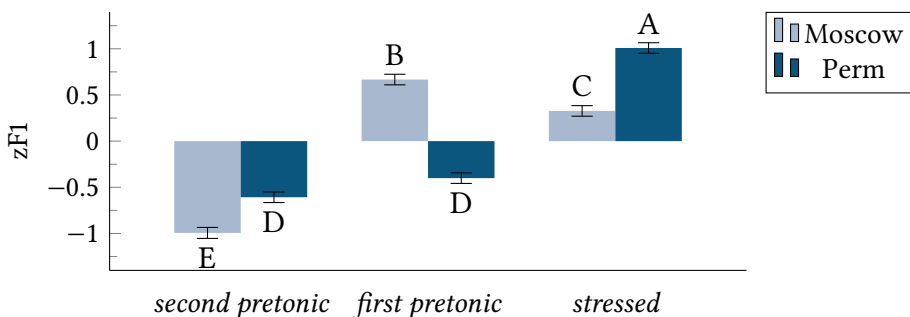
Figure 3: Mean zF1 values according to LOCATION and DISTANCE-TO-STRESS (both genders, light-coloured bars = Moscow; dark-coloured bars = Perm), with letter report from the pairwise comparison Tukey test

(dark-coloured bars), an opposition is again found only between stressed position on the one hand (level A in the Tukey pairwise comparison) and unstressed position on the other (level D for both a_{-2} and a_{-1}). In Moscow, the F1 values are significantly different in each position, like the durations, a_{-2}, a_{-1} and $á_0$ receiving the letters E, B, and C, respectively. Unlike for duration, a_{-1} has not lower, but higher mean F1 values than $á_0$ in stressed position, and differs most from a_{-2}, the other unstressed position.

The statistical measurements on our F2 data show no significant difference between the locations, Moscow and Perm. They reveal a main effect only of DISTANCE-TO-STRESS ($F [2, 485] = 384,7697$, $p < .05$; Figure 4).
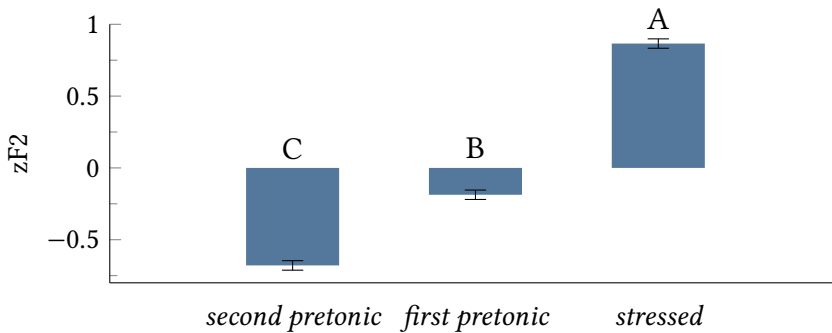
Figure 4: Mean zF2 values according to DISTANCE-TO-STRESS (both genders and both locations), including letter report from the pairwise comparison Tukey test

Our data do not support Kasatkina's (2005) claim of low F2 values in a_{-1} in varieties with a Northern substrate, but they do not disprove them either, given the limited value of our data for the comparison of actual formant values.

We also plotted F1 against F2 (Figure 5, not normalized), to visualize the actual measurements. The plots show a close to full overlap of the two pretonic vowels (square and triangle markers) in Perm, but much less overlap in Moscow, especially for F1 (*y*-axis). The highest degree of variation is found for a_{-1} (triangle markers) in Perm, covering a much larger space, especially in F2, than the stressed vowels (circle markers). The female (top) and male speakers (bottom) show the same patterns.

Figure 5: Vowel plots of F1 (*y*-axis) and F2 (*x*-axis) of the second pretonic (squares; dashed line), first pretonic (triangles; dotted line) and stressed (circles; solid line) vowels in the three target words produced by female speakers and male speakers in Moscow and Perm, with ellipses representing one standard deviation from the mean

4 Discussion

4.1 Durational differences: Larger than expected

The durational data confirm our hypothesis of a larger difference between the realizations of a_{-2} and a_{-1} in Moscow than in Perm. In fact, the difference between Moscow and Perm is larger than could be expected from previous literature. In Moscow, a_{-1} is, on average, twice as long as a_{-2}, whereas a_{-2} and a_{-1} hardly differ in Perm. In our data from adolescents, the relative difference in duration between the two pretonic vowels in Moscow is larger than in previous literature on Central Standard Russian (e.g. Zlatoustova 1981, Kuznecov 1997, Barnes 2006, Knjazev 2006) and even than in Vysotskij's (1973) data on traditional Moscow speech (cf. Post 2025). As remarked earlier, our speakers from Moscow do not necessarily have a Standard Russian pronunciation and might have a stronger local accent,

with a stronger prominence of the first pretonic vowel, than the participants of earlier studies of CSR. However, the first pretonic vowels are still not very prominent. They are shorter than in most previous studies (Zlatoustova 1981, *inter alia*), also relative to the duration of the stressed vowel, and both pretonic vowels are much shorter than the surrounding consonants, at least in the utterance in which we measured the durations of both vowels and consonants (cf. Post 2025).

The relative difference in duration between the pretonic vowels in our data from Perm is similar to the difference in the small data set from Perm measured by Erofeeva (2005) (see Post 2025) and proved not statistically significant. Unlike our data, Erofeeva's data were based on spontaneous speech from four speakers with an audible local accent, where one can expect a stronger degree of local colouring and, therefore, a smaller difference between the two pretonic vowels, than in a formal reading task by speakers who were not selected for locally coloured pronunciation. Nevertheless, in our reading task – recorded in a formal setting at school, which encourages the use of normative speech – the readers used the same locally coloured word prosodic structure.

4.2 Interpreting formant values in our data

The first and second formants of a vowel give an indication of its quality, F1 of its height (high or low tongue position) and F2 of the vowel's frontness (but cf. Whalen et al. 2022 for a recent warning not to equate formant measurements with vowel quality). The F1 and F2 values in our data should be approached with caution, since our reading task was not specifically designed to measure formant data. The vowels are surrounded by different consonants, giving different coarticulation effects (cf. Bondarko 1977: 65 on coarticulation from preceding consonants; Kasatkina 2005 on effects from consonants following the vowel). Consonant coarticulation is strong in Russian, especially on short unstressed vowels, like the vowels in our data. Another complicating factor is that our reading task contains both /o/ and /a/ in unstressed position (/a/ only in *potakat'*). These vowels merge categorically in unstressed position in Moscow speech and in CSR (Barnes 2006), but it is not certain that they have merged completely in Perm. Erofeeva (2005) found a slightly different distribution of unstressed /a/ and /o/ over vowel allophones in Perm, with a lower span of F2 values for unstressed /o/ than for unstressed /a/ (2005: 220). We cannot exclude that unstressed /o/ is still produced differently from unstressed /a/ by our speakers from Perm (or by some of them). Our reading task does not allow a direct comparison of pretonic /a/ with /o/ because of the different adjacent consonants. However, while non-merger might influence the absolute formant values of our data, it does not affect

our main research question, which concerns the relative difference between the two pretonic positions, which is very small in Perm in any case, not only for F1, but for F2 as well, as the following section 4.3 will argue.

4.3 Qualitative differences

Our results from the measurements of the first formant are very similar to those on duration as regards the relative difference between the two pretonic positions: Both measures show a large distance between the second and first pretonic vowels in Moscow, suggesting a very different opening grade for a_{-2} (level E) and a_{-1} (level B in the pairwise comparison test; Figure 3), but no significant difference between them in Perm (with both vowels on level D). The large difference in F1 values in Perm between the unstressed positions – both level D – with the stressed position – level A – suggests that the unstressed vowels are actually highly reduced in quality in Perm as well as in quantity, but in only one degree.

In Moscow, the main difference lies this time between a_{-2} and the other two positions. The F1 value of a_{-1} is just as high or even higher than in $á_0$ (Figure 3), suggesting it is not qualitatively reduced, although a_{-1} is substantially reduced in duration. F1 values as high in a_{-1} as in $á_0$, representing a fully open, low vowel, have earlier been found among speakers of Central Standard Russian, most, but not all of them, Muscovites (Kuznecov 1997, Knjazev 2006; three of the four speakers in Barnes 2006). The even higher values in prestressed than in stressed position can be caused by the vowel raising effect of the final palatalized consonant following $á_0$.

The pretonic vowels have different mean F2 values in both cities (Figure 4), but this appears to be due to factors other than relative distance to stress, for the difference between a_{-2} and a_{-1} disappears both in Moscow and Perm when one compares the mean values of the vowels in the same CVC string only – *-pot-* in *po$_{-2}$t(akat')* vs. *(to)po$_{-1}$t(at')*. The likely cause of the different F2 values when the other strings are included is consonant coarticulation, especially in Perm, where the vowels are even shorter than in Moscow, thus facilitating even stronger influence of the consonants.[10] In Perm, there might be an additional cause for the difference between the two pretonic positions: Incomplete merger of /a/ and /o/ would lead to higher F2 values for the phoneme /a/, which in our data only occurs in first pretonic position.

---

[10]The labial [p], frequent in second pretonic position, lowers F2, contrary to the laminal [t] and velar [k], the consonants surrounding a_{-1} in *potakat'* (cf. Bondarko 1977: 65).

4.4 Phonological status of two-degree reduction: Categorical difference in Moscow, not in Perm?

Trubetzkoy (1969, originally published in 1939) claimed that the allophones of /a, o/ in the two positions in Standard Russian are in complementary distribution. This is clearly not the case in (Standard) Belarusian, where we saw that the vowels are low irrespective of their distance to stress. Contrary to Trubetzkoy, Barnes (2006) and Iosad (2012) argue that the distinction between various reduction grades in Standard Russian is mainly one of duration (Iosad 2012: 531–532) and that the target vowel of /a, o/ after non-palatalized consonants in Standard Russian is a low vowel in all unstressed positions. This fits well with the Standard Belarusian reduction pattern. Barnes (2006) argues that the difference in quality found in Standard Russian is merely a result of phonetic implementation rules and caused by vowel undershoot: The vowels in second pretonic position are too short, he argues, to reach the low target level, which requires a minimal duration of approx. 60 ms in his data (Barnes 2006). What did we find in our data from Moscow and Perm?

Our data strongly suggest that the durational difference of /a, o/ between the two pretonic positions is categorical in Moscow speech. Not only is the first pretonic vowel on average twice as long as the second pretonic vowel, but, additionally, the asymmetry between the two positions is very stable: A closer look at the data shows that the second pretonic vowel is shorter than the first pretonic in all but a single token – in 77 out of the 78 analysed word tokens recorded in Moscow. This stability is remarkable, since these 78 tokens were pronounced with various nuclear and prenuclear pitch accents and focus patterns, resulting in a variety of tonal configurations and degrees of prominence on the target words. Our data obviously confirm Kasatkina's (2005) observation that the reduction in two degrees is not confined to utterances with a high tone on the first pretonic syllable, so if tone plays a role in the Central Russian word prosodic structure (Dubina 2012, Mołczanow 2015), it does indeed only on an abstract phonological level.

The large difference in quality between the pretonic positions in Moscow might be categorical as well, and it is unlikely to be caused solely by vowel undershoot.[11] The high average F1 values of a_{-1} among our Moscow speakers suggest that our young speakers need less time than 60 ms to reach the target level of a vowel, since their mean duration is only 51 ms. The target in second pretonic

---

[11]The difference in first formant values between the two positions is not only large. In addition, our Moscow speakers tend to produce higher F1 values for /a, o/ in a_{-1} than in a_{-2} even in those rare cases when both vowels have similar durations. A presentation of these data and a further discussion of the relation between F1 and duration is outside the scope of this paper.

position in Moscow speech might not be a high vowel, as Barnes (2006) suggests for CSR, but a vowel with a quality different from the one in first pretonic position (cf. Trubetzkoy 1969), thereby showing both qualitative and quantitative vowel dissimilation, not between the tonic and first pretonic vowels, as in the dissimilating Russian and Belarusian rural dialects, but between the second pretonic and first pretonic position. Further studies are needed to find out whether this pattern is found among Moscow speakers only.

In Perm, on the other hand, the differences between the pretonic positions in duration and in F1 are small and not statistically significant. The Perm speakers make a clear distinction between stressed and unstressed vowels, in both duration and in quality, but there is no evidence of a phonological distinction between first and second pretonic position.

5 Conclusions

In Central Standard Russian the prosodic word has a strong nucleus, consisting of the first pretonic and the tonic syllable, which leads to vowel reduction in two degrees, expressed in both quantity and quality, at least when the vowels are /a/ or /o/ after non-palatalized consonants. Our study is, to our knowledge, the first acoustic study of a considerable number of speakers that compares the vowel reduction patterns of these vowels in speech from Central European Russia, represented by Moscow, with modern urban speech from a different region, represented by Perm.

Our main research question concerns the relative difference in duration and quality between the second and first pretonic vowels in the two cities. In our reading task, they have a significantly different word prosodic pattern. Our data corroborate earlier suggestions that the difference between the two degrees is expressed less clearly in non-central varieties of modern urban Russian. In Moscow, the distinction between the first and second pretonics is remarkably large, even larger than in previous studies of Central Standard Russian. On average, the durational relation between second pretonic and first pretonic vowel is almost $1 : 2$ in our data. However, although a_{-1} – /a, o/ in first pretonic position – is twice as long as second pretonic a_{-2}, it is still much shorter than the stressed vowel $á_0$. In Perm, the difference between the two pretonic vowels was small and not statistically significant. In Moscow, the vowels also differ greatly in quality: the F1 formant values of a_{-2} are much lower than for a_{-1}, where a_{-1} can have an even higher mean F1 value, suggesting a lower tongue position, than the stressed vowel, even though it is much shorter. In Perm, however, the mean F1 values are low in both

pretonic syllables, and much lower than in stressed /a/, confirming Erofeeva's (2005) findings of considerable qualitative reduction, in addition to a very strong quantitative reduction.

Our data imply that the clear distinction between two degrees of durational reduction of /a, o/ (after non-palatalized consonants) is part of the phonology of Moscow speech. A further study of the relation between F1 and duration might confirm the indications that the qualitative distinction between the pretonic positions is part of Moscow phonology as well, and not a result of phonetic implementation, as Barnes (2006) suggests for Central Standard Russian. In Perm, on the other hand, the differences between the pretonic positions are so small that they suggest there might not be any phonological reduction in two degrees at all. This is remarkable, given the fact that in previous research even all traditional rural dialects had two-degree durational reduction to some degree, with a possible exception of a subgroup of Northern European rural dialects (Vysotskij 1973). One would not expect to find this most extreme pattern in modern urban Russian speech. Unlike these dialects, however, Perm speech shows much qualitative reduction of unstressed /a/ and /o/.

Thus, the normative pronunciation standard (Avanesov 1984) of vowel reduction in two clearly distinguished degrees is not a general feature of modern urban Russian speech, even though modern urban Russian shows little regional variation. This norm is not followed by our young speakers from Perm, not even in a reading task, where the tendency to follow high status norms is higher than in informal spontaneous speech. We found no gender differences in the word prosodic structure. Our comparison shows that today's urban youth still have local prosodic traits in their speech. Both Moscow and Perm speech are known for their local accent, so more research would be welcome on speech from other regions and from the other countries where East-Slavic languages are spoken, to give a better picture of the variation in vowel reduction processes in East Slavic.

Abbreviations

1	first person	NEG	negation
F	feminine	NOM	nominative
FUT	future	PST	past tense
INF	infinitive	SG	singular
IPFV	imperfective		

Acknowledgments

References

Al'muxamedova, Zel'fa M. & Ravza È. Kul'šaripova. 1980. *Redukcija glasnyx i prosodija slova v okajuščix russkix govorax*. Kazan': Kazan' University Press.

Andreev, Andrej N. 1984. Akustičeskie xarakteristiki glasnyx v belorusskom literaturnom jazyke. *Vesci Akadèmii navuk BSSR. Seryja hramadskix navuk* 1. 109–117.

Andrews, David R. 1995. Subjective reactions to two regional pronunciations of Great Russian: A matched-guise study. *Canadian Slavonic Papers* 37(1–2). 89–106. DOI: 10.1080/00085006.1995.11092083.

Avanesov, Ruben I. 1984. *Russkoe literaturnoe proiznošenie*. 6th ed. Moscow: Prosveščenie.

Avanesov, Ruben I. & Sof'ja V. Bromlej (eds.). 1986. *Dialektologičeskij atlas russkogo jazyka: Centr Evropejskoj časti SSSR*. Moscow: Nauka.

Barnes, Jonathan. 2006. *Strength and weakness at the interface: Positional neutralization in phonetics and phonology*. Berlin & New York: Mouton de Gruyter.

Bethin, Christina Y. 2006. Stress and tone in East Slavic dialects. *Phonology* 23(2). 125–156. DOI: 10.1017/S0952675706000868.

Boersma, Paul & David Weenink. 2023. *Praat: Doing phonetics by computer [Computer programme]*. http://www.praat.org/.

Bondarko, Lija V. 1977. *Zvukovoj stroj sovremennogo russkogo jazyka*. Moscow: Prosveščenie.

Bondarko, Lija V., Ljudmila A. Verbickaja & Lev R. Zinder. 1966. Akustičeskie xarakteristiki bezudarnosti. In Vjačeslav V. Ivanov (ed.), *Strukturnaja tipologija jazykov*, 56–64. Moscow: Nauka.

Borise, Lena. 2017. Prominence redistribution in the Aŭciuki dialect of Belarusian. In Yohei Oseki, Masha Esipova & Stephanie Harves (eds.), *Annual workshop on formal approaches to Slavic linguistics: The NYU Meeting 2015*, 94–109. Ann Arbor: Michigan Slavic Publications.

Broch, Olaf. 1916. *Govory k zapadu ot mosal'ska*. Petrograd: Tipografija Imperatorskoj akademii nauk.

Černjavskij, Aleksej R. 2012. Orfoèpičeskie pravila akan'ja v sovremennom belorusskom jazyke v sopostavlenii s dialektami. In Ljudmila È. Kalnyn' (ed.), *Issledovanija po slavjanskoj dialektologii XV: Osobennosti sosuščestvovanija dialektnoj i literaturnoj form jazyka v slavjanojazyčnoj srede*, 121–125. Moscow: Institut slavjanovedenija Rossijskoj akademii nauk.

Comrie, Bernard & Greville G. Corbett. 1993. *The Slavonic languages*. London & New York: Routledge.

Crosswhite, Katherine. 2000. Vowel reduction in Russian: A unified account of standard, dialectal, and "dissimilative" patterns. In Katherine Crosswhite & Joyce McDonough (eds.), *University of Rochester working papers in the language sciences, vol. Spring 2000, no. 1*, 107–171. Rochester: University of Rochester.

Dubina, Andrei. 2012. *Towards a tonal analysis of free stress* (LOT Dissertation Series 313). Utrecht: LOT.

Erofeeva, Elena V. 1993. *Fonetičeskaja variativnost' fonologičeskoj sistemy glasnyx (sopostavitel'nyj analiz peterburgskogo i permskogo vokalizma)*. Saint Petersburg: Saint Petersburg State University.

Erofeeva, Elena V. 2005. *Idiomy kak verojatnostnyje struktury: Sociolingvističeskij aspekt (na materiale fonetičeskogo urovnja)*. Saint Petersburg: Saint Petersburg State University. (Doctoral dissertation).

Erofeeva, Tamara I., Elena V. Erofeeva & I. I. Gračeva. 2000. *Gorodskie sociolekty: Permskaja gorodskaja reč': Zvučaščaja xrestomatija*. Perm' & Bochum: Ruhr-Universtät Bochum.

Grammatčikova, E. V., Sergej V. Knjazev, Sof'ja K. Požarickaja & L. V. Luk'janova. 2013. Ritmičeskaja struktura slova i mesto realizacii tonal'nogo akcenta v regional'nyx variantax sovremennogo russkogo literaturnogo jazyka. In Alek-

sandr V. Arxipov & Irina M. Kobozeva (eds.), *Aktual´nye voprosy teoretičeskoj i prikladnoj fonetiki*, 69–90. Moscow: Buki-Vedi.

Iosad, Pavel. 2012. Vowel reduction in Russian: No phonetics in phonology. *Journal of Linguistics* 48(3). 521–571. DOI: 10.1017/S0022226712000102.

Kasatkina, Rozalija F. 1996. Srednerusskie govory i ritmika slova. In Tat'jana M. Nikolaeva (ed.), *Prosodičeskij stroj russkoj reči*, 222–235. Moscow: Russian Language Institute of the Russian Academy of Sciences.

Kasatkina, Rozalija F. 2005. Moskovskoje akan'e v svete nekotoryx dialektnyx dannyx. *Voprosy jazykoznanija* 2. 29–45. https://vja.ruslang.ru/ru/archive/2005-2/29-45.

Knjazev, Sergej V. 2006. *Struktura fonetičeskogo slova v russkom jazyke: Sinxronija i diaxronija*. Moscow: MAKS-Press.

Kocharov, Daniil, Tatiana Kachkovskaia & Pavel Skrelin. 2015. Position-dependent vowel reduction in Russian. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow*, Paper number 0402. London: International Phonetic Association. https : / / www . internationalphoneticassociation . org / icphs - proceedings / ICPhS2015/Papers/ICPHS0402.pdf.

Kodzasov, Sandro. 1999. Russian. In Harry van der Hulst (ed.), *Word prosodic systems in the languages of Europe*, 852–870. Berlin: De Gruyter.

Kuznecov, Vladimir I. 1997. *Vokalizm svjaznoj reči: Èksperimental'noe issledovanie na materiale russkogo jazyka*. Saint Petersburg: Saint Petersburg University Press.

Labov, William. 2001. *Principles of linguistic change, volume 2: Social factors*. Oxford: Blackwell.

Lobanov, Boris M. 1971. Classification of Russian vowels spoken by different listeners. *Journal of the Acoustical Society of America* 49. 606–608. DOI: 10.1121/1.1912396.

Łukaszewicz, Beata & Janina Mołczanow. 2018. The role of vowel parameters in defining lexical and subsidiary stress in Ukrainian. *Poznan Studies in Contemporary Linguistics* 54(3). 355–375. DOI: 10.1515/psicl-2018-0014.

Łukaszewicz, Beata, Janina Mołczanow & Anna Łukaszewicz. 2022. Pretonic lengthening as the lexical stress domain extension. In Sónia Frota, Marisa Cruz & Marina Vigário (eds.), *Proceedings of Speech Prosody 2022, 23–26 May 2022, Lisbon, Portugal*, 372–376. DOI: 10.21437/SpeechProsody.2022-76.

Mołczanow, Janina. 2015. The interaction of tone and vowel quality in Optimality Theory: A study of Moscow Russian vowel reduction. *Lingua* 163. 108–137. DOI: 10.1016/j.lingua.2015.05.007.

Padgett, Jaye & Marija Tabain. 2005. Adaptive dispersion theory and phonological vowel reduction in Russian. *Phonetica* 62(1). 14–54. DOI: 10.1159/000087223.

Post, Margje. 2017. Why regional prosodic variation is worth studying: An example from Russian. *Bergen Language and Linguistic Studies* 7. 164–182. DOI: 10.15845/bells.v7i0.1138.

Post, Margje. 2025. Regional prosodic variation in the speech of young urban Russians: Quantitative vowel reduction in Moscow and Perm. In Susanne Wagner & Ulrike Stange-Hundsdörfer (eds.), *(Dia)lects in the 21st century: Selected papers from Methods in Dialectology XVII* (Language Variation), 241–268. Berlin: Language Science Press. DOI: 10.5281/zenodo.15006613.

Post, Margje & Bistra Andreeva. 2023. Polar question intonation in Russian speech from Moscow and Perm. In Oliver Niebühr & Malin Svensson Lundmark (eds.), *Proceedings of the 13th International Conference of Nordic Prosody*, 147–154. Sciendo. DOI: 10.2478/9788366675728-012.

Potebnja, Aleksandr A. 1866. *Dva issledovanija o zvukax russkogo jazyka: I. O polnoglasii. II. O zvukovyx osobennostjax russkix narečij.* Voronež: Tipografija V. Goľdštejna.

Trubetzkoy, Nikolay S. 1969. *Principles of phonology.* Berkeley: University of California Press.

Vardøy, Benedikte Fjellanger. 2021. Mapping young Russians' perceptions of regional variation in Russian. *Journal of Linguistic Geography* 9(1). 50–68. DOI: 10.1017/jlg.2021.5.

Verbickaja, Ljudmila A. 1977. Variantnosť sovremennoj proiznositeľnoj normy russkogo literaturnogo jazyka. *Vestnik Leningradskogo gosudarstvennogo universiteta* 8. 133–137.

Verbickaja, Ljudmila A., Ljudmila V. Ignatkina, N. F. Litvačuk, T. A. Sergeeva, Marija V. Cvetkova & Vasilij G. Ščukin. 1984. Regionaľnye osobennosti realizacii russkoj reči (na fonetičeskom urovne). *Vestnik Leningradskogo universiteta* 8(2). 71–80.

Vysotskij, Sergej S. 1973. O zvukovoj strukture slova v russkix govorax. In Sof ja V. Bromlej (ed.), *Issledovanija po russkoj dialektologii*, 17–41. Moscow: Nauka.

Whalen, D. H., Wei-Rong Chen, Christine H. Shadle & Sean A. Fulop. 2022. Formants are easy to measure; resonances, not so much: Lessons from Klatt (1986). *The Journal of the Acoustical Society of America* 152(2). 933–941. DOI: 10.1121/10.0013410.

Zlatoustova, Ljubov' V. 1981. *Fonetičeskie edinicy russkoj reči.* Moscow: Moscow State University Press.

Chapter 17

Morphosemantic mismatches with pronouns as a consequence of their internal structure

Zorica Puškar-Gallien

Leibniz-Centre General Linguistics

In addition to differences in their form and position in a sentence, strong pronouns and clitics in Bosnian/Croatian/Montenegrin/Serbian show systematic form-meaning mismatches. Strong pronouns license only animate referents and strict identity readings, whereas clitics show no such restrictions. This paper focuses on two exceptional contexts in which inanimate interpretation and sloppy identity readings are permitted on strong pronouns: focus contexts (acknowledged in previous literature) and prepositional phrases (novel contribution). The seemingly unrelated properties of pronominal elements can be accounted for under a unified approach to (pro)nominal syntactic structure. I will argue for a hierarchy of nominal projections: base $\succ$ ϕ-features $\succ$ case, whereby ϕ-features further split into a hierarchy (person $\succ$ number $\succ$ gender). Under the additional assumption that the pronominal base (nP) is a phase, and that it encodes referentiality and individuation features, its absence from the structure (due to deletion) will account for the spell-out of clitics and sloppy identity readings, while the blocking or deletion will allow for the same with strong pronouns in PPs and focus contexts.

1 Introduction

The goal of this paper is to develop a formal description of the morphological distinctions, distribution and form-meaning mismatches of pronominal elements in Bosnian/Croatian/Montenegrin/Serbian (BCMS), based on a unified model of the form, locus and function of their ϕ- and case features. BCMS personal pronouns distinguish between the so-called STRONG PRONOUNS (pronouns in their

full form) and clitics. The main claim that this paper will advance is that some seemingly unrelated properties of pronominal elements, which will be inspected throughout the paper, can be accounted for as a consequence of a unified approach to (pro)nominal syntactic structure, which relies on the key notion of HIERARCHY.

Pronominal elements in BCMS differ across two dimensions: local person (1st and 2nd person) vs. 3rd person pronouns on the one hand, and strong pronouns vs. clitics on the other. Looking at their morphological structure, clitics are morphologically reduced forms of strong pronouns. For instance, the accusative forms of third person singular pronouns are *nje-ga* '3.SG.M.ACC', *nje* '3.SG.F.ACC', *nje-ga* '3.SG.N.ACC', while the corresponding clitics are realised by a portmanteau morpheme expressing gender, number and case, omitting the base *nj(e)-*, i.e *ga*, *je*, *ga*. On a different dimension, local person pronouns seem to spell out all their phi-features in the form of a portmanteau and their case separately, while third person pronouns spell out the base separately from gender, number and case, resembling lexical nouns and adjectives.

Strong pronouns have been argued to license only animate referents and strict identity readings whereas clitics show no such restrictions. While the lack of animacy in focus contexts was acknowledged in previous literature, I will present novel data from prepositional phrases which further blur this seemingly sharp divide by demonstrating that strong pronouns in the complement of a P position may in fact be inanimate and license sloppy identity readings.

This disparate set of distributional properties of pronominal elements in BCMS raises the question whether there is a way to unite them under a single analysis. The first step towards such an analysis requires us to look at the properties outlined above in further detail, which will be the task of Section 2 below. The core of the proposal will be based on the claim that the internal structure of a pronoun involves several hierarchies: (i) Within the pronominal extended projection, consisting of a nominal base, followed by ϕ-feature-encoding projections, followed in turn by case ([Case [Φ [NP]]]); (ii) within ϕ-features (Harley & Ritter 2002), such that person precedes number, which itself precedes gender ([gender [number [person]]]); and (iii) within case features (Caha 2009), which distinguishes between the following types of case – unmarked (NOM) ≻ dependent (ACC, GEN) ≻ oblique (DAT) ≻ prepositional (INS, LOC). I will further propose that these hierarchies are structurally encoded in the syntax (Béjar & Řezáč 2009, van Koppen 2012). Distribution of nominal features across them and the locality domains they define will be shown to have consequences on the morphology of pronouns (Moskal 2015b), interpretation and ability to move. In particular, local-person pronouns will differ from third-person pronouns in whether they encode

grammatical gender (Puškar-Gallien 2019); while the former cannot do it, for the latter it is one of their defining properties. Clitics and strong pronouns share the same structure, but clitics crucially lack the NP base. As I will argue, due to the location of features [animate] and [human] on the NP, and their deterministic role in establishing individuation, as well as N's role in establishing reference, the absence of N (modelled as deletion after van Urk 2018) will allow for certain semantic flexibility which will lead to the possibility of sloppy readings of clitics.

The paper is structured as follows. Section 2 introduces the pronominal paradigms and morphosemantic mismatches. A short overview of previous literature and certain issues raised from it will be presented in Section 3. The proposal on the internal structure of pronominal elements will occupy Section 4. Subsequently, Section 5 will inspect the consequences of the proposal for syntax and interpretation in more detail. Section 6 summarises and concludes.

2 Properties of pronominal elements in BCMS

2.1 Morphological form

An overview of the BCMS personal pronouns and clitics is presented in Table 1; clitics are outlined in boldface. First and second person pronouns share the same set of case endings, and realise their base (comprising of π (person) and # (number)) separately from their case features. I will consider the morphemes *-en-* and *-eb-* in the singular to be the so-called "support morphemes" (Cardinaletti & Starke 1999), which distinguish the strong pronoun forms from their clitic counterparts. The clitic forms of those pronouns are the simple *me* and *te*, without this extension. The base of first person pronouns undergoes suppletion in all non-nominative cases (cf. *ja* vs. *m-* / *na-*), as well as in the plural, while second person pronouns undergo suppletion in the plural (*ti* vs. *vi*). The third person pronouns' base undergoes suppletion in non-nominative environments, resulting in the *nj(e)-* allomorph. This morpheme is followed by a portmanteau morpheme that realises gender, number and case features, which shares its paradigm with adjectival inflection.

As for clitics, they are available in genitive, accusative and dative. Local-person clitics spell out the person, number and case features without the support morpheme, whereas third-person clitics amount to the spellout of the gender, number and case suffix, without the pronominal base *on-/nj(e)-*.

Table 1: Strong pronouns vs. clitics in BCMS

	1SG	2SG	1PL	2PL	3SGM/N	3SGF	3PL
NOM	*ja*	*ti*	*mi*	*vi*	*on-∅/-o*	*on-a*	*on-i/-e/-a*
GEN	*m-en-e*	*t-eb-e*	*na-s*	*va-s*	*nje-ga*	*nj-e*	*nj-ih*
DAT	*m-en-i*	*t-eb-i*	*na-ma*	*va-ma*	*nje-mu*	*nj-oj*	*nj-ima*
ACC	*m-en-e*	*t-eb-e*	*na-s*	*va-s*	*nje-ga*	*nj-u*	*nj-ih*
INS	*m-n-om*	*t-ob-om*	*na-ma*	*va-ma*	*nj-im*	*nj-om*	*nj-ima*
LOC	*m-en-i*	*t-eb-i*	*na-ma*	*va-ma*	*nje-mu*	*nj-oj*	*nj-ima*

2.2 Restrictions on reference

2.2.1 Animacy

As noted in previous literature (e.g. Despić 2011), a clitic can be interpreted as referring to either an animate (or rather human), or an inanimate referent, in contrast to a strong pronoun, which can only be interpreted as denoting a human entity.

(1) *Clitics vs. pronouns, animacy/humanness (Despić 2011: 240)*

  a. Čuo   sam   je .
      heard.M.SG AUX.1.SG   CL.3.F.SG.ACC
      'I heard her/it.'   [+HUM] [-HUM]

  b. Čuo   sam   nju .
      heard.M.SG AUX.1.SG   3.F.SG.ACC
      'I heard her.'   [+HUM] *?[-HUM]

Exceptions to this generalization have been shown to appear in prepositional phrases and focus contexts. Specifically, in a PP, it is not possible to realise a clitic, instead a strong pronoun is necessary (2) (as also discussed by Abels 2012, Milićev & Bešlin 2019).[1]

---

[1] See Stegovec (2019) for a tripartite distinction between Slovenian strong, clitic and P-pronouns, present in earlier stages of BCMS.

(2) *Clitics vs. pronouns in a PP*
Slavica kupuje poklon za
Slavica buys   present for
njega/nju/*ga/*ju .
3.M.SG.ACC/3.F.SG.ACC/CL.3.M.SG.ACC/CL.3.F.SG.ACC
'Slavica is buying a present for him/her.'

What has, to my knowledge, hitherto escaped closer scrutiny is that such a strong pronoun in a complement of P position can in fact refer to an inanimate entity. The following sentences illustrate this for genitive (3a), dative (3b), and accusative case (3c).

(3) *Strong pronouns as complements of P*

a. Dok vozi, Ljubica uglavnom koristi svoj telefon   za navigaciju,
   while drives Ljubica mostly   uses   her phone.M.SG for navigating

   a Tamara se   dobro snalazi i [PP bez   njega].
   but Tamara REFL good manages and   without 3.M.SG.GEN
   'While driving, Ljubica mostly uses her phone for navigating and Tamara manages well without it.'   (GEN, INANIM)

b. Jelena mnogo voli svoj novi posao , a   Jovana oseća izrazitu
   Jelena a.lot loves self's new job.M.SG but Jovana feels distinct

   odbojnost [PP prema   njemu].
   revulsion   towards 3.M.SG.DAT
   'Jelena likes her new job a lot and Jovana finds it repulsive.'
                                                          (DAT, INANIM)

c. Mladen je prošao kroz   svoja pitanja   za kontrolni, a   i
   Mladen is went   through self's questions.N.PL for test   but and

   Saša   je takođe prošao [PP kroz   njih].
   Sasha is also   went   through 3.N.PL.ACC
   'Mladen went through his questions for the test and Sasha went through them too.'   (ACC, INANIM)

Additionally, instrumental and locative strong pronouns (those without clitic counterparts), show the same behaviour. This has also been noted for Slovenian by Stegovec (2019), and can be illustrated by the examples in (4). By analogy with (3), I will use this to argue that instrumental and locative are in fact PPs in BCMS.

(4) *Strong pronouns in instrumental and locative*

    a. Slavica uglavnom putuje bez   svog velikog ⟦ruksaka⟧,   a
        Slavica mostly   travels without self's big   backpack.M.SG but

        Jovan obavezno   putuje [PP s   ⟦njim⟧].
        Jovan necessarily travels   with 3.F.SG.INS
        'Slavica mostly travels without her big backpack, but Jovan
        necessarily travels with it.'   (INS, INANIM)

    b. Lena se   rado igra u svojoj ⟦sobi⟧,   a Matija samo uči
        Lena REFL gladly play in self's room.F.SG but Matija only studies

        [PP u ⟦njoj⟧].
           in 3.F.SG.LOC
        'Lena likes to play in her room and Matija only studies in it.'
                                (LOC, INANIM)

Finally, if a strong pronoun is marked as discourse prominent by focus or topicalisation, it may also be inanimate. The following example illustrates this for a focused pronoun. Compare (5) to (1) above.

(5) *Focused inanimate pronoun (Despić 2011: 246)*
    Čuo   sam   čak i   ⟦nju⟧.
    heard.M.SG AUX.1.SG even and 3.F.SG.ACC
    'I heard even it (lit. her).'   [+HUM] [-HUM]

It should also be noted that strong pronouns referring to inanimate entities can appear in argument positions even without focus particles, but in this case they normally introduce a contrastive topic, cf. (6). The generalisation however remains that information structure properties facilitate inanimate interpretations of strong pronouns.

(6) *Topical inanimate pronoun*
    Ovo je moj novi bicikl. Njega   su   mi   poklonili
    this is my new bicycle 3.M.SG.ACC AUX.3.PL CL.1.SG.DAT given
    roditelji za rođendan.
    parents for birthday
    'This is my new bicycle. It was given to me by my parents for my
    birthday.'

2.2.2 Sloppy identity readings

Another property that distinguishes strong pronouns from clitics in BCMS is their ability to function as bound variables. Specifically, while strong pronouns may only strictly refer to their antecedent, clitics can license sloppy identity readings (in addition to strict ones).[2] According to Franks (2013), factors that affect the availability of sloppy identity readings include animacy, modification of the antecedent and regional variant, however Runić (2014) argues that all that is necessary is the appropriate context, e.g. (7) (see also Ruda 2021a,b for Polish). Note that examples (7a)–(7b) may not seem to be entirely parallel, due to the second position requirement on the clitic placement, however see Section 5.1 for further detail.[3]

(7)   *Clitics vs. pronouns regarding sloppy readings*

    a. Nikola je   vidio zanimljivog klovna, a   vidio ga
       Nikola AUX.3.SG saw   interesting clown   and saw CL.3.SG.M.ACC

       je   i   Danilo.
       AUX.3.SG and Danilo
       'Nikola saw an interesting clown and Danilo saw him/one too.'
       (✓ Nikola saw an interesting clown and Danilo saw him (=the same clown that Nikola saw))
       (✓ Nikola saw an interesting clown and Danilo saw one (=a different clown from Nikola's.)

    b. Nikola je   vidio zanimljivog klovna, a   njega   je
       Nikola AUX.3.SG saw   interesting clown   and 3.SG.M.ACC AUX.3.SG

       vidio i   Danilo.
       saw   and Danilo
       'Nikola saw an interesting clown, and Danilo saw him/*one too.'
       (✓ Nikola saw an interesting clown and Danilo saw him (=the same clown that Nikola saw).)
       (✗ Nikola saw an interesting clown and Danilo saw one (=a different clown from Nikola's).)   (Runić 2014: 123-124)

---

[2]The discussion here is restricted to third-person clitics.

[3]The context for sloppy reading in (7) as suggested by Runić (2014: 123) is the following: 'Nikola and Danilo are cousins who live in two different cities in Serbia. Specifically, Nikola lives in Belgrade, while Danilo lives in Niš. They are both five years old and their parents take them to circus performances whenever a circus is in town. A circus is in both Belgrade and Niš at the same time. Both Nikola and Danilo saw an interesting clown in the circus, albeit not the same one'.

A novel observation I put forward is that BCMS strong pronouns in complement of P position may also allow for sloppy readings, as the examples repeated in (8) show. Example (8a) illustrates this for genitive case, (8b) for dative, and (8c) for accusative.

(8) *Sloppy readings of strong pronouns as complements of P*

    a. Dok vozi, Ljubica uglavnom koristi svoj telefon za navigaciju,
        while drives Ljubica mostly uses her phone.M.SG for navigating
        a Tamara se dobro snalazi i [PP bez njega].
        but Tamara REFL good manages and without 3.SG.M.GEN
        'While driving, Ljubica mostly uses her phone for navigating and Tamara manages well without Ljubica's phone/Tamara's phone.'

    b. Jelena mnogo voli svoj novi posao , a Jovana oseća izrazitu
        Jelena a.lot loves self's new job.M.SG but Jovana feels distinct
        odbojnost [PP prema njemu].
        revulsion towards 3.M.SG.DAT
        'Jelena likes her new job a lot and Jovana finds it (Jelena's job/ Jovana's job) repulsive.'

    c. Mladen je prošao kroz svoja pitanja za kontrolni, a i
        Mladen is went through self's questions.N.PL for test but and
        Saša je takođe prošao [PP kroz njih].
        Sasha is also went through 3.N.PL.ACC
        'Mladen went through his questions for the test and Sasha went through them (Sasha's/Mladen's questions) too.'

The same holds for instrumental and locative, as repeated in (9).

(9) *Sloppy readings of strong pronouns in instrumental and locative*

    a. Slavica uglavnom putuje bez svog velikog ruksaka , a
        Slavica mostly travels without self's big backpack.M.SG but
        Jovan obavezno putuje [PP s njim].
        Jovan necessarily travels with 3.F.SG.INS
        'Slavica mostly travels without her big backpack, but Jovan necessarily travels with it (Slavica's/Jovan's backpack).'

    b. Lena se rado igra u svojoj sobi , a Matija samo uči
        Lena REFL gladly play in self's room.F.SG but Matija only studies

[PP u njoj].
in 3.F.SG.LOC
'Lena likes to play in her room and Matija only studies in it (Lena's/ Matija's room).'

The sentences in (8)–(9) were included in an informal survey, completed by 35 native speakers, recruited through the online community (a Facebook group) *Kako biste VI rekli?* 'How would YOU say?'. Based on a short context, the participants were asked to rate the sentence (thus probing the acceptance of animacy restrictions) and choose the appropriate interpretation in a multiple-choice task (choice between the strict and the sloppy interpretation, or both). For instance, (8a) received an overall rating of 4/5 and 25/35 speakers chose the sloppy identiy reading as the preferred interpretation. This confirms that the context plays a big role, but so does the sentence structure. A more formal and balanced further study is planned in order to confirm and elaborate on these results, considering additional factors such as the position of the PP. Nevertheless, the fact that BCMS speakers accept sloppy identity readings of strong pronouns in this context indicates that the divide between strong pronouns and clitics may not be as sharp as is normally drawn, which any theory that models them should be able to account for.

2.2.3 Information structure

An additional distinction between strong pronouns and clitics in BCMS associates strong pronouns with focus, and clitics with topical interpretation. In BCMS, only strong pronouns may express new-information or contrastive focus (or require an antecedent that carries focus, see Despić 2011, Jovović 2024), as illustrated in (10), where the sentence-final position is normally the one where contrastive focus is introduced.

(10) *Strong pronouns and focus*
      Who did you see?

      a. # Video   sam   ga .
          seen.M.SG AUX.1.SG CL.3.M.SG
          'I saw him.'

      b. Video   sam   njega .
          seen.M.SG AUX.1.SG 3.M.SG
          'I saw him.'   (Despić 2011: 245)

Clitics, on the other hand, are topical elements, or require antecedents that express discourse-given information (Jovović 2024). If contrastive focus is present, a strong pronoun must be used as in (11b). Note that (11b) remains ungrammatical even if the clitic is moved to its (expected) second position in the clause (11c).

(11) *Clitics and topicality*

    a. Svaki predsednik$_i$ misli da | ga$_i$/??njega$_i$ | svi
       every president   thinks that CL.3.M.SG.ACC/3.M.SG.ACC everyone
       vole.
       love
       'Every president$_i$ thinks that everybody loves him$_i$.'

    b. Svaki predsednik$_i$ misli da samo {njega$_i$ / *ga$_i$}
       every president   thinks that only   3.M.SG.ACC   CL.3.M.SG.ACC
       svi   vole.
       everyone love
       'Every president$_i$ thinks that everyone loves only him$_i$.'

    c. *Svaki predsednik$_i$ misli da ga   samo svi   vole.
       every president   thinks that CL.3.M.SG.ACC only everyone love
       Intended: 'Every president$_i$ thinks that everyone loves only him$_i$.'
                                      (Despić 2011: 243)

Focus in BCMS requires prosodic prominence, which clitics always lack, which in turn makes them illicit in a focus position.[4] If a focused pronoun allows for inanimate reference as in (12)–(13), Despić (2011: 244) argues that such a pronoun is merely a clitic that has to be spelled out as a strong pronoun due to the phonological requirements on focused constituents. Such a 'camouflaged clitic' (Despić 2011: 244) should also be able to act as a bound variable, as illustrated by (11b) above.

(12) *Focused inanimate pronoun*
    Čuo   sam   čak i | nju |.
    heard.M.SG AUX.1.SG even and 3.F.SG.ACC
    'I heard even it/her.'   [+HUM] [-HUM] (Despić 2011: 246)

---

[4]See Browne (1974), Zec & Inkelas (1991), Franks & Progovac (1994), Godjevac (2000) on clitics lacking prosodic prominence, Godjevac (2000) on focus requiring prosodic prominence, and Despić (2011: 244) on further interactions between the two.

(13) *Focused inanimate pronoun*
      Malo ko   obilazi muzeje   oko   gradske crkve*ᵢ*. Nju*ᵢ*   *(samu),
      few   who visits   museums around city   church 3.ꜰ.sɢ.ᴀᴄᴄ   alone
      opet   dnevno poseti oko   50 turista.
      again daily   visits   around 50 tourists
      'A few people visits museums around the city church. (As for the church
      itself), an average of 50 tourists visits it a day.'   (Despić 2011: 247)

The animacy properties, the ability to be bound and the sloppy readings outlined
in Section 2.2.2 indicate a lack of inherent referentiality of strong pronouns in
these contexts. This may be the reason why Cardinaletti & Starke (1999) treat
them as weak pronouns, or why Despić (2011: 244) treats them as clitics in dis-
guise.

3 Theoretical puzzles and their treatment in the literature

The data presented above pose several basic questions that a unified theory of
pronominal elements should be able to answer. For a start, we would like to know
how the morphosyntactic differences between strong pronouns and clitics can
be accounted for, while specifying how referential properties of strong pronouns
vs. clitics should be modelled. In relation to their referential properties, the ques-
tion arises how animacy is represented, as well as why clitics allow for sloppy
interpretations, and how the exceptions in PPs can be accounted for. This should
directly extend to the behaviour of pronouns in focus contexts.

All of the issues raised here have been discussed in relation to the categorial
status of the pronoun by being tied to the debate on whether nominal elements
in BCMS project a DP. Specifically, Despić (2011) and Runić (2014), among oth-
ers, argue that pronouns in BCMS are NPs. Some of their arguments come from
pronominal modification, argument ellipsis, the ability of clitics to license sloppy
readings, etc. Yet, Bešlin (in press) advocates for a parametrised view of nominal
categories in BCMS, under which lexical nouns are NPs, but pronouns are DPs
in this language. Part of her argument is based on pronominal modification and
the fact that Left-Branch Extraction of a nominal modifier is possible with a lexi-
cal NP but not with a pronoun. As we will see shortly below, using modification
of a pronoun as a diagnostic has shown to lead to inconclusive results, which
makes the parametrised view require closer scrutiny. Finally, some authors re-
ject the NP/DP distinction as a culprit for the difference in the behaviour of nom-
inal and pronominal elements altogether in BCMS, arguing that factors other

than the presence of articles in a language may be employed to explain some of Bošković's (2008) typological generalizations. For instance, Jovović (2024) does this for binding and Condition B violations present in BCMS (and absent in languages without articles), showing that the empirical picture is more complex and dependent on factors such as information structure, and not necessarily nominal size.

One way to resolve this puzzle is to apply tests in order to probe the structure of the pronominal phrase. Déchaine & Wiltschko (2002) argue that this structure can be threefold, namely pronouns may be mere NPs (Pro-NP), or DPs (Pro-DP), or of an intermediate size, which they term Pro-PhiP. Unfortunately, the tests provided in their work prove to be inconclusive for BCMS. For instance, for a pronoun to count as a DP, is should allow modification of the type *we linguists* or *you poor thing*, where the pronoun would be the overt realisation of the D head. BCMS pronouns do allow modification (see Progovac 1998, Bošković 2008, Despić 2011, Runić 2014, Arsenijević 2017, Bešlin in press for detailed descriptions, as well as Höhn 2016 on such constructions in general), as illustrated in (14).

(14)   *Modified personal pronouns*

    a. Dobri   ti   me   retko {zove   / zoveš}.
       good.M.SG 2.SG.NOM 1.SG.ACC rarely call.3.SG   call.2.SG
       'The good you rarely call(s) me.'   (Arsenijević 2014)

    b. Ja   volim   onog   tebe   kojeg poznajem.
       1.SG.NOM love.1.SG that.M.SG 2.SG.ACC who know.1.SG
       'I love that you that I know.'   (Pereltsvaig 2007: 28)

Nevertheless, as observed by Arsenijević (2017), the mere fact that pronouns can be modified in BCMS and in English is insufficient to diagnose the presence or absence of a DP layer. Arsenijević (2017: 13) argues (contra Bošković 2008, Runić 2014) that even English pronouns can be modified by adjectives (e.g. *Last night's him was so unlike the him that Sepi had first met*). And since they can be preceded by an article, this would indicate that they do not move to D, contrary to Déchaine & Wiltschko (2002). Moreover, Arsenijević (2014) recognises that there are semantic restrictions on the adjectives that can modify pronouns, such that only non-restrictive adjectives can combine with pronouns. Adjectives that are used restrictively can combine with pronouns only if the pronouns themselves semantically shift in interpretation, acquiring the interpretation of nouns (i.e. from type e to $\langle e, t \rangle$, as evident in the different agreement possibilities that such a pronoun can license, demonstrated in (14a)).

Furthermore, a Pro-DP behaves as an R-expression, while a Pro-PhiP behaves as a bound variable, which would qualify strong pronouns as DPs and clitics as PhiPs. However we have seen above that strong pronouns may license sloppy readings in PPs and act as bound variables in focus contexts, which would simultaneously make them PhiPs. Finally, according to Déchaine & Wiltschko (2002), a Pro-DP cannot be used as a predicate, but only as an argument. Clitics in BCMS can only be used as arguments (15), which would qualify them as DPs, while strong pronouns can appear in both contexts (15)–(16), which would make them Pro-PhiPs. However note that the very claim that DPs cannot function as predicates, put forward by Longobardi (1994), and followed by Déchaine & Wiltschko (2002) has been disputed in the literature (see for instance Pereltsvaig 2007: 21f. and references therein for Slavic).[5]

(15) Video   sam   {tebe   / te}.
     see.PRT.M.SG AUX.1.SG 2.SG.ACC CL.2.SG.ACC
     'I saw you.'

(16) Postala   sam   ti.
     become.PRT.F.SG AUX.1.SG 2.SG.NOM
     'I became you.'

There thus seems to be a lack of clear evidence on what category the pronominal elements could be, but more evidence favours their being PhiPs, than DPs. I will thus take an intermediate position, which is on the one hand, that the DP is not crucial to our understanding of the properties of personal pronouns, and on the other, that ϕ-features are one of their defining properties. As such, the DP will not play a crucial role in our analysis and will be left out of the pronominal structures proposed below (which will also be in line with recent proposals by Stegovec 2019, Ruda 2021a, but also the bulk of recent literature on the morphological realisation of pronouns advocated for by Moskal 2015b; Smith et al. 2019; McFadden 2018). Their PhiP status will prove to be convenient in accounting for the similarities and differences between strong pronouns and clitics. Eliminating the DP will require other ways to deal with their referentiality, but see Trenkić (2004), Stanković (2014a,b) on reference not requiring D in BCMS. The existence

---

[5]Cardinaletti and Starke (1999) argue for a tripartite distinction between strong, weak, and clitic pronouns; their tests are also insufficient – we could treat argument pronouns as strong and PP pronouns such as those in example (3) as weak (since they allow for inanimate referents, unlike strong pronouns in argument position), but they should also disallow coordination (see Bešlin in press and Despić 2011 for discussion and counterexamples).

of the DP in the structure and its location in relation to other phases will thus not be essential for the analysis.

Having established that ϕ-features are a crucial part of pronouns, we may further inquire about their exact structural encoding and relation to case and animacy features. Several works in the literature have tackled this issue, including Progovac (1998), Franks (2013), Despić (2017), Stegovec (2019), Caha (2021), Ruda (2021a). Assuming that they are distributed along the nominal spine, the consensus is mostly on a structure that involves an NP, followed by ϕ-features and case features on top of them, which I will follow, with some adjustments. As for animacy and humanness, they are tied to referential/individuation specification and also connected to natural and grammatical gender and number distinction, as well as person, which makes them generally problematic for the Y-model of syntax. They have been tied to person by Sichel & Toosarvandani (2024a,b), or to gender and classifiers by Harley & Ritter (2002), Puškar (2018), Puškar-Gallien (2019), Arsenijević (2021), or referential index (Stegovec 2019). Any successful analysis of the data presented above should be able to account for the optionality of animacy on clitics.

In what follows, I aim to provide an account of the properties of pronouns (animacy restrictions and sloppy readings) outlined above that will be based on a unified syntactic structure with well-defined locality domains.

4 Proposal: The internal structure of pronouns

In this section, I will outline a proposal for the internal structure of pronominal categories based on a combination of the feature geometry approach (Harley & Ritter 2002), the size of nominal phrase (Déchaine & Wiltschko 2002, Caha 2021), separate encoding of ϕ-features and predefined locality domains (e.g. Moskal 2015b, van Urk 2018).[6]

The general idea is that the (pro)nominal phrase consists of three general zones, a lexical one, followed by ϕ-feature-hosting projections, topped by case-bearing projections ([KP [ϕP [NP]]]). The ϕP will be further dissected into a person phrase (PersP), number phrase (NumP) and a gender phrase (ClassP). Finally, the case phrases will distinguish between unmarked, dependent, oblique, and prepositional case.

---

[6]Puškar-Gallien (forthcoming) offers a proposal on full syntactic decomposition of pronouns and their subfeatures, as well as their morphological realisation in the Distributed Morphology framework, which is why these will be largely put aside in the discussion below.

The base of the noun consists of a nominal root and a nominalizing head *n* (see Kramer 2015 and references therein). Following the claims of Moskal (2015a,b) and Smith et al. (2019) that the pronominal base crucially differs from the one of nouns in lacking a lexical root, I will treat the pronominal *n*P as consisting solely of the categorizing head *n* (van Urk 2018, building on Postal 1969, Elbourne 2005; but also Déchaine & Wiltschko 2002, van Koppen 2012).

4.1 Phi-features and their distribution

In analysing the syntactic representation of ϕ-features, I will rely on the proposal of Harley & Ritter (2002), who argue that ϕ-features have complex internal structure in the form of hierarchically organised sub-features. Their proposal is reproduced in Figure 1. An important aspect of the hierarchy is feature entailment. Having a deeper-embedded feature implies having the feature dominating it. For instance, if a pronoun has the feature [Addressee] from Figure 1, it will also contain the feature [Participant]. Such a structured geometric representation of morphological features, modelled after that of the phonological ones, is claimed to help constrain pronoun and agreement systems and present interdependence of features in a systematic way.
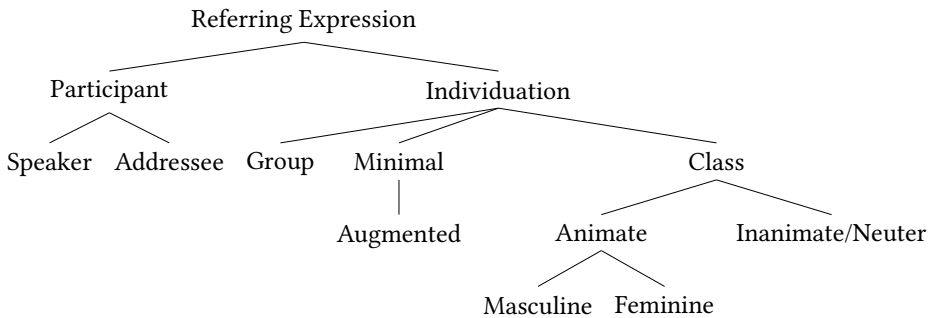
Figure 1: Structural hierarchy of ϕ-features (Harley & Ritter 2002: 486)

Accounts that distribute these features across the nominal spine have mostly focused on two types of features, person and number, or number and gender (see Béjar & Řezáč 2009, van Koppen 2012, Puškar 2018, Puškar-Gallien 2019, Caha 2021). I intend to offer a unified proposal for structural encoding of the hierarchy in Figure 1 within the nominal phrase that includes all the feature types present in it.

As a starting assumption, I take it that each feature type is hosted by a separate phrase. Taking the incremental bottom-up approach to syntactic structure building very literally, I interpret the root node of the pronoun, one that the entire hierarchy is built on (the "Referring Expression" in Figure 1), as the *n*P base. This models the idea that *n*P is responsible for the referentiality of the pronoun.[7]

Disagreement in the literature is present not only in the encoding of referentiality, but also in the encoding of individuation (another complex node in the hierarchy in Figure 1). Referentiality and individuation are connected such that reference taking and quantification are dependent on individuation (see e.g. Sichel & Toosarvandani 2024a), which differentiates nouns from other lexical categories (Baker 2003: 94-189). Individuation as a property has received different treatments in the literature. While Harley & Ritter (2002) separate it from person and make it a precondition for having number and gender features (cf. Figure 1), Sichel & Toosarvandani (2024a,b) employ a separate syntactic projection to encode this property, which to them mediates between person and animacy features and accounts for their interdependence. The locus of animacy is thus also a matter of debate, or rather crosslinguistic variability. It has been related to person (see also, e.g., Lochbihler et al. 2021), but also to gender by Foley & Toosarvandani (2022), or Puškar (2018), Puškar-Gallien (2019) for BCMS.

I follow Puškar (2018), Puškar-Gallien (2019) in assuming that individuation is related to animacy, both of which are a part of '.' Puškar (2018) integrates animacy into the representation of natural gender, which is argued to be located on '.' Encoding animacy as part of natural gender on *n* (as opposed to morphological gender which is higher in the structure, see below) correctly derives all available, and rules out unavailable patterns of agreement in BCMS such as hybrid agreement and Corbett's (1979) Agreement Hierarchy. Puškar-Gallien (2019) extends this to agreement with honorific pronouns by arguing that animacy is also an integral part of natural number, which is encoded together with natural gender on '.' They are located under a common node, labeled "IND", standing for "individuation". I will thus assume that individuation (in addition to referential index) is a property encoded on the nominal base. Recall that Baker (2003: 94-189) claims that individuation and reference taking differentiate nouns from other lexical categories. Distributed Morphology models this difference by building different

---

[7]Precursors for this idea include Caha (2021), who models RefP as an additional syntactic projection above the *n*P, albeit without providing much detail on its purpose or interpretation. Sichel & Toosarvandani (2024b) use a more abstract σP for individuation purposes, while Ruda (2021a,b) utilizes a PersP. See also Stegovec (2019), who employs a (morphologically) empty node *Index* to introduce the referential index on the pronoun. This node is assumed to be higher in the structure.

categories on different categorising heads (and sharing their extended projections). Making *n* responsible for individuation and reference thus models this connection. More concretely, I will assume that individuation is dependent on properties such as [animate] and [human], which can appear as features of the pronominal base.[8]

Disassociating individuation from number and gender requires a reorganisation of the hierarchy in Figure 1 such that it can ultimately be encoded in terms of syntactic phrase structure. That person features reside lower than number features has been argued by Noyer (1992), Trommer (2002), Harbour (2007, 2008a, 2016), Arregi & Nevins (2012). Their argument comes from the ordering of person and number affixes, where it was noticed that person affixes strongly tend to be linearised closer to the stem of the word, and number affixes further from them. Under the Mirror Theory (Baker 1985, Brody 2000, Brody & Szabolcsi 2003), this points to a lower base position of person with respect to number. Additionally, under Harbour's (2016) theory of person and number encoding, person being introduced higher than number makes wrong predictions for possible and impossible pronoun inventories. Following van Urk (2018); Smith et al. (2019), I assume π to be local to the pronominal base. I take person to head its own projection, πP, above the *n*P, following recent proposals of Ruda (2021a) for Polish and Stegovec (2019) for Slovenian. Specifically, I assume that 1st person comprises the features [π, Participant, Speaker], 2nd person lacks the [Speaker] feature and 3rd person is represented by the person [π] node alone, as illustrated in figure Figure 2 below.

Number heads a projection further up, which I will label as #P (Picallo 1991, Bernstein 1993, Borer 2005, Acquaviva 2009, Harbour 2008b). Since BCMS has a simple binary number system, it suffices to assume that it includes the generalised feature [#], which can have a [PL] feature as its dependant. Singular will be treated as the absence of number (Nevins 2011, Pesetsky 2013; see Despić 2017 for a claim that singular number is unmarked with respect to plural in Serbian). Technically, #P will be postulated only in case it specifies plural number, i.e. #P is not projected if the noun is singular (Kratzer 2007).

Grammatical gender heads its own projection CL(ass)P above #P. Here, CLASS will be used as mnemonic for gender, which admittedly has more complex structure and whose further modelling is outside of the scope of this paper. I will simply assume that CLP hosts the morphologically realised GENDER. In locating morphological gender above number I also follow Puškar (2018), Puškar-Gallien (2019), who argues that this constellation is indispensable for BCMS in order to

---

[8] Puškar-Gallien (forthcoming) offers a revision of this model and provides further detail on how animacy and humanness can be encoded on the *n* base.

derive the variability of agreement patterns found with different nominals. This position of number in between grammatical gender and individuation (in her case natural gender and number) has a blocking effect on agreement, which can derive agreement mismatches of nouns such as *vladika* 'bishop', which agree as masculine in the singular (natural gender), but as feminine in the plural (grammatical gender). This way of modeling gender is also a precondition to deriving all other agreement patterns in the language.[9]

To sum up the discussion thus far, Figure 2 presents the proposal for the basic pronominal functional spine in BCMS. I assume that the features themselves are the syntactic heads that project the corresponding phrases. These features can also include a small hierarchy of sub-features below them.[10]

4.2 Case features and their distribution

Following Bittner & Hale (1996), Caha (2009), Neeleman & Szendrői (2007), Moskal (2015a,b), Smith et al. (2019), I assume that case is introduced by a separate projection K(P). K can have a complex structure that encodes Caha's (2009) *Case Hierarchy*:

---

[9] GENDER as a category can be dispersed across the nominal spine. For the distinctions in encoding grammatical and natural gender see Steriopolo & Wiltschko (2010), Pesetsky (2013), Landau (2016), Kučerová (2018), Steriopolo (2018a,b), Fassi Fehri (2018), but also Arsenijević (2021) for an alternative view, and in particular Puškar (2018), Puškar-Gallien (2019) for arguments why natural gender must be located lower in the structure.

[10] One necessary addition to this model is the representation of natural gender on '.'. I assume that it additionally involves a feature [CL] and a feature [F] as its dependant. This directly links gender and the features [ANIM] and [HUM]. For instance, nouns of feminine natural gender will involve all of the available nodes in the hierarchy: [CL[ANIM[HUM]][F]], while grammatically feminine nouns will lack the animate and human specification, leaving them with [CL[F]]. Nouns of masculine grammatical gender will only involve the [CL] node, as an unmarked gender feature. Masculine natural gender will involve the [ANIM] and [HUM] features as well, accounting for the general bias in language under which the default referent of human nouns is male Finally, the absence of the [CL] node signals the absence of gender, thereby modelling neuter gender. As such, gender can also participate in agreement, as 1st and 2nd person pronouns control natural gender agreement.

(i)
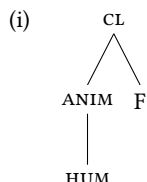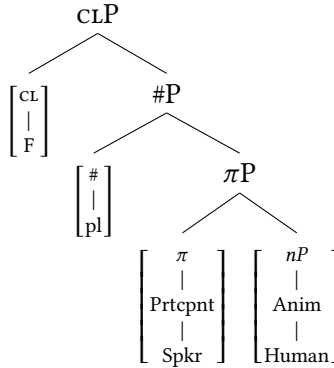
Figure 2: Basic pronominal functional spine in BCMS

NOMINATIVE ≻ ACCUSATIVE ≻ GENITIVE ≻ DATIVE ≻ INSTRUMENTAL ≻ COMI-TATIVE. Smith et al. (2019) simplify this somewhat by assuming a distinction between the DEPENDENT CASE (DEP; here encompassing ACC and GEN) and the OBLIQUE CASE (OBL, here DAT). To this I add the assumption that BCMS also includes two cases that are realised as prepositional phrases, namely INSTRUMENTAL and LOCATIVE (see Milićev & Bešlin 2019 for instrumental in BCMS; the assumption on locative is straightforward for BCMS, as it is always syncretic with dative and obligatorily preceded by a preposition).

(17)   [$_{PP}$ P [$_{K_{OBL}P}$ K_{OBL} [$_{K_{DEP}P}$ K_{DEP} [$_{K_{UNM}P}$ K_{UNM} [$_{CLP}$ CL [$_{\#P}$ # [$_{\pi P}$ π [$_{nP}$ n]]]]]]]]

To the structure above McFadden (2018) adds the proposal that NOM is the absence of case (built on Bittner & Hale 1996, McFadden & Sundaresan 2009, i.a.), which he models as the absence of the case-bearing projection(s). This eliminates K_{UNM}, leaving nominative pronouns without any case projections.[11]

---

[11]Modelling case features closely follows the assumptions from nanosyntax on the containment of case projections. A reviewer notices though that KP layers differ from the other layers in the NP as they are interdependent. In order to streamline the nature of the projections, it can be assumed that KP is projected by the feature [DEP], thus KP would only be present when the feature [DEP] is. Other case features, such as [OBL] may be introduced as sub-features of [DEP], such that the case hierarchy is present within the head node on this projection, just like with ϕ-features. This would model the dependence of oblique case on the dependent case, as well as the absence of case in the nominative. See Bárány (2017) for a similar approach.

4.2.1 Interim summary

To sum up, Figure 3 represents the complete structure of a BCMS nominal phrase in the most complex case. This provides a way to distribute the Harley & Ritter (2002) hierarchy across the pronominal spine (see also van Koppen 2012, Fassi Fehri 2000).
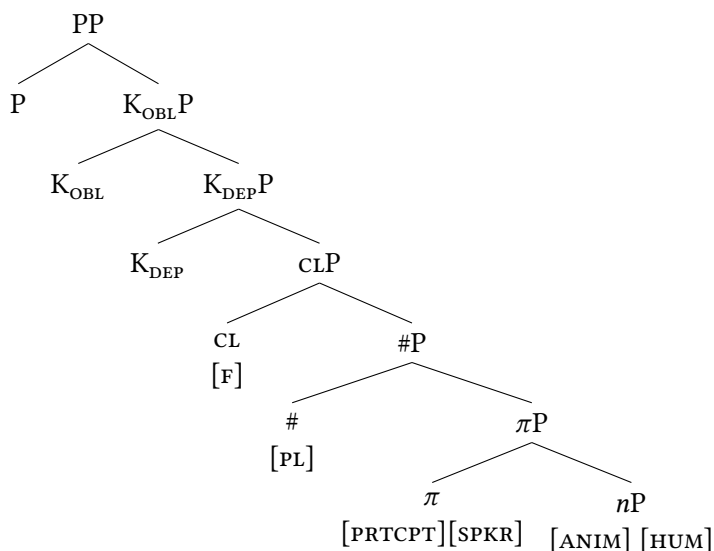
Figure 3: Proposed model of the structure of a BCMS pronoun

4.3 The representation of pronoun types

The complete structure of a pronoun given in Figure 3 offers possibilities for parametrisation, as not all pronouns will require all the available nodes. I propose that local-person pronouns lack CLP in general, which models the lack of grammatical gender. Their singular forms also lack #P. The πP is projected, since they must have at the minimum the [PRTCPT] feature. The structures in Figures 4–5 represent the local-person pronouns in the nominative case (hence the lack of KP). First person pronouns differ from second person ones in having the additional [SPKR] feature.[12]

---

[12]A reviewer wonders how local-person pronouns can control gender agreement without having overt grammatical gender features. Recall from Section 4.1 and footnote 10 that I assume that natural gender is present on the *n*P of local-person pronouns, following Puškar (2018), Puškar-Gallien (2019). From there it can enter agreement relations.

This structure offers additional possibilities for parametric variation. While BCMS does not show gender distinctions on local person due to an assumed lack of CLP, Slovenian does contain this phrase and consequently distinguishes feminine (*m-e* '1-F.PL') and masculine (*m-i* '1-M.PL') versions of local person. Notice that Slovenian incidentally offers evidence for ordering person before number and gender, as the gender and number portmanteau follows the person morpheme.[13]
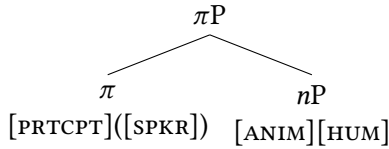
Figure 4: Singular local-person pronoun
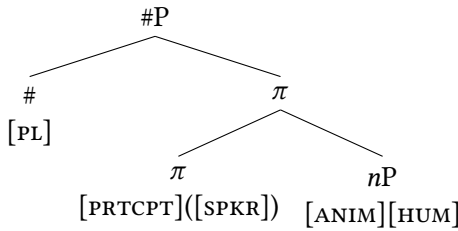
Figure 5: Plural local-person pronoun

The proposed structures for 3rd-person pronouns are presented in Figures 6–7. In the singular, due to the absence of number, their *n*P will be dominated by *π*P and CLP, which bears the [F] node for grammatically feminine nouns or just the [CL] node for masculine ones. In the plural, the CLP will be projected above the #P. The combination of these two phrases will define the inflectional affixes of the pronouns. The *n*P lacks features if the pronoun denotes an inanimate entity. With an animate (or human) referent, these features will be present on the *n*P.

The system proposed above may be extended straightforwardly to other languages of the Slavic family. As for further extensions to possible and impossible pronominal systems, the proposal would make similar predictions as those

---

[13]Alternatively, we may assume grammatical gender to be universally present and that it gets deleted under Impoverishment in local person contexts, as suggested by Noyer (1992) for Arabic, or Despić (2017) for Serbian.
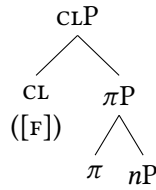
Figure 6: Singular 3rd-person pronoun
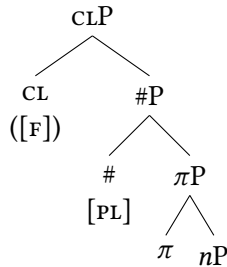
Figure 7: Plural 3rd-person pronoun

made by Harley & Ritter (2002) under the assumption that what they call "activation" of a particular node is implemented as the presence of that node in the syntax. Just like their model, my model keeps person and number features separate, and the variation in pronominal systems depends on the activation of the (sub-)hierarchies of these nodes. If the two nodes [Participant] and [#] are activated together, their combination may yield particular types of person, such as those with inclusive/exclusive distinctions. According to them, the presence of particular features in the pronominal hierarchy may be motivated by the presence of a feature in other areas of grammar too. E.g. Pirahã, Maxakalí and Kwakiutl do not show number distinctions and consequently do not make use of the Individuation node in their hierarchy. Thus in my system a language that makes person and number distinctions would project πP and #P, whose subnodes would further model distinctions such as inclusive/exclusive, paucal, etc.

As for gender, Harley & Ritter admit that the CL node in their hierarchy would need further modelling and elaboration due to wide crosslinguistic variation in the representation of gender features. They note that "1st or 2nd person features should combine freely with any of the number and gender features, since the latter are dependents of a separate organizing node" (Harley & Ritter 2002: 508).

Representation of gender across different (lexical and functional) categories, interaction of gender with other ϕ- and and case features and interaction of gender with animacy and humanness is thus a task under current research that is outside the scope of this paper.[14]

4.4 A note on the morphological realisation of strong pronouns vs. clitics

The general intuition that I would like to outline here is that the spell-out rules for local-person pronouns target the base and ϕ-features together, whereas in third-person pronouns, the base is spelled out separately from the inflectional affixes, cf. Figures 8–9. This is what in principle makes third-person pronouns similar to nouns. The spell-out rules will have to be made more precise in order to be able to account for the suppletion patterns presented in Section 2.1, however this is outside the scope of the current paper.
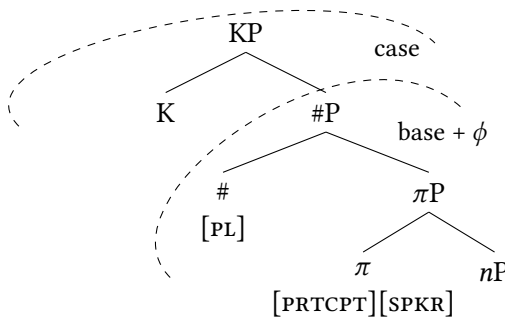
Figure 8: Local person

We will furthermore see that spelling out nP independently, i.e. effectively deleting it, is what enables a certain amount of flexibility to clitics that strong pronouns lack. Under the assumption that the nP is a locality domain and as

[14]First steps of further research involve a crosslinguistic study of pronouns that show gender distinctions on local person. So far, I have identified 54 languages with gender on local person, belonging to 18 families and 2 isolates, based on the World Atlas of Language Structures (Siewierska 2013). My system predicts that in polymorphemic pronouns, gender should follow person and number, and languages that conform to this include Andi, Arabic, Berber, Bora, Djeebbana, Gagadu, Nama, Provencal, Spanish, Lithuanian, Slovenian, Korana. Other candidates to be studied further include Aramaic, Beja, Coptic, Zari, Paez, Sha, Baniata, Dumo, Murui Huitoto and Tunica. This sample should offer further insight into feature entailment relations by identifying patterns of gender encoding and its limitations.
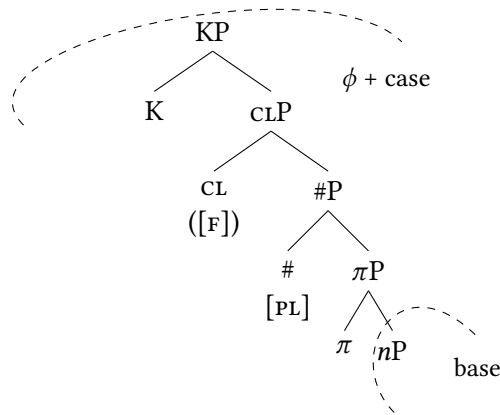
Figure 9: 3rd person

such it is transferred to the interfaces independently of the rest of the structure, the remaining structure is spelled out in the next cycle as a clitic. Figures 10–13 illustrate the part of the structure that gets realised as a clitic after *n*P deletion. I will build on this below in exploring the syntactic consequences of the given structures.
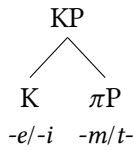
Figure 10: Local person clitic singular
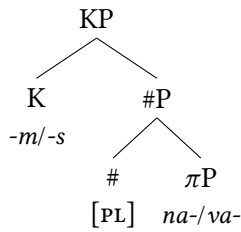
Figure 11: Local person clitic plural

KP
K CLP
CL πP
-ga/-mu/-je/-joj/-ju
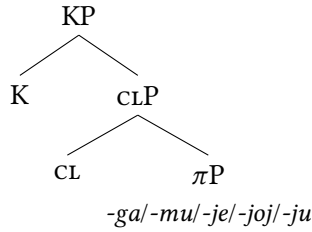
Figure 12: 3rd person clitic singular

KP
K CLP
CL #P
πP
[PL] -ih/-*im*
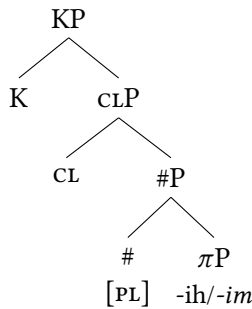
Figure 13: 3rd person clitic plural

To sum up, what unifies strong pronouns and clitics is their internal structure, which can be parametrised. What differentiates strong pronouns from clitics is the presence of the *n*P, such that with clitics it is not realised.

4.5 Consequences for animacy and referentiality

The proposal above has direct consequences for the interpretational properties of pronouns presented in Section 2.2. Since clitics lack the *n*P, and with it the animate and human features, they are in principle compatible with either interpretation. Recall that clitics also behave as bound variables, which allows for sloppy readings and the ability to be bound. Due to the lack of *n*P, they also lack strict reference, and are thus more flexible.

Before continuing on to the syntactic consequences of this proposal, a comment on the interpretation of ϕ-features is in order. As interpretable features, ϕ-features have been widely assumed to trigger presuppositions (Cooper 1983, Heim 2008, Kratzer 2009, Jacobson 2012, Sudo 2012). Pronouns carry a referential

index which determines their interpretation (e.g. speaker, hearer, participant in a speech act), and ϕ-features, which are considered to introduce presuppositions to the values provided by the index (see Sauerland 2013). Even though presuppositions triggered by free and bound pronouns may differ in some aspects, they have been subject to unified analyses (see Sudo 2012, Sauerland 2013).

Since I treat animacy as a part of natural gender, I will follow Merchant (2014); Murphy et al. (2018); Sudo & Spathas (2020), Arsenijević (2021), all of whom assume that natural gender features trigger presuppositions on the gender of the referent, although they differ in their treatment of grammatical gender (no presuppositions by Merchant 2014, Murphy et al. 2018, presuppositions but no assertions by Sudo & Spathas 2020, or weak presupposition by Arsenijević 2021). Arsenijević (2021) and Arsenijević et al. (2022) argue that features like [human] can also be presupposition triggers in BCMS, mostly in conjunction with and in relation to gender. In particular, they argue that [human] contributes to intepretation of gender by triggering a moderate male presupposition (due to cultural bias). In principle, the absence of a gender presupposition (or an assertion thereof) makes a noun compatible with either male or female referents. In the same vein, we can assume that the absence of animacy and humanness information on the *n*P leads to a pronoun's compatibility with both animate and inanimate referents. This would mean that the deletion mechanism proposed below applies at LF as well. I will leave further formalisation of this for future research and explore some of the technical consequences below.

5 Consequences for syntax and interpretation

This section explores the syntactic consequences of the structures proposed above. In particular, I will argue that the availability of sloppy readings of strong pronouns is related to their inability to move out of the PP. Section 5.1 explores the general properties of movement of (pro)nominal elements, and Section 5.2–Section 5.3 develop an account on the interactions of this movement with the pronominal structure and its locality domains.

5.1 Pronoun movement

Recall that if a pronoun follows a preposition, it can only appear in its strong form, no clitics are allowed, as illustrated above in (2). Yet such strong pronouns in the complement of PP show clitic-like behaviour: They may be inanimate and allow for sloppy readings, as illustrated by examples (3)–(4) and (8)–(9) above. I

will argue that such clitic-like behaviour of pronouns in this context is due to a ban on movement out of the PP.

As a starting point, let us examine the general behaviour of (pro)nominal elements in BCMS with respect to movement. Unlike nouns, pronouns in BCMS have been argued to move outside of the VP, as illustrated in (18a) for pronouns and (18b) for nouns. As Bešlin (in press: 3) suggests, a potential context for (18a) could be something like 'When will Mary meet John next?'. A lexical NP may move, with an effect on its interpretation (the moved instance of *Jovan* in (18b) is topical, while the postverbal *in-situ* one is new information focus, as reported in Bešlin in press). Clitics in BCMS are also known to undergo movement to the second position in a sentence (18c) (see Bošković 2001, 2004, Talić 2018).

(18)   *Pronoun movement*

    a. Marija {njega}   sreće {?*njega}   svaki dan.
       Marija 3.M.SG.ACC meets   3.M.SG.ACC every day.
       'Marija meets him every day.'

    b. Marija {Jovana} sreće {Jovana} svaki dan.
       Marija Jovan   meets Jovan   every day.
       'Marija meets Jovan every day.' (Stojanović 1997: 307; Bešlin in press)

    c. Marija {ga}   sreće {*ga}   svaki dan.
       Marija CL.3.M.SG.ACC meets CL.3.M.SG.ACC every day.
       'Marija meets him every day.'

Based on the position of the pronoun relative to adverbs and negation, Bešlin (in press) proposes that the landing site of the moved pronoun is somewhere in the middle field, between *v*P and TP (19b). Although the movement of clitics is further affected by phonological considerations such as second position in a prosodic word (see Talić 2018 and references therein), assuming that clitics behave like pronominal elements, they should be able to move at least as high as strong pronouns otherwise do. Since the exact position to which the pronominal elements move is not crucial for the further discussion, it will be left for further research.

(19)   *Pronoun movement*

    a. Marko (juče)   ni-je   {NJU   / nju}   mudro
       Marko yesterday NEG-AUX.3.SG 3.SG.F.ACC 3.SG.F.ACC wisely
       savetovao.
       advise.PRT.M.SG
       'Yesterday, Marko did not advise { HER / her} in a wise manner.'

b.   [$_{TP}$ yesterday [$_{TP}$ NEG-AUX [$_{XP}$ HER/her$_i$ [$_{vP/VP}$ wisely [$_{vP/VP}$ advised t$_i$]]]]]   (Bešlin in press: 6)

Proposals on the trigger for such a movement include semantically-triggered object shift (moving out of the VP to avoid existential closure and receive a definite interpretation; Stojanović 1997), or categorially-driven movement (pronouns, unlike lexical nouns, are DPs and as such have to move to Spec, AgrOP to check the D-feature, Bešlin in press). Although the source of the trigger requires more elaborate research, it seems to me that the most probable explanation is the one that Bešlin (in press) rejects, namely information structure. Even though in (18a) it is argued that the interpretation of the pronoun is neutral (under the context assumed by Bešlin, the pronoun should refer to the topic of the previous discourse), compared to (18b), the strong pronoun still carries some sort of contrastive interpretation. Thus whereas focus might not necessarily be at play, some sort of contrast is definitely involved, as for instance in a contrastive topic. And these may require movement in BCMS. I will leave this issue for further research and come back to it briefly below in Section 5.3.

5.2 Pronouns in PP position

5.2.1 Assumptions

Having established that pronouns as complements of verbs move from their base position, we may extend this to pronouns in general, including those that are in the complement of P position. However, with the latter this movement will be blocked by the preposition. Below I will argue that this is exactly what leads to inanimate interpretations and sloppy readings in these specific contexts.

I will largely build my account on van Urk's (2018) proposal for pronoun copying, based on pronoun copying in Dinka Bor (Nilotic).[15] This language allows constructions in which a pronoun doubles a noun or another pronoun. This poses the challenge of having multiple copies of the same element in a sentence (as for instance in constructions with multiple copies of a verb that has undergone movement, see Abels 2001 for Russian, Landau 2006 for Hebrew). What is more, a mismatch can happen as in (20). Both examples involve an overt copy of a fronted object pronoun, realised as the 3.PL *kêek*. This pronoun matches the fronted pronoun only partially – in number, but not in person.

---

[15]See also Bošković (2001) for a copy-based account of clitic placement in Serbo-Croatian.

(20) a. wɔ̂ɔk c�came bôl   {kêek / *wɔ̂ɔk} t̪íiŋ
          1.PL PRF.OV Bol.GEN 3.PL   1.PL   see.INF
          'Us, Bol has seen.'

      b. wêek c̣íi   bôl   {kêek / *wêek} t̪íiŋ
          2.PL PRF.OV Bol.GEN 3.PL   2.PL   see.INF
          'You all, Bol has seen.'   (Dinka Bor; van Urk 2018: 940)

Van Urk (2018) thus needs to account for pronoun movement and multiple-copy spellout. Building on Landau (2006), van Urk's analysis employs the copy theory of movement and a spellout algorithm that enables prononuciation of multiple copies. There are two conditions on copy-spellout, namely recoverability and economy. Recoverability requires that a copy be pronounced if it is associated with phonetic content and economy ensures that as little structure is spelled out as possible, amounting to one copy in a chain ("all unique phonetic content is realised at least once"; van Urk 2018: 964). Association with phonetic content is met either if an item has its own phonetic content, or if it appears in a position specified with some phonological requirement (Landau 2006: 31). These two conditions normally ensure that only one copy in a chain is pronounced and the others deleted. The spellout of multiple copies in Dinka is motivated by the peculiarities of phonological requirements related to the EPP features on *v*P and CP edges, which was taken to be a matter of parametric variation.

In a movement chain some copies will undergo full deletion (a precondition on deletion is that a unit must be a phase). For pronouns, van Urk also proposes a so-called PARTIAL DELETION. The *n*P may be a phase, which is taken to be a cross-linguistic parameter, and as such it can undergo copy deletion independently of the rest of the NP. The deletion operation includes the phase head as well, see van Urk (2018: 968f.). Deleting the *n*P thus leaves the rest of the projections in the pronoun intact, which results in a partial copy, including KP and NumP in his case. Since person information gets deleted together with *n*P (the locus of π under his account), the remaining copy need not match in person. In my account below, deleting the *n*P will exactly amount to spelling out a clitic, and I will assume that deleting the *n*P also deletes all of the contents of its sub-hierarchy.

5.2.2 Derivation

Following van Urk (2018), I will assume that pronominal *n*P in BCMS is a phase. I also assume that the target for movement and copying is the KP as in Figure 14. This ensures that only objects move. The pronoun moves through the edges of phases, stopping (at least) at the *v*P edge. Such a movement operation may create

multiple copies, some of which must be deleted. I posit that the difference in whether we will get a strong pronoun or a clitic depends on the phonological requirements related to their landing sites (e.g. if a pronoun is in a focus position, *n*P will be realised, resulting in a strong pronoun; if it is in a topical position, it will be deleted, resulting in a clitic). As a result of partial deletion, only the structure between *n*P and the highest KP gets realised, but not the *n*P itself. In my system this amounts exactly to a realisation of a clitic, as illustrated in Figure 14.
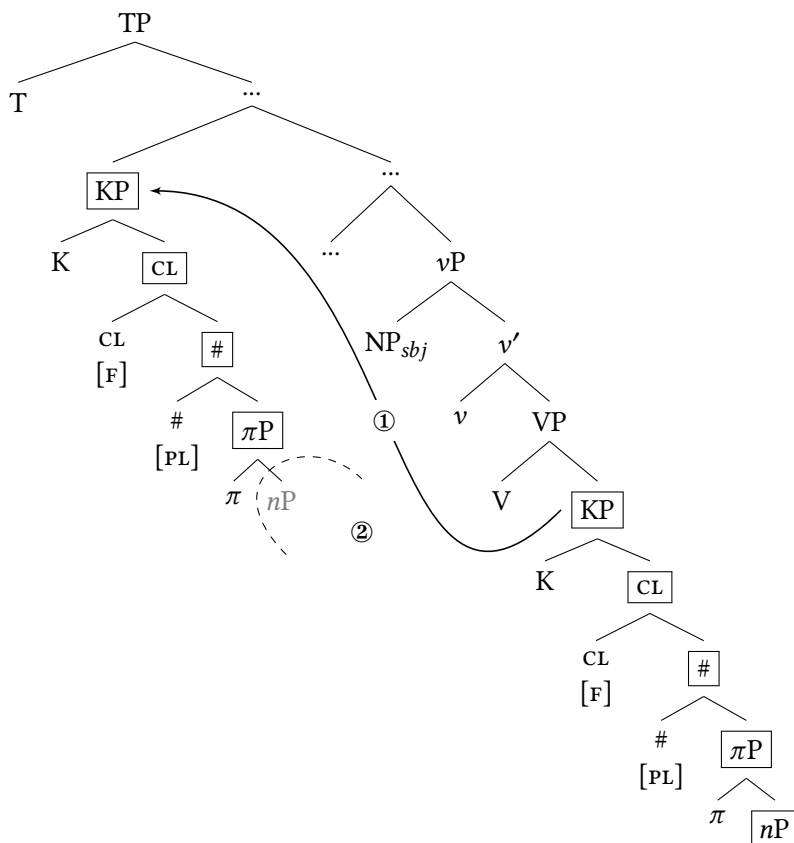
Figure 14: Pronoun movement, resulting in a clitic (here e.g. 3.ꜰ.ᴘʟ)

The deletion of the *n*P makes the animacy and humanness features unavailable, leaving the clitic more flexible in terms of its interpretation by virtue of lacking the individuation information.

Applying the process above to pronouns in the complement of PP position will result in the preposition blocking the first step of the process. Assuming that PP is a phase, I will take the cause of the impossibility of extraction to be antilocality (Abels 2012, Milićev & Bešlin 2019). The moved pronoun would have to pass through the Spec, PP position, which is too short a movement step. This will in turn enforce the spellout of the full pronoun.

(21) PP blocking movement
$$[_{PP} <KP> P [\boxed{_{KP}} K [\boxed{_{CLP}} CL [\boxed{_{\#P}} \# [\boxed{_{\pi P}} \pi [_{nP} n]]]]]]$$
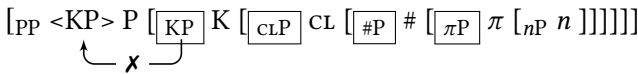
As a result, due to an inherent lack of stress on the prepositions under discussion, a clitic remains without a phonological host (see e.g. Talić 2018) or the possibility to move. The spellout of a strong pronoun may in this case be thought of as a last-resort strategy due to recoverability in order to satisfy the phonological requirements within the PP. As a result, the nP must be realised, and exactly in these contexts the pronoun can also be inanimate and have a sloppy reading (8) (i.e. formally a strong pronoun may functionally be a clitic). As an extension, if instrumental and locative are treated as PPs instead of KPs (e.g. Milićev & Bešlin 2019 for instrumental, or Stegovec 2019 for Slovenian), the behaviour of their complement pronouns (inanimate reference and sloppy readings, as in other PPs) follows automatically.

A further benefit of this analysis is that a clitic need not be animate or human, since those features remain stranded on the nP base and undergo deletion with it. A clitic may also act as a bound variable since the projections that are responsible for establishing reference are missing (see also Ruda 2021a,b for a claim that PersP is responsible for specificity and definiteness, which is absent in pronouns with a non-specific reading; on reference not requiring D in BCMS, see Trenkić 2004, Stanković 2014a,b, Arsenijević et al. 2022). In addition to this, the position of the DP in the structure is not crucial for the analysis.[16]

---

[16]The final issue is the nature and timing of the copy-deletion process. Van Urk (2018: 968) entertains the possibility that deletion may be seen as non-Transfer, under the assumption that Transfer applies to phasal units (e.g. as in Fox & Pesetsky 2005). He admits that this view raises an operation-ordering issue in terms of timing of Transfer and copy deletion, as copy deletion would have to precede Transfer, even though it is assumed to be a PF operation. He also admits that there is an issue of how long the copies actually have to stay visible in the derivation in order to evaluate which one in the chain will be spelled out. Adopting this premise would require that deleting the nP essentially means that it avoids Transfer to PF and LF. The absence of the features [ANIMATE] and [HUMAN] would allow for a more flexible interpretation since they cannot trigger presuppositions on the referent. PF would still need to have access to the nP somewhat longer though, at least until the next phase head is merged. This would result

5.3 Pronouns in focus position

This section provides a brief discussion on the extensions of the analysis above on pronouns in focus constructions. Recall that in BCMS only strong pronouns may express contrastive focus (or require a focused antecedent), while clitics are topical elements. We assumed above that if a strong pronoun is present in a context where a clitic is usually banned (PPs, focus contexts), such pronouns can be treated as clitics in disguise (Despić 2011: 244).

Under my proposal, the presence of focus on the pronoun should somehow be able to prevent the deletion of the *n*P or enforce its phonological realisation. Recall from examples (12)–(13) from Section 2.2.3 that a pronoun can be focused either by being in a particular position in a sentence (e.g. at the beginning or at the end) or by appearing with a particle. In the former case, under the account above, the focus position would impose a PF requirement that the element in this position must carry stress, thus a strong pronoun will be realised, as per recoverability and economy principles.

If a pronoun appears with an element that carries stress, as in example (13) above, one way to implement this technically is to assume that a pronominal phrase may include an additional functional layer, an FP, which may serve as a landing site for the movement of the clitic, as proposed by van Alem (2025). Van Alem justifies this by the existence of nouns with focus particles in Dutch, which can be accounted for under this kind of structure. This FP essentially adds focus to the DP and provides an escape hatch for the clitic to move through. If Spec, FP is already occupied by the focus material, the clitic cannot move out. Instead, it has to be pronounced *in situ*, which has different effects in different Dutch dialects. Despić (2011: 217) proposes a similar analysis especially for examples like (13) which include an overt focus element, such as the intensifier *sam*, although in his account this element projects its own phrase above the nominal projections. See Despić (2011) for further examples and discussion.

Applied to the case at hand, the specifier of the FP above KP introduces focus material, such as the intensifier *sam* (22), which would disable the movement of the KP. As a focus environment, just like a PP, requires a strong pronoun, the *n*P will have to be pronounced as last resort. Note that in the absence of a DP, movement of the KP to Spec, FP would also independently be banned due to antilocality (Abels 2012).

(22) FP blocking movement

---

in the possibility of realising the nP within the PP phase due to recoverability and economy, while the animacy features would be inaccessible.

$$[_{FP} \text{ XP F } [_{\boxed{KP}} \text{ K } [_{\boxed{ClP}} \text{ CL } [_{\boxed{\#P}} \# [_{\boxed{\pi P}} \pi [_{nP} n]]]]]]$$

Recall that sometimes it is not strictly focus, but some sort of contrastive interpretation that is also involved in these kinds of structures. I will tentatively assume that such constructions involve the same kind of structure as presented in (22), however further research is necessary to establish their exact nature.[17]

6 Conclusion and outlook

The aim of this paper was to develop a unified model of the form and structure of pronominal elements in BCMS in order to account for a wide set of their distributional properties, including morphological realisation, animacy restrictions, ability to function as bound variables, and the distribution in focus (and contrastive) contexts. In addition to presenting an overview of the data available in the literature on these various properties, I have introduced novel data that show that strong pronouns in the complement of PP position may be inanimate, and may allow for sloppy identity readings, contrary to expectation. The data are based on an informal survey, but nevertheless suggestive of the flexibility of the strong pronouns that has previously been overlooked.

I have argued that the behaviour of strong pronouns in PPs and focus contexts in terms of allowing for animate referents and bound variable interpretations makes them more clitic-like in these contexts. The mismatch between their form and distribution was resolved based on a proposal for their unified syntactic structure and restrictions on morphological realisation, based on a particular theory of pronominal copying.

---

[17]As noted by a reviewer, Slovenian clitics differ from BCMS ones. For instance, they can stand alone as answers to polar questions, and they can carry stress and appear in focus positions (see Dvořák 2007 for a full spectrum of variation and peculiar behaviour of Slovenian clitics). I would nevertheless expect them to behave the same in terms of animacy restrictions and sloppy readings, given their clitic status. The locus of variation would lie in the phonological requirements on the realisation of stress, such that in Slovenian it can be carried by the clitic itself, while in BCMS the realisation of the base is unavoidable. On the other hand, Slovenian makes use of a further type of pronouns such as *zá_nj* 'for him', which make use of the pronominal base in a PP, with a shift of the stress from the base onto the preposition. Note that it is not so clear-cut what portion of structure these pronouns actually involve, since the feminine version is syncretic with the strong pronoun *zá_njo* 'for her' (P-pronoun) vs. *za njó* 'for her' (PP). I will leave this issue as an avenue for further extension (Stegovec 2019 analyses these as lacking a referential index and the KP layer).

One of the main contributions of this paper is the proposal for a decomposed structure of pronominal elements in BCMS, that is applicable to other Slavic languages, but potentially also wider. I have argued that all pronouns are based on an *n*P, followed by ϕ-feature-bearing projections, such that person is local to the base, number follows it and gender tops them both ([CL [# [π]]]). These are followed by case-bearing projections, of which the nominative one is missing, and the others encode DEPendent case below OBLique one. Crucially for us, the features [ANIM] and [HUM] are encoded on the *n*P, and as such tied to individuation and referential properties of pronouns.

As a direct consequence, in case that the pronominal base undergoes deletion, the remaining structure becomes more flexible in terms of its interpretation. Specifically, leaving out the *n*P leaves us with a clitic, interpreted as either animate or inanimate, and either sloppy or strict. The deletion of the *n*P was implemented using van Urk's (2018) theory of pronominal copying. A benefit of this analysis was that cases where the *n*P had to be realised due to phonological reasons (PPs and focus/contrastive contexts) were exactly those in which strong pronouns show clitic-like behaviour. Another benefit of the approach is that it allowed us to treat locative and instrumental as PPs in BCMS, based on the parallels in the behaviour of strong pronouns between them and other cases.

One issue that remains open concerns dative clitics and sloppy readings. In particular, Runić (2014) notices that in BCMS only accusative clitics allow for sloppy identity readings, while with dative clitics this is impossible. We have however seen that strong pronouns in the complement of a preposition that inherently assigns dative case do not face such a restriction. One way to account for this may be to assume that the K_{OBL} phrase functions as some sort of a locality-domain-determining phrase and as such also restricts the interpretation of dative clitics. This issue will be left for further research. In addition to that, the next steps would include validating this proposal based on the data from other Slavic languages, as well as a broader range of crosslinguistic data.

Abbreviations

1	first person	LOC	locative
2	second person	M	masculine gender
3	third person	N	neuter gender
F	feminine gender	NEG	negative
ACC	accusative	NOM	nominative
ANIM	animate	OV	Object Voice
AUX	auxiliary	PL	plural
CL	class	PRF	perfect
CL	clitic	PRT	participle
DAT	dative	PRTCPT	participant
GEN	genitive	REFL	reflexive
HUM	human	SG	singular
INANIM	inanimate	SPKR	speaker
INF	infinitive	π	person
INS	instrumental	#	number

Acknowledgments

References

Abels, Klaus. 2001. The predicate cleft construction in Russian. In Steven Franks, Tracy Holloway King & Michael Yadroff (eds.), *Formal approaches to Slavic linguistics (FASL) 9*, 9–19. Ann Arbor: Michigan Slavic Publications.

Abels, Klaus. 2012. *Phases: An essay on cyclicity in syntax*. Berlin: de Gruyter.

Acquaviva, Paolo. 2009. Roots and lexicality in Distributed Morphology. In Alexandra Galani, Daniel Redinger & Norman Yeo (eds.), *York-Essex morphology meeting 2*, 1–21. York: University of York.

Arregi, Karlos & Andrew Nevins. 2012. *Morphotactics: Basque auxiliaries and the structure of spellout*. Dordrecht: Springer.

Arsenijević, Boban. 2014. The agreement of adjectives with the honorific pronoun in Serbo-Croatian. *Philologia Mediana* VI(6). 39–50.

Arsenijević, Boban. 2017. What the prosody of Serbian short and long form adjectives tells us about the nominal structure. In Gordana Djigić & Sofija Miloradović (eds.), *Godišnjak za srpski jezik*, 9–21. Niš: Faculty of Philosophy, University of Niš.

Arsenijević, Boban. 2021. No gender in "gender agreement": On declension classes and gender in Serbo-Croatian. *Balcania et Slavia* 1(1). 11–46. DOI: 10.30687/BES/0/2021/01/001.

Arsenijević, Boban, Ivana Mitić & Zorica Puškar-Gallien. 2022. *Judge* vs. *fool*: Restrictive use of a noun predicts hybrid agreement. Talk presented at the Formal Approaches to Slavic Linguistics 31, 22.06.2022.

Baker, Mark C. 1985. The mirror principle and morphosyntactic explanation. *Linguistic Inquiry* 16(3). 373–415. https://www.jstor.org/stable/4178442.

Baker, Mark C. 2003. *Lexical categories: Verbs, nouns and adjectives*. Cambridge: Cambridge University Press.

Bárány, András. 2017. *Person, case, and agreement: The morphosyntax of inverse agreement and global case splits*. Oxford: Oxford University Press.

Béjar, Susana & Milan Řezáč. 2009. Cyclic Agree. *Linguistic Inquiry* 40(1). 35–73. https://www.jstor.org/stable/40071465.

Bernstein, Judy. 1993. *Topics in the syntax of nominal structure across Romance and Germanic languages*. New York: City University of New York. (Doctoral dissertation).

Bešlin, Maša. In press. DP in a model NP language: Evidence from Serbo-Croatian personal pronouns. *MIT Working Papers in Linguistics*.

Bittner, Maria & Kenneth Hale. 1996. The structural determination of case and agreement. *Linguistic Inquiry* 27(1). 1–68. https://www.jstor.org/stable/4178925.

Borer, Hagit. 2005. *Structuring sense, vol. 1: In name only*. Oxford: Oxford University Press.

Bošković, Željko. 2001. *On the nature of the syntax-phonology interface: Cliticization and related phenomena* (North Holland Linguistic Series: Linguistic Variations 60). Amsterdam: Elsevier.

Bošković, Željko. 2004. Clitic placement in South Slavic. *Journal of Slavic Linguistics* 12(1/2). 37–90. https://www.jstor.org/stable/24599868.

Bošković, Željko. 2008. What will you have, DP or NP? In Emily Elfner & Martin Walkow (eds.), *Proceedings of the North East Linguistic Society 37*, 101–114. Amherst: University of Massachusetts.

Brody, Michael. 2000. Mirror theory: Syntactic representation in perfect syntax. *Linguistic Inquiry* 31(1). 29–56. DOI: 10.1162/002438900554280.

Brody, Michael & Anna Szabolcsi. 2003. Overt scope in Hungarian. *Syntax* 6(1). 19–51. DOI: 10.1111/1467-9612.00055.

Browne, Wayles. 1974. On the problem of enclitic placement in Serbo-Croatian. In Richard D. Brecht & Catherine V. Chvany (eds.), *Slavic transformational syntax* (Michigan Slavic Materials 10), 36–52. Ann Arbor: University of Michigan.

Caha, Pavel. 2009. *The nanosyntax of case*. Tromsø: University of Tromsø. (Doctoral dissertation).

Caha, Pavel. 2021. Modeling declensions without declension features: The case of Russian. *Acta Linguistica Academica* 68(4). 385–425. DOI: 10.1556/2062.2021.00433.

Cardinaletti, Anna & Michal Starke. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In Henk van Riemsdijk (ed.), *Clitics in the languages of Europe*, 145–233. Berlin: Mouton de Gruyter.

Cooper, Robin. 1983. *Quantification and syntactic theory*. Dordrecht: Reidel.

Corbett, Greville G. 1979. The agreement hierarchy. *Journal of Linguistics* 15(2). 203–224. https://www.jstor.org/stable/4175494.

Déchaine, Rose-Marie & Martina Wiltschko. 2002. Decomposing pronouns. *Linguistic Inquiry* 33(3). 409–442. DOI: 10.1162/002438902760168554.

Despić, Miloje. 2011. *Syntax in the absence of Determiner Phrase*. Storrs: University of Connecticut. (Doctoral dissertation).

Despić, Miloje. 2017. Investigations in mixed agreement: Polite plurals, hybrid nouns and coordinate structures. *Morphology* 27(3). 253–310. DOI: 10.1007/s11525-017-9301-3.

Dvořák, Boštjan. 2007. Slovenian clitic pronouns and what is so special about them. *Slovenski jezik / Slovene Linguistic Studies* 6. 209–233. DOI: 10.17161/SLS.1808.4408.

Elbourne, Paul D. 2005. *Situations and individuals*. Cambridge: MIT Press.

Fassi Fehri, Abdelkader. 2000. Distributing features and affixes in Arabic subject verb agreement paradigms. In Jacqueline Lecarme, Jean Lowenstamm & Ur Shlonsky (eds.), *Research in Afroasiatic grammar*, 79–100. Amsterdam: John Benjamins.

Fassi Fehri, Abdelkader. 2018. *Constructing feminine to mean: Gender, Number, numeral and quantifier extensions in Arabic*. Lanham: Lexington Books.

Foley, Steven & Maziar Toosarvandani. 2022. Extending the person-case constraint to gender: Agreement, locality, and the syntax of pronouns. *Linguistic Inquiry* 53(1). 1–40. DOI: 10.1162/ling_a_00395.

Fox, Danny & David Pesetsky. 2005. Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31. 1–45. DOI: 10.1515/thli.2005.31.1-2.1.

Franks, Steven. 2013. Orphans, doubling, coordination, and phases: On nominal structure in Slovenian. *Slovenski jezik / Slovene Linguistic Studies* 9. 55–92.

Franks, Steven & Ljiljana Progovac. 1994. On the placement of Serbo-Croatian clitics. In George Fowler, Henry Cooper & Jonathan Ludwig (eds.), *Indiana Slavic studies 7, Proceedings of the 9th biennial conference on Balkan and South Slavic linguistics, literature and folklore*, 69–78. Bloomington: Indiana University.

Godjevac, Svetlana. 2000. *Intonation, word order and focus projection in Serbo-Croatian*. Columbus: The Ohio State University. (Doctoral dissertation).

Harbour, Daniel. 2007. Against PersonP. *Syntax* 10. 223–243. DOI: 10.1111/j.1467-9612.2007.00107.x.

Harbour, Daniel. 2008a. Discontinuous agreement and the syntax-morphology interface. In Daniel Harbour, David Adger & Susana Béjar (eds.), *Phi-theory: Phi-features across modules and interfaces*, 185–220. Oxford: Oxford University Press.

Harbour, Daniel. 2008b. *Morphosemantic number: From Kiowa noun classes to UG*. Dordrecht: Springer.

Harbour, Daniel. 2016. *Impossible persons*. Cambrigde: MIT Press.

Harley, Heidi & Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78(3). 482–526.

Heim, Irene. 2008. Features on bound pronouns. In Daniel Harbour, David Adger & Susana Bejar (eds.), *Phi theory: Phi-features across modules and interfaces*, 35–56. Oxford: Oxford University Press.

Höhn, Georg F.K. 2016. Unagreement is an illusion: Apparent person mismatches and nominal structure. *Natural Language & Linguistic Theory* 34(2). 543–592. DOI: 10.1007/s11049-015-9311-y.

Jacobson, Pauline. 2012. The direct compositionality and "uninterpretability": The case of (sometimes) "uninterpretable" features on pronouns. *Journal of Semantics* 29(3). 305–343. DOI: 10.1093/jos/ffs005.

Jovović, Ivana. 2024. Condition B and other conditions on pronominal licensing in Serbo-Croatian. *Linguistic Inquiry* 55(2). 402–421. DOI: 10.1162/ling_a_00475.

Kramer, Ruth. 2015. *The morphosyntax of gender*. Oxford: Oxford University Press.

Kratzer, Angelika. 2007. On the plurality of verbs. In Tatjana Heyde-Zybatow & Johannes Dölling (ed.), *Event structures in linguistic form and interpretation*, 269–299. Berlin: Mouton de Gruyter.

Kratzer, Angelika. 2009. Making a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry* 40(2). 187–237.

Kučerová, Ivona. 2018. Phi-features at the syntax-semantics interface: Evidence from nominal inflection. *Linguistic Inquiry* 49(4). 813–845. DOI: 10.1162/ling_a_00290.

Landau, Idan. 2006. Chain resolution in Hebrew V(P) fronting. *Syntax* 9. 32–66. DOI: 10.1111/j.1467-9612.2006.00084.x.

Landau, Idan. 2016. DP-internal semantic agreement: A configurational analysis. *Natural Language & Linguistic Theory* 34(3). 975–1020. DOI: 10.1007/s11049-015-9319-3.

Lochbihler, Bethany, Will Oxford & Nicholas Welch. 2021. The person-animacy connection: Evidence from Algonquian and Dene. *Canadian Journal of Linguistics / Revue canadienne de linguistique* 66(3). 432–442. DOI: 10.1017/cnj.2021.14.

Longobardi, Giuseppe. 1994. Reference and proper names: A theory of N-movement in syntax and Logical Form. *Linguistic Inquiry* 25. 609–665.

McFadden, Thomas. 2018. *ABA in stem-allomorphy and the emptiness of the nominative. *Glossa: a journal of general linguistics* 3(1). 8. DOI: 10.5334/gjgl.373.

McFadden, Thomas & Sandhya Sundaresan. 2009. Subject distribution and finiteness in Tamil and other languages: Selection vs. case. *Journal of South Asian Linguistics* 2. 5–34.

Merchant, Jason. 2014. Gender mismatches under nominal ellipsis. *Lingua* 151. 9–32. DOI: 10.1016/j.lingua.2014.01.008.

Milićev, Tanja & Maša Bešlin. 2019. Instrumental case: Why it is absent from the clitic system in Serbian/Croatian. In Vesna Lopičić & Biljana Mišić Ilić (eds.), *Jezik, književnost, teorija*, 153–168. Niš: Faculty of Philosophy.

Moskal, Beata. 2015a. *Domains on the border: Between morphology and phonology*. Storrs: University of Connecticut. (Doctoral dissertation).

Moskal, Beata. 2015b. Limits on allomorphy: A case study in nominal suppletion. *Linguistic Inquiry* 46(2). 363–375. DOI: 10.1162/LING_a_00185.

Murphy, Andrew, Zorica Puškar & Matías Guzmán Naranjo. 2018. The structure of hybrid nouns in Bosnian/Croatian/Serbian: Experimental evidence from ellipsis. In Denisa Lenertová, Roland Meyer, Radek Šimík & Luka Szucsich (eds.), *Advances in formal Slavic linguistics 2016*, 357–379. Berlin: Language Science Press. DOI: 10.5281/zenodo.2545533.

Neeleman, Ad & Kriszta Szendrői. 2007. Radical pro drop and the morphology of pronouns. *Linguistic Inquiry* 38. 671–714.

Nevins, Andrew. 2011. Multiple Agree with clitics: Person complementarity vs. omnivorous number. *Natural Language & Linguistic Theory* 29(4). 939–971. DOI: 10.1007/s11049-011-9150-4.

Noyer, Rolf. 1992. *Features, positions, and affixes in autonomous morphological structure*. Cambridge: MIT. (Doctoral dissertation).

Pereltsvaig, Asya. 2007. *Copular sentences in Russian: A theory of intra-clausal relations*. Dordrecht: Springer.

Pesetsky, David. 2013. *Russian case morphology and the syntactic categories*. Cambridge: MIT Press.

Picallo, M. Carme. 1991. Nominals and nominalization in Catalan. *Probus* 3(3). 279–316.

Postal, Paul. 1969. On so-called "pronouns" in English. In David Reibel & Sanford Schane (eds.), *Modern readings in transformational grammar*, 201–224. Englewood Cliffs: Prentice-Hall.

Progovac, Ljiljana. 1998. Determiner phrase in a language without determiners. *Journal of Linguistics* 34. 165–179.

Puškar, Zorica. 2018. Interactions of gender and number agreement: Evidence from Bosnian/Croatian/Serbian. *Syntax* 21(3). 275–318. DOI: 10.1111/synt.12154.

Puškar-Gallien, Zorica. 2019. Resolving polite conflicts in predicate agreement. *Glossa: a journal of general linguistics* 4(1). 33. DOI: 10.5334/gjgl.587.

Puškar-Gallien, Zorica. Forthcoming. Disassembling and reassembling pronouns: A case study of Bosnian/Croatian/Montenegrin/Serbian. *Journal of Slavic Linguistics*.

Ruda, Marta. 2021a. Polish personal pronouns: [PersP Pers [NumP Num [n]]] and [NumP Num [n]]. *Studies in Polish Linguistics* 16(1). 23–40. DOI: 10 . 4467 / 23005920SPL.21.002.13956.

Ruda, Marta. 2021b. Strict and sloppy readings of pronominal objects in Polish. *Studies in Polish Linguistics* 16(2). 121–144. DOI: 10.4467/23005920SPL.21.006. 13960.

Runić, Jelena. 2014. *A new look at clitics, clitic doubling, and argument ellipsis: Evidence from Slavic*. Storrs: University of Connecticut. (Doctoral dissertation).

Sauerland, Uli. 2013. Presuppositions and the alternative tier. *Proceedings of SALT* 23. 156–173. DOI: 10.3765/salt.v23i0.2673.

Sichel, Ivy & Maziar Toosarvandani. 2024a. Nominal intervention and the extended PERSON domain. In Robert Autry, Gabriela de la Cruz, Luis A. Irizarry Figueroa, Kristina Mihajlovic, Tianyi Ni, Ryan Smith & Heidi Harley (eds.), *Proceedings of the 39th West Coast Conference on Formal Linguistics*, 662–669. Somerville: Cascadilla Proceedings Project.

Sichel, Ivy & Maziar Toosarvandani. 2024b. The featural life of nominals. *Linguistic Inquiry*. DOI: 10.1162/ling_a_00517.

Siewierska, Anna. 2013. Gender distinctions in independent personal pronouns. In Matthew S. Dryer & Martin Haspelmath (eds.), *WALS online (v2020.3) [Data set]*. Zenodo. DOI: 10.5281/zenodo.7385533.

Smith, Peter W., Beata Moskal, Ting Xu, Jungmin Kang & Jonathan David Bobaljik. 2019. Case and number suppletion in pronouns. *Natural Language & Linguistic Theory* 37. 1029–1101. DOI: 10.1007/s11049-018-9425-0.

Stanković, Branimir. 2014a. Arguments for DP-analysis of Serbo-Croatian nominal expressions. In Ludmila Veselovská & Markéta Janebová (eds.), *Nominal structures: All in complex DPs*, 29–48. Olomouc: Palacký University.

Stanković, Branimir. 2014b. *Sintaksa i semantika odredjenog i neodredjenog pridevskog vida u srpskom jeziku.* Kragujevac: University of Kragujevac. (Doctoral dissertation).

Stegovec, Adrian. 2019. Crop to fit: Pronoun size and its relation to strict/sloppy identity. Handout from a talk presented at the 93rd Annual Meeting of the Linguistic Society of America (LSA), New York, NY, January 3–6. https://drive.google.com/file/d/1H4ip3kRCqr9cVwblyi1nuMfdEYDU4cxR/view.

Steriopolo, Olga. 2018a. Mixed gender agreement in the case of Russian hybrid nouns. *Questions and Answers in Linguistics* 5(2). 91–105. DOI: 10.2478/qal-2018-0001.

Steriopolo, Olga. 2018b. Morphosyntax of gender in Russian sex-differentiable nouns. *Journal of Slavic Linguistics* 26(1). https://www.jstor.org/stable/26742398.

Steriopolo, Olga & Martina Wiltschko. 2010. Distributed gender hypothesis. In Gerhild Zybatow, Philip Dudchuk, Serge Minor & Ekaterina Pschehotskaya (eds.), *Formal studies in Slavic linguistics: Proceedings of FDSL 7.5*, 155–172. New York: Peter Lang.

Stojanović, Danijela. 1997. Object shift in Serbo-Croatian. In James R. Black & Virginia Motapanyane (eds.), *Clitics, pronouns and movement*, 301–320. Amsterdam: John Benjamins.

Sudo, Yasutada. 2012. *On the semantics of phi features on pronouns.* Cambridge: MIT. (Doctoral dissertation).

Sudo, Yasutada & Giorgos Spathas. 2020. Gender and interpretation in Greek: Comments on Merchant (2014). *Glossa: a journal of general linguistics* 5(1). 129. DOI: 10.5334/gjgl.1173.

Talić, Aida. 2018. Spelling out enclitics and giving their tone a voice: Cyclic clitic incorporation in BCS and breaking the cycle. *The Linguistic Review* 35(2). 307–370. DOI: 10.1515/tlr-2017-0026.

Trenkić, Danijela. 2004. Definiteness in Serbian/Croatian/Bosnian and some implications for the general structure of the nominal phrase. *Lingua* 114. 1401–1427. DOI: 10.1016/j.lingua.2003.09.005.

Trommer, Jochen. 2002. The interaction of morphology and syntax in affix order. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2002*, 283–324. Dordrecht: Kluwer.

van Alem, Astrid. 2025. Complementizer agreement is clitic doubling: Evidence from intervention effects in Frisian and Limburgian. *Natural Language & Linguistic Theory* 43. 115–149. DOI: https://doi.org/10.1007/s11049-024-09621-9.

van Koppen, Marjo. 2012. The distribution of phi-features in pronouns. *Natural Language & Linguistic Theory* 30(1). 135–177. DOI: 10.1007/s11049-011-9159-8.

van Urk, Coppe. 2018. Pronoun copying in Dinka and the copy theory of movement. *Natural Language & Linguistic Theory* 36(3). 937–990. DOI: 10.1007/s11049-017-9384-x.

Zec, Draga & Sharon Inkelas. 1991. The place of clitics in the prosodic hierarchy. In Dawn Bates (ed.), *Proceedings of the 10th West Coast Conference on Formal Linguistics*, 505–519. Stanford: CSLI Publications.

Chapter 18

Animacy influences segmental phonology: The velar–sibilant alternation in BCMS

Marko Simonović

University of Graz

Bosnian/Croatian/Montenegrin/Serbian velar–sibilant alternation is a morphologised process that varies in application rates depending on the context. This article focuses on assibilation in DAT/LOC.SG of the nouns which end in *-a* in the citation form, where varying assibilation ratios are encountered. Two corpus studies targeting nouns with velar-final stems were conducted to establish the influence of phonological factors, animacy, and the presence of a non-alternating /i/ elsewhere in the paradigm on the alternation ratios. The results show that animacy comes out as a significant predictor of the alternation ratios in DAT/LOC.SG in both data sets.

1 Introduction

Bosnian/Croatian/Montenegrin/Serbian (BCMS) assibilation, whereby velars /k, g, x/ alternate with sibilants /ts, z, s/ in front of an /i/-initial affix, is a highly morphologised process whose application rates vary from context to context.[1] Table 1 shows four morphemes which all have the segmental content /i/. The imperative morpheme unexceptionally triggers the alternation, the NOM/VOC.PL morpheme triggers the alternation productively, but exceptions are attested. On

---

[1] To my knowledge, there are no major differences between the four varieties when it comes to assibilation. As clarified in Section 3, the empirical basis for this study originates from a Croatian and a Serbian corpus.

the other hand, the DAT/LOC.SG morpheme is one of the examples where it is hard to determine whether application or non-application is more common: the morpheme triggers the alternation in some words, fails to do so in others, and triggers it optionally in yet other words. Finally, the GEN.PL morpheme never triggers assibilation.

Table 1: Four assibilation contexts in BCMS

Application ratio	Morphological context	Examples
Categorical	IMP	/leg-i/ → [lezi] 'lie down'
High	NOM/VOC.PL (nouns)	/kirurg-i/ → [kirurzi] 'surgeons' (except in very few cases such as /detʃk-i/ → [detʃki] 'guys')
Medium	DAT/LOC.SG (nouns)	/bajk-i/ → [bajtʃi] 'fairy tale (DAT/LOC)' /alg-i/ → [algi] 'alga (DAT/LOC)' /fresk-i/ → [freski] / [frestʃi] 'fresco (DAT/LOC)'
Zero	GEN.PL (nouns)	/bajk-i/ → [bajki] 'fairy tales (GEN)' /alg-i/ → [algi] 'algae (GEN)'

The primary focus of this contribution is on the DAT/LOC.SG ending /i/. I adhere to the assumption of traditional approaches to BCMS, asserting that the underlying form of the exponent of the DAT/LOC.SG morpheme remains consistent. Various factors, then, determine whether this exponent triggers assibilation. In other words, I do not adopt an overabundance analysis (see Thornton 2019 for the general approach and Lečić 2015 for an analysis in terms of overabundance in Croatian). An overabundance analysis would posit two DAT/LOC.SG endings differing solely in their assibilation behavior. The rationale behind this choice not to employ an overabundance analysis lies in the observation that variable assibilation is characteristic of a plethora of unrelated morphological contexts in BCMS. Assuming two different endings competing in all these contexts would face the problem of accounting for the fact that precisely these two endings compete in all these unrelated contexts.

The primary objective of the paper is to establish the factors determining the occurrence of assibilation in DAT/LOC.SG. The factors, as outlined in standard descriptions, will serve as predictors in a statistical model to anticipate assibilation

encountered in the corpus (see Lečić 2016 for an overview of statistical modeling of variation observed in corpus data). In addition to phonological and morphological factors, special attention will be given to the influence of animacy on the velar–sibilant alternation in BCMS. This attention is warranted due to its theoretical relevance – the impact of a semantic factor on a segmental alternation.

The rest of the paper is organised as follows. Section 2 provides an overview of the predictors of assibilation and their treatment in the descriptive literature on BCMS. In Section 3, I describe the two corpus studies and the rationale behind them. Section 4 presents the results of the corpus studies, with the main finding being that animacy is a strong predictor of assibilation. Section 5 discusses the results and their theoretical implications. Finally, Section 6 concludes the paper.

2 Predictors of assibilation

Virtually all traditional descriptions of BCMS have a dedicated section on the application of assibilation in each of the morphological contexts. There are three important methodological obstacles in using these descriptions when modelling modern BCMS. Firstly, they often mix prescriptive objectives with descriptive ones, failing to distinguish between the two domains. Secondly, even when fully descriptive, they frequently lack a description of the empirical basis for the descriptions. Finally, they contain long lists of classes or intersections of classes in which assibilation is either favoured or blocked, without any indication of the strength of the generalisations or the size of classes in question. In what follows, I will provide a brief overview of the main phonological, morphological, and semantic factors that influence assibilation in BCMS. I will focus on generalisations that are applicable to a significant number of cases and can be meaningfully tested in a quantitative analysis. The ultimate goal is to define a list of properties which can be used in the quantitative analysis based on corpus data.

2.1 Phonological factors

The most general phonological factor, described in quite some detail in Težak (1986), is the FINAL VELAR, i.e. the difference between the three velars. /k/ is most prone to assibilation, /x/ assibilates in a minority of cases, whereas /g/ takes an intermediate position. This mirrors the size of the relevant classes within the declension class: *k*-final stems are much more common than *g*-final stems, which are in turn much more common than *x*-final stems (e.g., Petrović & Gudurić 2010: 475–478).

Several descriptions account for the non-application of assibilation in some classes by the fact that "the alternation would be experienced as moving away from the citation form of the word" (translation mine) (Pešikan et al. 2010: 47). While this criterion is extremely vague and therefore difficult to implement, Barić et al. (1997: 154) argue that the danger of excessively altering the stem is especially relevant for words with monosyllabic stems (simply because they have less stem material). Following this reasoning, we can implement the factor MONOSYLLABIC STEM as one of the predictors of non-assibilation.

All other phonological factors refer to stem-final consonant clusters, many of which block assibilation. While most such generalisations are tendencies, there is an unexceptional generalisation (first described as "self-evident" in Maretić 1963: 169): Assibilation never applies if the result would lead to total identity with the preceding consonant. This means that stems ending in *-tsk-, -zg-* and *-sx-* never alternate. Unfortunately, this generalisation only applies to a handful of items.

The long lists of rules concerning more frequent clusters (all of which are *k*-final) can be summarised as follows: all blockers have an obstruent stop or an affricate as the first member of the cluster, clusters which have a fricative as the first member allow both assibilation and non-assibilation, whereas clusters with more sonorous consonants tend to favour assibilation. We can conclude that the sonority of the consonant preceding the stem-final velar (C_1-SONORITY) is a predictor of assibilation. The more sonorous the first consonant of the cluster is, the more it is probable that assibilation will occur.

2.2 Morphological factors

The most important morphological predictor of assibilation is the specific morphological context. Since the present study only focuses on one specific context, this factor is controlled for. Still, I will take a brief look at two other morphemes which trigger assibilation in the nominal paradigm: NOM/VOC.PL *-i* and DAT/LOC/INS.PL *-ima*. The discussion of these two morphemes will be helpful in formulating a hypothesis concerning the morpheme in focus here.

As mentioned in Section 1, the NOM/VOC.PL morpheme *-i* triggers assibilation, with very few exceptions. This morpheme only shows up in the paradigms of masculine nouns, illustrated by the paradigm of [kirurg] 'surgeon' in Table 2. Such paradigms always contain another form with an assibilation-triggering ending: the DAT/LOC/INS.PL ending *-ima*. Interestingly, the DAT/LOC/INS.PL *-ima* also shows up in the paradigm of neuter nouns, where it is the only assibilation-triggering ending, as illustrated by the paradigm of [blago] 'treasure' in Table 2.

The actual acceptability of the DAT/LOC/INS.PL forms of the handful of neuter nouns with velar-final stems still needs to be established. The parallel forms with and without assibilation cited in the table are based on the description in Marković (2018: 136), who also points out that the forms with assibilation are in this case somewhat more marked. My personal judgement is ineffability in these case forms for all three nouns mentioned by Marković (2018). In other words, for all three nouns ([klupko] 'ball (of yarn)', [blago] 'treasure' and [ruxo] 'attire'), I cannot derive an acceptable form with the ending *-ima.* Either way, it is clear that DAT/LOC/INS.PL *-ima* triggers assibilation much more successfully in masculine paradigms than it does in neuter ones.

Table 2: Assibilation in masculine and neuter paradigms as illustrated by [kirurg] 'surgeon' and [blago] 'treasure' based on Marković (2018)

	Masculine		Neuter	
	SG	PL	SG	PL
NOM	kirurg	**kirurz-i**	blag-o	blag-a
GEN	kirurg-a	kirurg-a	blag-a	blag-a
DAT/LOC	kirurg-u	**kirurz-ima**	blag-u	**blag-ima/blaz-ima**
ACC	kirurg-a	kirurg-e	blag-o	blag-a
VOC	kirurg-u	**kirurz-i**	blag-o	blag-a
INS	kirurg-om	kirurz-ima	blag-om	**blag-ima/blaz-ima**

One possible way of understanding the empirical picture described above is that the tendency towards assibilation is stronger in cases where multiple assibilation-triggering endings occur in a paradigm. This could be due to the cumulative effect of these endings, which may enable the licensing of allomorphy. If this reasoning is correct, we would expect assibilation to be even more limited in paradigms containing another *-i* ending that does not trigger assibilation. The feminine nouns in focus here are actually the ideal testing ground for this hypothesis, because some of them have the GEN.PL ending *-i*, which never triggers assibilation, as mentioned in Section 1.

The distribution of the GEN.PL endings in the feminine declension in focus here can be summarised as follows. Nouns with a single stem-final consonant have the ending *-a* (e.g., [sʊraka] 'magpie.GEN.PL').[2] On the other hand, in nouns

---

[2]In all traditional descriptions, this GEN.PL ending contains a long vowel and causes a lengthening of the preceding vowel, e.g. [sʊraka] 'magpie.NOM.SG' vs. [sʊraːkaː] 'magpie.GEN.PL'. I am

with a stem-final consonant cluster three endings are attested: *-i* (e.g. in [kriŋki] 'disguise.ɢᴇɴ.ᴘʟ'), *-aa*, whose first vowel breaks up the consonant cluster (e.g. in [banaka] 'bank.ɢᴇɴ.ᴘʟ'), and, somewhat marginally, *-a* (e.g., in [superscript: ?][kriŋka] 'disguise.ɢᴇɴ.ᴘʟ' and [superscript: ?][baŋka] 'bank.ɢᴇɴ.ᴘʟ'). If endings can play a role in favouring/blocking allomorphy in other members of the paradigm, nouns which have ɢᴇɴ.ᴘʟ in *-i* should be less prone to assibilation in ᴅᴀᴛ/ʟᴏᴄ.ꜱɢ than nouns which have other endings in ɢᴇɴ.ᴘʟ. In other words, if -ɪ[ɢᴇɴ.ᴘʟ] is a strong factor that blocks assibilation, most nouns should behave either as [kriŋka] 'disguise' (i.e., have a ɢᴇɴ.ᴘʟ in *-i* and no assibilation in ᴅᴀᴛ/ʟᴏᴄ.ꜱɢ) or as [baŋka] 'bank.ɢᴇɴ.ᴘʟ' (i.e., not have a ɢᴇɴ.ᴘʟ in *-i* and exhibit assibilation in ᴅᴀᴛ/ʟᴏᴄ.ꜱɢ).

Table 3: Assibilation in [baŋka] and lack of assibilation in [kriŋka]

	SG	PL	SG	PL
ɴᴏᴍ	baŋk-a	baŋk-e	kriŋk-a	kriŋk-i
ɢᴇɴ	baŋk-e	banak-a	kriŋk-e	kriŋk-i
ᴅᴀᴛ/ʟᴏᴄ	**ban**ts-**i**	baŋk-ama	**kriŋk-i**	kriŋk-ama
ᴀᴄᴄ	baŋk-u	baŋk-e	kriŋk-u	kriŋk-e
ᴠᴏᴄ	baŋk-o	baŋk-e	kriŋk-o	kriŋk-e
ɪɴꜱ	baŋk-om	baŋk-ama	kriŋk-om	kriŋk-ama

2.3 Semantic factors

All traditional descriptions contain lists of classes in which assibilation is blocked, often defined by one semantic and one formal criterion. Almost all such classes are restricted to animates. For instance, Težak & Babić (1992: 92–93) include classes such as: personal male and female names, surnames regardless of their origin, names of pets and domestic animals, terms of endearment, ethnonyms derived using the suffixes *-ka*, *-nka* and *-čanka*, nouns in *-jka* which mean a female person with a certain characteristic or are derived from loanwords, and many others. While none of the descriptive works that I am aware of suggest that there is a direct link between animacy and lack of assibilation, overviews as the one sketched above justify the implementation of ᴀɴɪᴍᴀᴄʏ as one of the factors that blocks assibilation.

---

ignoring both vowel length and suprasegmental information here, because there is considerable variation in this respect, including large numbers of speakers who do not have distinctive vowel length.

Indirect evidence that animacy plays a role in blocking assibilation is contained in the discussions of minimal pairs which emerge when words that typically refer to animates (often female inhabitants) acquire an additional inanimate referent, e.g., a factory, a restaurant, etc. In such cases, all sources report that assibilation becomes possible with the inanimate referent. Examples are *Podravka* (the inhabitant of the Podravina region or a factory in Koprivnica, Croatia), *Beograđanka* (the female inhabitant of Belgrade or a building in Belgrade), and *Japanka* (a Japanese woman) vs. *japanka* (flip flop). Most normative sources argue against assibilation in such cases (e.g., Hudeček 2022 and Pešikan et al. 2010: 47), while others just describe it (e.g., Barić et al. 1997: 154). A potential example of an extension in the other direction would be the word *stranka*, which most commonly means 'party' (e.g., political party) but in some contexts can mean 'client'. My intuition is that assibilation is only possible in the former meaning.

3 Methodology

Given the shortcomings of the existing descriptions, in order to get a realistic picture of the data, I obtained data from two web corpora of BCMS: hrWaC and srWaC (Ljubešić & Klubička 2014). Corpus data are especially valuable when studying phenomena that exhibit a significant amount of variation, because they allow for the calculation of the relative frequencies of the specific options. In the case of assibilation in DAT/LOC.SG, which allows for a considerable amount of variation, the relevant construct to be employed here is the ASSIBILATION RATIO. The assibilation ratio of a word is the proportion of the DAT/LOC.SG forms with assibilation, calculated as the number of DAT/LOC.SG forms with assibilation divided by the total number of DAT/LOC.SG forms. For instance, if three DAT/LOC.SG tokens of the word [loziŋka] 'password' were extracted and two of them are [lozintsi], whereas one is [loziŋki], then the ASSIBILATION RATIO for this noun is 0.67.

The outcome variable, ASSIBILATION RATIO, can be computed for any noun. The same is true for ANIMACY, FINAL VELAR and MONOSYLLABIC STEM. However, some of the predictors can only be meaningfully applied to a subset of nouns. Specifically, C_1-SONORITY and -I[GEN.PL] (implemented here as -I[GEN.PL] RATIO) can only be applied to nouns with stems that end in a consonant cluster. I therefore address the nouns with stem-final consonant clusters in a separate study.

Marko Simonović

3.1 Study 1: Assibilation with CC-final stems in hrWaC

This study was based on data from the Croatian web corpus hrWaC. In order to obtain nouns with CC-final stems, I first conducted a CQL search for lemmas ending in -CGa, where C is any consonant and G is any velar. The results were ranked by frequency and the 204 most frequent nouns were copied to a separate table.[3] Each noun was then annotated for FINAL VELAR, MONOSYLLABIC STEM, C_1-SONORITY and ANIMACY.

The annotation for ANIMACY was implemented analogously to the category of animacy in the masculine declension, where animacy influences the exponence of the ACC.SG (ACC.SG = GEN.SG for animates, ACC.SG = NOM.SG for inanimates). For instance, the noun [lutka] 'puppet' was annotated as animate because its masculine counterpart [lutak] 'male puppet' declines as animate.

For simplicity, the predictor C_1-SONORITY was implemented as a binary variable. A value of 0 was assigned to cases where the first consonant of the stem-final cluster is an obstruent stop or an affricate, while a value of 1 was assigned to all other cases.

An initial overview of the data showed that FINAL VELAR could not be meaningfully included as a factor, because, among 204 most frequent nouns, there were no *x*-final stems and only eight *g*-final stems. I therefore decided to only include *k*-final items in this study and replaced the eight *g*-final items with the next eight *k*-final items from the frequency list.

For each of the targeted nouns the counts of all the possible DAT/LOC.SG and GEN.PL forms were obtained by processing the results of CQL queries. Based on these counts the values for ASSIBILATION RATIO and -I[GEN.PL] RATIO were calculated. Specifically, since the morphological tags were found to be unreliable, CQLs were used to find strings in which the word in question is preceded by two congruent adjectival words. This method proved to yield a sufficiently precise sample, which could be manually cleaned within the constraints of the available time and manpower. The CQL used for the DAT/LOC.SG form of the word [freska] 'fresco' is shown in (1a), while (1b) shows the CQL used for the GEN.PL forms of the same noun.[4]

(1) a. `[word = ".*oj"][word = ".*oj"][word = "fres(c|k)i"]`

---

[3] The numbers of items eventually included in the study depended on the available time and manpower. However, it should be pointed out that the sample did include low frequency nouns, whose meaning needed to be looked up.

[4] The employed endings from the adjectival declension uniquely identify the relevant paradigm cells: *-oj* only appears in DAT/LOC.SG, whereas *-ih* [ix] only appears in GEN.PL.

```
b. [word = ".*ih"][word = ".*ih"][word = "fres(ki|aka|ka)"]
```

The nouns for which one of the queries yielded an empty result were removed and supplanted by the following word from the frequency ranking.

For the statistical analysis, the data were transformed so that each attestation of the DAT/LOC.SG form in the corpus constituted a separate observation (row in the table). This allowed us to treat the outcome variable ASSIBILATION as a binary variable. All relevant data, including ASSIBILATION, MONOSYLLABIC STEM, C_1-SONORITY, ANIMACY, and -I[GEN.PL] RATIO, are published in Simonović (2024). These data were inputted into a mixed-effects logistic regression model in R, where ASSIBILATION served as the outcome variable, and MONOSYLLABIC STEM, C_1-SONORITY, ANIMACY, and -I[GEN.PL] RATIO were treated as fixed effects. Additionally, the specific noun was included as a random factor to account for random variance between different nouns.

3.2 Study 2: Assibilation with VC-final stems in srWaC

This study was based on data from the Serbian web corpus srWaC.[5] In order to obtain nouns with VG-final stems, we first conducted a CQL search for lemmas ending in -VGa, where V is any vowel and G is any velar. The results were cleaned and ranked by frequency. The 349 most frequent nouns were copied to a separate table and annotated for FINAL VELAR, MONOSYLLABIC STEM and ANIMACY.[6]

The annotation for ANIMACY was implemented as in Study 1. Since the morphological tags were found to be unreliable, the values for ASSIBILATION RATIO were obtained by processing results of two CQL queries. Specifically, CQLs were used to find strings in which the target word is preceded by one of the typical prepositions (2a illustrates this for [baraka] 'barrack') and strings in which the word in question is preceded by a congruent adjectival word (2b).

(2) a. [lemma = "(o|u|na|prema|k|ka)"][word = "bara(c|k)i"]
     b. [word = ".*oj"][word = "bara(k|c)i"]

The search results were manually cleaned and the ASSIBILATION RATIO was calculated for each noun. The nouns for which both queries yielded an empty result

---

[5]Data collection for this study was conducted in collaboration with participants of the course Collecting and Analyzing Corpus and Experimental Data in Hypothesis-Driven Linguistic Research at the University of Novi Sad.

[6]As with the previous study, the numbers of items eventually included in the study depended on the available time and manpower. However, it should be pointed out that the sample did include low frequency nouns, whose meaning needed to be looked up.

were removed and supplanted by the following word from the frequency ranking.

As in Study 1, the data were transformed so that each attestation of the DAT/LOC.SG form constituted a separate observation (row in the table). This allowed us to treat the outcome variable ASSIBILATION as a binary variable. All relevant data, including values for FINAL VELAR, MONOSYLLABIC STEM, ANIMACY and ASSIBILATION are published in Simonović (2024). These data were inputted into a mixed-effects logistic regression model in R, where ASSIBILATION served as the outcome variable, and FINAL VELAR, MONOSYLLABIC STEM, ANIMACY were treated as fixed effects. Additionally, the specific noun was included as a random factor to account for random variance between different nouns.

4 Results

4.1 Study 1

Before presenting the results of the statistical model, a brief overview of the mean values for the ASSIBILATION RATIO is provided. In this study, the overall mean ASSIBILATION RATIO is 0.33. The means for all groups identified by single values of the binary variables, along with the number of items in these groups, are presented in Table 4.

The mean ASSIBILATION RATIO for animate nouns exhibits a notably low value, also indicating a significant difference of means concerning ANIMACY. Similarly, and as expected, a considerable difference in means is observed for C_1-SONORITY. Specifically, stems in which the first consonant of the stem-final cluster is an obstruent stop or an affricate, display, on average, a lower ASSIBILATION RATIO compared to stems with different consonant configurations.

Interestingly, the difference in means for MONOSYLLABIC STEM is relatively small, but it also deviates from the expected pattern: monosyllabic stems exhibit a higher mean ASSIBILATION RATIO than polysyllabic ones.

The binary predictor variables mentioned earlier, along with the continuous predictor variable -I[GEN.PL]RATIO (with a mean of 0.87 in the dataset), were incorporated as fixed factors in a generalised linear mixed model. The binary variables MONOSYLLABIC STEM, C_1-SONORITY and ANIMACY were stored as factors, while -I[GEN.PL]RATIO was the only numeric factor. Individual lemmas, were included as a random factor, with by-noun varying intercepts. In (3) I provide the formula for the model as implemented in R using the package lme4 (Bates et al.

Table 4: Mean ᴀssɪʙɪʟᴀᴛɪᴏɴ ʀᴀᴛɪᴏ for each value of the binary variables

Variable	Mean AR for 1 (ɴ)	Mean AR for 0 (ɴ)	Difference
ᴀɴɪᴍᴀᴄʏ	0.04 (77)	0.51 (127)	−0.47
C_1-sᴏɴᴏʀɪᴛʏ	0.39 (160)	0.13 (44)	0.26
ᴍᴏɴᴏsʏʟʟᴀʙɪᴄ sᴛᴇᴍ	0.40 (80)	0.29 (124)	0.11

2015). The complete script is published in Simonović (2024). The summarised results can be found in Table 5.[7]

(3)
```
model1 <- glmer(assib ~ mono + anim + c1son + igenpl + (1 |
    noun), family = binomial(link = "logit"), data = analysis1)
```

Table 5: Generalised linear mixed model results

Variable	Coefficient	Std. Error	z value	Pr($> |z|$)	Odds Ratio
(Intercept)	3.3012	1.0109	3.266	0.00109**	27.1447
ᴍᴏɴᴏsʏʟʟᴀʙɪᴄ sᴛᴇᴍ	0.1950	0.6303	0.309	0.75703	1.2153
ᴀɴɪᴍᴀᴄʏ	−6.7532	0.8073	−8.366	$< 2 \times 10^{-16}$***	0.0012
C_1-sᴏɴᴏʀɪᴛʏ	−4.9472	0.7643	−6.473	9.62×10^{-11}***	0.0071
-ɪ[ɢᴇɴ.ᴘʟ] ʀᴀᴛɪᴏ	−2.3075	0.9823	−2.349	0.01882*	0.0995

The *Coefficient* column in Table 5 provides the estimated coefficients, revealing the log-odds change in the outcome variable for a one-unit change in each predictor. These estimates offer valuable insights into the direction and magnitude of the predictors' impact. Accompanying the estimates, the *Std. Error* column indicates the standard error of each coefficient estimate. This information is crucial for assessing the precision and reliability of the estimated coefficients. The z-statistic is calculated as the coefficient estimate divided by the standard error, where larger values indicate a larger estimated effect size. The *p*-value represents the probability of observing an effect at least as large as the one found assuming the null hypothesis is true. Using an *alpha*-level of .05, we consider effects where

---

[7]The significance codes used in this report follow the standard R output format: '***' for *p*-values $\leq$ 0.001, '**' for *p*-values $\leq$ 0.01, '*' for *p*-values $\leq$ 0.05, '.' for *p*-values $\leq$ 0.1, and no extra symbol for *p*-values > 0.1. These codes are retained from the R output for consistency in reporting results.

$p < .05$ to be significant. Lastly, the *Odds Ratio* column shows the exponentiated coefficients, offering a clear understanding of the multiplicative change in odds for a one-unit change in each predictor. This column provides practical insights into the implications of the predictors on the odds of the outcome.

Summarising the findings, Table 5 indicates, that among the four predictors, only two exhibit a highly statistically significant relationship with ASSIBILATION: ANIMACY and C_1-SONORITY both demonstrate particularly strong negative associations, as evidenced by their low *p*-values. The predictor -I[GEN.PL] RATIO shows a significant negative association with ASSIBILATION, aligning notably in the expected direction. It should, however, be noted that the magnitude of the effect size (OR) is somewhat lower than all the other significant predictors. Finally, the predictor MONOSYLLABIC STEM does not show a statistically significant relationship with the outcome variable. Consequently, the unexpected positive difference of means observed earlier can be attributed to chance rather than a meaningful association.

4.2 Study 2

As with Study 1, I begin with a brief overview of the mean values for the ASSIBILATION RATIO. In this study, the overall mean ASSIBILATION RATIO is 0.75, which is much higher than in the previous study. The means for all groups identified by single values of the binary variables, along with the number of items in these groups, are presented in Table 6.

The mean ASSIBILATION RATIO for animate nouns exhibits a notably low value, indicating a significant disparity in means concerning ANIMACY. Interestingly, the difference in means for MONOSYLLABIC STEM is also relatively high, and it goes in the expected direction: monosyllabic stems exhibit a lower mean ASSIBILATION RATIO than polysyllabic ones.

Table 6: Mean ASSIBILATION RATIO for both binary variables

Variable	Mean SR for 1 (N)	Mean SR for 0 (N)	Difference
ANIMACY	0.13 (40)	0.83 (309)	−0.70
MONOSYLLABIC STEM	0.44 (101)	0.87 (248)	−0.42

Table 7 shows the mean ASSIBILATION RATIO for the 3 values of the variable FINAL VELAR. As expected, *k*-final stems have the highest mean, whereas the *x*-final stems have the lowest mean.

Table 7: Mean ᴀꜱꜱɪʙɪʟᴀᴛɪᴏɴ ʀᴀᴛɪᴏ for the three values of ꜰɪɴᴀʟ ᴠᴇʟᴀʀ

Mean AR for k (ɴ)	Mean AR for g (ɴ)	Mean AR for x (ɴ)
0.86 (248)	0.54 (82)	0.19 (19)

Both binary predictor variables discussed above, as well as the categorical predictor variable ꜰɪɴᴀʟ ᴠᴇʟᴀʀ were included in a generalised linear mixed model. Additionally, the specific noun was entered as a random variable. In (4) I provide the formula for the model as implemented in R using the package lme4. The complete script is published in Simonović (2024). The results are summarised in Table 8.

(4)
```
model2 <- glmer(assib ~  finalvelar + anim + mono + (1 | noun),
    family = binomial(link = "logit"), data = analysis2)
```

Table 8: Generalised linear mixed model results

Variable	Coefficient	Std. Error	z value	Pr(> |z|)	Odds Ratio
(Intercept)	9.4696	0.4308	21.980	$< 2 \times 10^{-16}$***	1.296×10^4
ꜰɪɴᴀʟ ᴠᴇʟᴀʀ:G	−5.2410	0.8887	−5.897	1.23×10^{-9}***	0.00529
ꜰɪɴᴀʟ ᴠᴇʟᴀʀ:X	−10.0896	1.6726	−6.032	1.62×10^{-9}***	4.15×10^{-5}
ᴀɴɪᴍᴀᴄʏ	−9.9452	1.2072	−8.238	$< 2 \times 10^{-16}$***	4.80×10^{-5}
ᴍᴏɴᴏꜱʏʟʟᴀʙɪᴄ ꜱᴛᴇᴍ	−6.3654	0.7886	−8.072	6.90×10^{-16}***	0.00172

First, it is important to note that the categorial variable ꜰɪɴᴀʟ ᴠᴇʟᴀʀ was dummy coded, resulting in two of its values appearing in the list. The baseline value, ꜰɪɴᴀʟ ᴠᴇʟᴀʀ:k, is used as the reference category for comparison.

The model results indicate that all incorporated predictors exhibit a negative association with ᴀꜱꜱɪʙɪʟᴀᴛɪᴏɴ. Specifically, for the variable ꜰɪɴᴀʟ ᴠᴇʟᴀʀ, which was omitted in Study 1, both g and x demonstrate negative associations with the outcome. Similarly, ᴀɴɪᴍᴀᴄʏ maintains a negative association, consistent with the findings of Study 1. Notably, unlike in Study 1, ᴍᴏɴᴏꜱʏʟʟᴀʙɪᴄ ꜱᴛᴇᴍ also shows a negative association.

5 Discussion

Having presented the results of the two corpus studies, an evaluation of the three types of factors presented in Section 2 is in order.

Regarding the phonological factors, C_1-SONORITY has clearly come out as an important predictor in Study 1, as did the FINAL VELAR in Study 2. Indirectly, the difference between the mean ASSIBILATION RATIO in the two studies (0.33 vs. 0.75), although not statistically tested, points in the direction of a more general influence of the sonority of the segment preceding the stem-final velar.[8] The issue is somewhat less clear when it comes to the factor MONOSYLLABIC STEM, which only came out as significant in Study 2.

As regards the morphological factor -I[GEN.PL] RATIO, its relation with ASSIBILATION was found to be statistically significant, but of relatively weak magnitude compared to the other factors. Especially in the context of the ongoing debate on the relevance of paradigms for phonological computation (see, e.g., Bobaljik 2008), the presented findings cannot be taken as firm evidence that other paradigm cells influence assibilation.

Finally, ANIMACY unequivocally emerges as an influential factor in determining the application of assibilation. The described pattern then joins other, better described and understood, animacy effects in BCMS morphology. Animacy has been well described to influence the exponence of ACC.SG in the main masculine declension in BCMS, leading to minimal pairs such as, e.g., [tip-a] 'guy.ACC.SG' vs. [tip] 'type.ACC.SG'. The influence of animacy on BCMS tonal patterns has also been discussed in the literature, especially for the DAT/LOC.SG ending [-ú], which seems to realise its underlying High tone only in inanimate monosyllables, leading to minimal pairs such as [tíip-u] 'guy.DAT/LOC.SG' and [tiip-ú] 'type.DAT/LOC.SG' (vs. [tíip-a] 'guy/type.GEN.SG'; see Martinović 2012 for a recent quantitative analysis).

Prima facie, the assibilation pattern seems much more gradient than the other two animacy-controlled patterns. The closest we get to a categorical effect is the blocking of assibilation in animates. It is therefore worthwhile to take a closer look at the exceptional animates that display assibilation. The main insight is that there are extremely few animates that display assibilation more often than not (i.e. have assibilation ratios above 0.5). In Study 1, these are only 3 (out of 77 animates): [majka] 'mother', [pomajka] 'foster mother' and [djevojka] 'girl(friend)'. In Study 2, out of 40 animate nouns, 5 have assibilation ratios higher than 0.5: [supruga] 'wife', [unuka] 'granddaughter', [sluga] 'servant', [sʋastika] 'sister-in-

---

[8]The difference would have been even bigger if stems ending in consonant clusters in /g/ and /x/ had been included in Study 1. I am not aware of a single noun from this group that undergoes assibilation in modern BCMS. Barić et al. (1997: 154) mention [kaʋga] 'conflict' as the only Cg-final stem that undergoes assibilation, but most modern speakers seem to either not know this word or use it without assibilation. Including such items in Study 1 would have then additionally lowered the ASSIBILATION RATIO in this study

law' and [vladika] 'bishop'. The fact that all of these nouns refer to roles suggests that roles might belong to a distinct category between animates and inanimates. If this holds true, we can assert that there exists a clear prohibition on assibilation within the category of true animates in the DAT/LOC.SG, and that the possibility of assibilation emerges for entities falling lower on the animacy hierarchy, such as roles and (other) inanimates.

6 Conclusion

The present study aimed to investigate the influence of phonological, morphological and semantic factors on the application of assibilation in DAT/LOC.SG in BCMS nouns. Through a comprehensive analysis of data and statistical modeling, it has become evident that ANIMACY plays a central role in determining the occurrence of assibilation in this context.[9] A detailed analysis of the individual exceptions to the generalisation that animates do not allow assibilation showed that assibilation is restricted to animates that have the meaning of roles.

While providing a complete formal account of the observed pattern is reserved for future research, the results of this study facilitate the formulation of desiderata for such an account. In the spirit of advancing incrementally, the following steps are suggested to establish a connection between the phenomenon described here and its closest related phenomena.

The most closely related phenomenon appears to be the tonal pattern observed in the DAT/LOC.SG forms of the main masculine declension. The two phenomena both exhibit a more intimate phonological interaction with case endings in inanimates compared to animates. This interaction is manifested as a tonal shift in one case and as assibilation in the other. However, a notable distinction lies in the fact that masculine declension roles do not permit the imposition of the DAT/LOC.SG ending's tonal pattern.

The next in line closely related domain is the occurrence or absence of assibilation elsewhere in the nominal and adjectival declensions. The DAT/LOC.SG data presented above suggest the presence of a boundary that hinders phonological interactions between the stem and the case ending in animates. However, this boundary seems to disappear in the plural cases of the masculine declension, where animates undergo assibilation without restriction (e.g., in [t͡ʃex] 'Czech man.NOM.SG', [t͡ʃesi] 'Czech man.NOM.PL', [t͡ʃesima] 'Czech man.DAT/LOC/INS.PL').

---

[9]As argued by one of the reviewers, the statistical tests used here are telling in terms of the statistical significance of the coefficients rather than on predictive power for novel data. We leave the latter type of analysis to future works.

Conversely, it is noteworthy that the adjectival declension never allows assibilation, despite having numerous *i*-initial case endings.

Future research will also profit from more extensive data collection, not only from corpora, but also from elicited production, wug experiments, etc. It is worth noting, that although the present study encompassed a sizable data sample, certain nouns had to be excluded due to the absence of encountered forms, particularly in Study 1, where the absence of GEN.PL forms led to the exclusion of many nouns with attested DAT/LOC.SG forms. Moreover, it is possible that there are further factors which were not included in the analysis.

Finally, an important aspect that was not addressed here is the precise representation of animacy. The observed consistency in assibilation among animate entities suggests the possibility of formalising animacy as the presence of an additional feature or structure.

Overall, the findings of this study contribute to our understanding of the intricate relationship between animacy and phonological processes in BCMS. While animacy's influence on other aspects of BCMS morphology has been previously described, this study unveils a novel finding by demonstrating its comprehensive impact on the application of segmental phonological alternations.

Abbreviations

ACC	accusative	LOC	locative
DAT	dative	NOM	nominative
GEN	genitive	PL	plural
IMP	imperative	SG	singular
INS	instrumental	VOC	vocative

Acknowledgments

References

Barić, Eugenija, Mijo Lončarić, Dragica Malić, Slavko Pavešić, Mirko Peti, Vesna Zečević & Marija Znika. 1997. *Hrvatska gramatika* [Croatian grammar]. Zagreb: Školska knjiga.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using {lme4}. *Journal of Statistical Software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.

Bobaljik, Jonathan. 2008. Paradigms (optimal and otherwise): A case for skepticism. In Asaf Bachrach & Andrew Nevins (eds.), *Inflectional identity*, 29–54. Oxford: Oxford University Press.

Hudeček, Lana. 2022. O podravki koja radi u Podravci, Pepeljugi i Branki. *Hrvatski jezik: znanstveno-popularni Časopis za kulturu hrvatskoga jezika* 9(4). 28–32.

Lečić, Dario. 2015. Morphological doublets in Croatian: The case of the instrumental singular. *Russian Linguistics* 39. 375–393. DOI: 10.1007/s11185-015-9152-7.

Lečić, Dario. 2016. *Morphological doublets in Croatian: A multi-methodological analysis.* Sheffield: University of Sheffield. (Doctoral dissertation).

Ljubešić, Nikola & Filip Klubička. 2014. bs,hr,srWaC - web corpora of Bosnian, Croatian and Serbian. In Felix Bildhauer & Roland Schäfer (eds.), *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, 29–35. Gothenburg: Association for Computational Linguistics. DOI: 10.3115/v1/W14-0405.

Lüdecke, Daniel. 2024. *sjPlot: Data Visualization for Statistics in Social Science.* R package version 2.8.16. https://CRAN.R-project.org/package=sjPlot.

Maretić, Tomislav. 1963. *Gramatika hrvatskoga ili srpskoga književnog jezika.* Zagreb: Matica hrvatska.

Marković, Ivan. 2018. *Hrvatska morfonologija* [Croatian morphonology]. Zagreb: Disput.

Martinović, Martina. 2012. The interaction of animacy with phonetic and phonological factors in Neoštokavian pitch accents. In Jaehoon Choi, E. Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz & Alex Trueman (eds.), *Proceedings of the 29th West Coast Conference on Formal Linguistics*, 161–168. Somerville.

Pešikan, Mitar, Jovan Jerković & Mate Pižurica. 2010. *Pravopis srpskoga jezika, izmenjeno i dopunjeno izdanje* [Orthography of Serbian: Revised and expanded edition]. Novi Sad: Matica srpska.

Petrović, Dragoljub & Snežana Gudurić. 2010. *Fonologija srpskog jezika* [Phonology of Serbian]. Beograd: Institut za srpski jezik SANU / Beogradska knjiga / Matica Srpska.

Simonović, Marko. 2024. *Sibilarisation in the Dative/Locative Singular in BCMS (Data from hrWaC and srWac) [Data set].* DOI: 10.5281/zenodo.13825333.

Težak, Stjepko. 1986. Sibilarizacija u suvremenomu hrvatskom književnom jeziku. *Filologija* 14. 395–402.

Težak, Stjepko & Stjepan Babić. 1992. *Gramatika hrvatskoga jezika: Priručnik za osnovno jezično obrazovanje.* Školska knjiga.

Thornton, Anna M. 2019. Overabundance: A canonical typology. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation* (Studies in Morphology 5), 223–258. Cham: Springer.

Appendix A Plots for predicted probabilities

Below I report the plots for predicted probabilities for each predictor. These were obtained in R using the package sjPlot (Lüdecke 2024) in R. The formula was implemented as `plot_model(model1, type = "pred", terms = "name_of_the_predictor")`.

A.1 Study 1
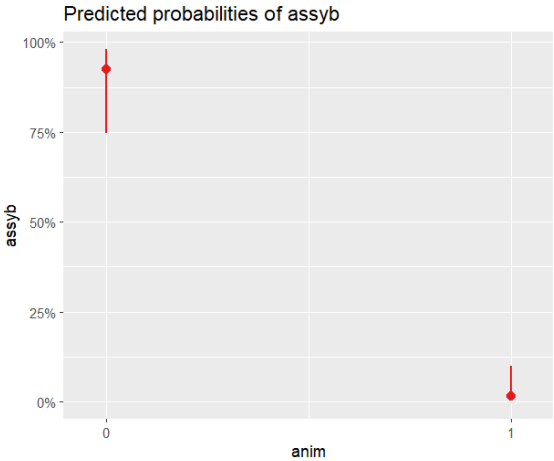
Figure 1: Animacy
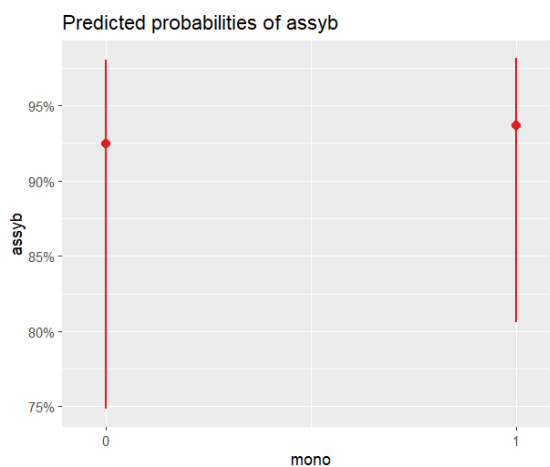
Figure 2: monosyllabic stem
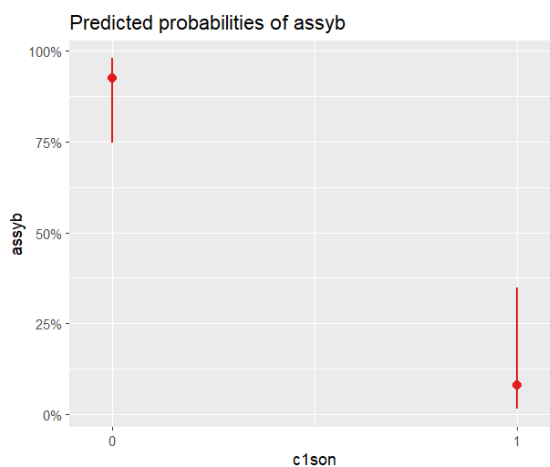
Figure 3: C_1-sonority
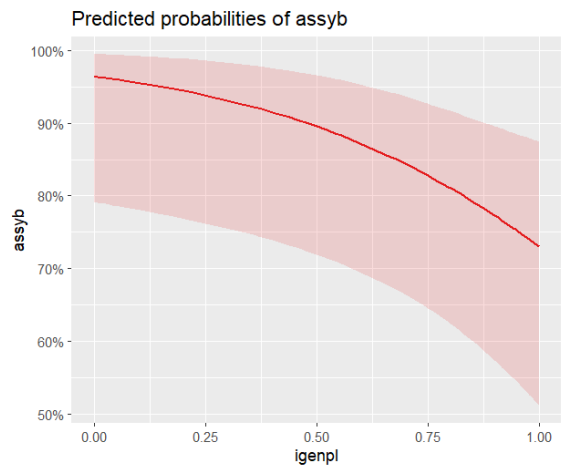
Figure 4: ɪ[ɢᴇɴ.ᴘʟ] ratio

A.2 Study 2
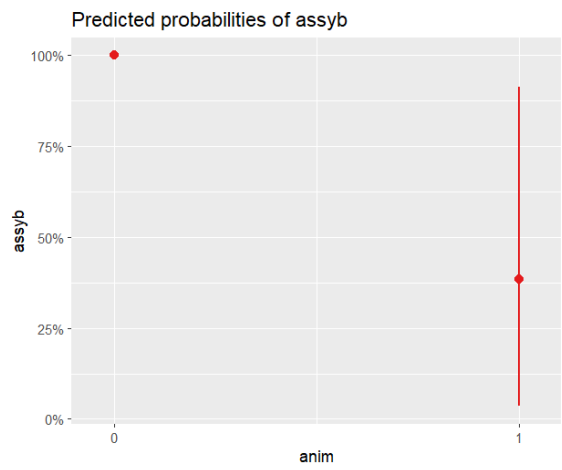
Figure 5: Animacy
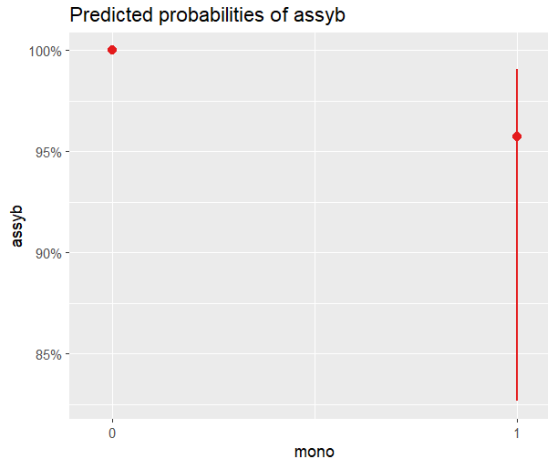
Figure 6: monosyllabic stem
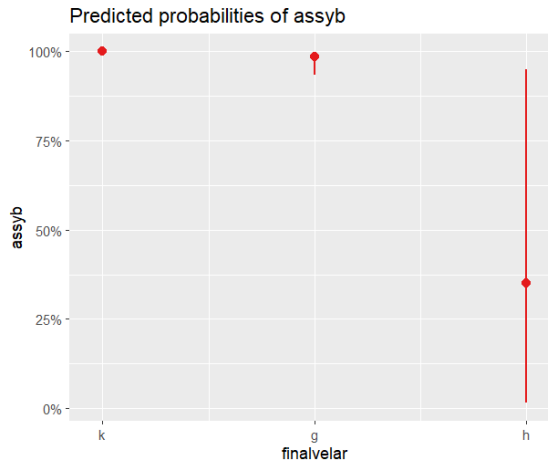
Figure 7: Final Velar

Appendix B R session info

```
R version 4.4.1 (2024-06-14 ucrt)
Platform: x86_64-w64-mingw32/x64
Running under: Windows 10 x64 (build 19045)
```

Matrix products: default
Locale:

```
[1] LC_COLLATE=Dutch_Netherlands.utf8
    LC_CTYPE=Dutch_Netherlands.utf8
    LC_MONETARY=Dutch_Netherlands.utf8
    LC_NUMERIC=C
    LC_TIME=Dutch_Netherlands.utf8
```

Time zone: Europe/Vienna
tzcode source: internal

Attached base packages:

```
[1] stats     graphics  grDevices utils     datasets  methods   base
```

Other attached packages:

```
[1] sjPlot_2.8.16   lme4_1.1-35.5   Matrix_1.7-0    apaTables_2.0.8
    devtools_2.4.5  usethis_3.0.0   psych_2.4.6.26  readxl_1.4.3
```

Loaded via a namespace (and not attached):

```
[1] gtable_0.3.5        xfun_0.47          ggplot2_3.5.1
    htmlwidgets_1.6.4   remotes_2.5.0      insight_0.20.3      lattice_0.22-6
    sjstats_0.19.0      vctrs_0.6.5        tools_4.4.1         generics_0.1.3
    datawizard_0.12.2   parallel_4.4.1     tibble_3.2.1        fansi_1.0.6
    pkgconfig_2.0.3     RColorBrewer_1.1-3 ggeffects_1.7.0     lifecycle_1.0.4
    farver_2.1.2        compiler_4.4.1     stringr_1.5.1       sjmisc_2.8.10
    munsell_0.5.1       mnormt_2.1.1       httpuv_1.6.15       htmltools_0.5.8.1
    later_1.3.2         pillar_1.9.0       nloptr_2.1.1        urlchecker_1.0.1
    tidyr_1.3.1         MASS_7.3-60.2      ellipsis_0.3.2      cachem_1.1.0
    sessioninfo_1.2.2   boot_1.3-30        nlme_3.1-164        mime_0.12
    sjlabelled_1.2.0    tidyselect_1.2.1   digest_0.6.37       performance_0.12.2
    stringi_1.8.4       dplyr_1.1.4        purrr_1.0.2         labeling_0.4.3
    splines_4.4.1       fastmap_1.2.0      grid_4.4.1          colorspace_2.1-1
    cli_3.6.3           magrittr_2.0.3     pkgbuild_1.4.4      utf8_1.2.4
```

```
broom_1.0.6       withr_3.0.1       scales_1.3.0     promises_1.3.0
backports_1.5.0   cellranger_1.1.0  memoise_2.0.1    shiny_1.9.1
knitr_1.48        miniUI_0.1.1.1    profvis_0.3.8    rlang_1.1.4
Rcpp_1.0.13       xtable_1.8-4      glue_1.7.0       pkgload_1.4.0
rstudioapi_0.16.0 minqa_1.2.8       R6_2.5.1         fs_1.6.4
```

Chapter 19

How departicipial are "ʟ-participle" nominalisations in Western South Slavic

Marko Simonović[a], Petra Mišmaš[b], Stefan Milosavljević[a], Boban Arsenijević[a], Katarina Gomboc Čeh[b], Franc Lanko Marušič[b] & Rok Žaucer[b]

[a]University of Graz [b]University of Nova Gorica

We focus on nominalisations seemingly derived from ʟ-participles, illustrated by *lec*-nominalisations in Slovenian, in order to establish the nature and position of the ʟ-morpheme as well as the structure of these nominalisations in general. Our research is situated in the current debates on whether the item ʟ in ʟ-participles and ʟ-nominalisations is the same morpheme or two different morphemes, and if the former, whether ʟ-nominalisations are derived from ʟ-participles. We argue that the ʟ-morpheme is a root in both, but also show that it is not the case that *lec*-nominalisations contain ʟ-participles. The *lec*-nominalisations are argued to contain a smaller structure than the corresponding ʟ-participle, which is also reflected in the set of theme vowels possible in these nominalisations.

1 Introduction

One productive strategy to derive deverbal agentive nominalisations in Slovenian (Slo) is with the item *-lec*, which shares its first segment with the ʟ-participle (termed the past participle in much of the traditional literature). This is illustrated in (1).

(1) br-a-ti — br-a-l-a — br-a-l-ec   (Slo)
     read-TV-INF read-TV-L.PTCP.F.SG read-TV-L-*er*
     'to read'   '(she) read'   'reader'

This is not an isolated example of such nominalisations in Western South Slavic. The same type can be found in Bosnian/Croatian/Montenegrin/Serbian [BCMS], where the item *-lac* has the same structure, (2).

(2)   čit-a-ti   – čit-a-l-a   – čit-a-l-ac   (BCMS)
       read-TV-INF   read-TV-L.PTCP-F.SG   read-TV-L-*er*
       'to read'   '(she) read'   'reader'

In both languages, the string *-ec/-ac* (glossed as *-er* in the examples above) is also an attested nominal suffix, as illustrated in, e.g., *Slovenec/Slovenac* 'a Slovenian'. This fact prompts several authors to analyse the string *-lec* as consisting of two morphological units, L and *-ec/-ac* (see §2 for references or Birtić 2008 for an overview). And while the function/contribution of *-ec/-ac* seems to be unproblematic, the question whether *lec/lac*-nominalisations (and other comparable derivations) contain the L-participle has been posed and answered differently both in traditional descriptive work and in formal approaches.[1]

Given the pattern in (1), the "L-is-participial" analysis may be the most straightforward one. Such an approach would mean that *lec*-nominalisations join a broader class of departicipial nominalisations, which also include nominalisations illustrated in (3), standardly analysed as derived from the passive participles (e.g., Toporišič 2000).

(3)   anketir-a-ti   – anketir-a-n   – anketir-a-n-ec   (Slo)
       survey-TV-INF   survey-TV-PASS.PTCP   survey-TV-PASS.PTCP-*er*
       'to survey'   'surveyed'   'respondent'

Moreover, if derived from a participle, *lec*-nominalisations can be taken to be similar to other agentive nominalisations that have a form from the verbal paradigm as their base. Such an analysis is possible, for example, for agentive nominalisations in *-telj*, where the base seems to be the short infinitive.[2]

---

[1]In some contexts in Slovenian, *-lec* is written and pronounced as *-vec*, specifically, after some roots ending in a vowel (e.g., *pi-∅-ti, pi-vec* 'to drink, drinker') or in *-l* or *-lj*, e.g., *del-a-ti, del-a-vec* 'to work, worker' (Toporišič 2000: 163–164). We take this to be lexically conditioned allomorphy.

[2]Note that *-lec* and *-telj* are not allomorphs. First, there are a few pairs with these suffixes combined with the same base (e.g., the Slovenian *brani-telj – brani-lec* 'defender' from *braniti* 'to defend' or *hrani-telj – hrani-lec* 'custodian' from *hraniti* 'to keep in custody'). Second, *-telj* is much more consistently related to an agent interpretation (i.e., animate and human; there are only a few exceptions, such as *pokazatelj* 'indicator'). On the other hand, *-lec* can also be associated with an instrument interpretation, see §2. Finally, nominalisations with *-telj* are far less common than nominalisations with *-lec* (see Arsenijević et al. 2024).

(4) predav-a-t – predav-a-t-elj   (Slo)
    lecture-TV-S.INF   lecture-TV-S.INF-*er*
    'to lecture'   'lecturer'

(5) uč-i-t – uč-i-t-elj   (Slo)
    teach-TV-S.INF   teach-TV-S.INF-*er*
    'to teach'   'teacher'

As will be discussed in detail in §3, despite the many similarities, *lec*-nomina-lisations (and nominalisations from short infinitives) are not fully comparable to passive-participle nominalisations, since only the latter preserve the prosody of the participle and allow all theme-vowel classes in the verbal base. In this paper, we therefore revisit the issue of the nature and the contribution of the ʟ-morpheme. The empirical data and the proposed analysis tackle some of the foundational questions of morphology, in particular regarding the status of roots, cycles of computation, and their interactions. While we will focus on Slovenian data, the observations and the analysis can be extended to BCMS. In what follows, the examples are from Slovenian, unless marked otherwise.

Before we continue, a remark is in order on a type of nominalisation that is NOT attested in Western South Slavic, since this gap will inform our analysis. Nominalisations from the three bases shown above (approximately matching the ʟ-participle, passive (N/T)-participle and short infinitive) are, to the best of our knowledge, the only deverbal derivations that preserve the theme vowel of the base verb.[3] In other words, there are no deverbal derivations, such as the hypothetical derivations illustrated in (6), where the root and the theme vowel would directly combine with a hypothetical morpheme *-p* that would not show up in the paradigm of the verb.

(6) a. predav-a-ti – *predav-a-p
       lecture-TV-INF   lecture-TV-*p*
       'to lecture'

    b. uč-i-ti – *uč-i-p
       teach-TV-INF   teach-TV-*p*
       'to teach'

---

[3]The combination of the root and the theme vowel by themselves, without an overt derivational suffix, is also not attested as a derivational pattern (i.e., something like *predava* or *uči* does not occur as a nominalisation). This naturally means that zero-derived nominals in which the theme vowel is not present, e.g., the Slovenian *popis* 'inventory' (related to *popis-a-ti* 'to catalogue'), are not at issue here.

In sum, whenever a theme vowel appears in a nominalisation, it appears embedded under additional (seemingly) functional material. In what follows, we advance an account in which, in the nominalisations, this material, i.e., both the ʟ-morpheme and the passive-participle ɴ/ᴛ-morpheme, correspond to the same conceptually empty root.[4] Focusing on the ʟ, we argue that ʟ-participles contain a richer structure than the corresponding portion of ʟ-nominalisations, while no such difference is found with the passive-participle ɴ/ᴛ-morpheme. We leave agentive nominalisations that are derived from the short infinitive for future work.

The paper is organised as follows. In §2, we discuss (both old and new) reasons for splitting ʟ-initial deverbal suffixes into an ʟ-morpheme and another suffix that is added on top of it (e.g.,-*ec*). In §3, we discuss the nature of the ʟ-morpheme in the nominalisations under consideration. §4 presents our account of the structural position of the ʟ-morpheme and the theme vowel restrictions in the respective nominalisations. §5 concludes the paper.

2 Internal structure: Severing ʟ from *-ec* (& other affixes)

We start the discussion with the internal structure of *-lec* and related derivations in Slovenian. While various other agentive nominalisations also exist, deverbal nominalisations that contain the ʟ-morpheme preceded by a theme vowel are by far the most common in Slovenian (see Marvin 2015, 2016 quoting Stramljič Breznik 1999).[5] Example (7), taken from Marvin (2002: 98, (22)), illustrates three such nominalisations. According to Marvin, the three affixes added to the ʟ-morpheme (-*ec*, -*k* and -∅) are variants of the same affix deriving nouns of three different genders. All three nominalisations are generally related to an external argument, be it an agent or an instrument, with the neuter-gender nominalisation primarily having the instrument interpretation (Marvin 2002: 99, fn. 18, but see fn. 12 for examples in which *-lec* is not associated with agentivity).[6]

---

[4]The notion of a conceptually empty root corresponds to the notion of a light root in the sense of Quaglia et al. (2022), who use this label for secondary-imperfective suffixes; see also §4.2.

[5]Stramljič Breznik (1995) presents counts in which *-ač* emerges as the most frequent affix in agentive nominalisations. This is due to the fact that the author assumes *-ilec* and *-alec* to be two separate affixes. If *-ilec* and *-alec* are taken to instantiate the same item depending on the theme vowel of the base verb (as it is assumed in this paper), the unified item comes out as more frequent than *-ač* in her counts as well.

[6]As pointed out by a reviewer, it is relevant to show at this point that the examples in (7) indeed have nominal properties. The examples in (i) give the same nominalisations in the genitive case (case, number and gender being nominal properties in Slovenian) with an agreeing adjective.

(7) a. brus-i-l-ec-∅
         sharpen-ᴛᴠ-ʟ-*er*-ɴᴏᴍ.sɢ
         'a sharpener.ᴍ'

     b. brus-i-l-k-a
         sharpen-ᴛᴠ-ʟ-*er*-ɴᴏᴍ.sɢ
         'a sharpener.ꜰ'

     c. brus-i-l-∅-o
         sharpen-ᴛᴠ-ʟ-*er*-ɴᴏᴍ.sɢ
         'a sharpening device'

Moreover, it has long been observed that various nominalisations and adjectivisations share the same ʟ-final base. Some further affixes (both nominal and adjectival) that can combine with the ʟ-form are exemplified in (8). Note that the adjectivising items *-en* and *-n* are exponents of the same item, whereby the exponent *-en* [ən] includes an epenthetic vowel. The distribution of the epenthetic vowel is guided by the same rules as for *-ec* (see fn. 7). The two suffixes that we gloss as ᴘʟᴀᴄᴇ are two different items. While *-išč* consistently results in a place interpretation, *-ic* is only associated with this interpretation when in the context of -ʟ and *-n*. While these suffixes are in and of themselves interesting and underexplored, a more detailed account of them is beyond the scope of this paper.

(8) čak-a-l-en | čak-a-l-n-ic-a | čak-a-l-išč-e
     wait-ᴛᴠ-ʟ-ᴀᴅᴊ | wait-ᴛᴠ-ʟ-ᴀᴅᴊ-ᴘʟᴀᴄᴇ-ɴᴏᴍ.sɢ | wait-ᴛᴠ-ʟ-ᴘʟᴀᴄᴇ-ɴᴏᴍ.sɢ
     'waiting.ᴀ' | 'waiting room' | 'waiting spot'

These items cannot be modified by an adverb such as *hitro* 'fast.ᴀᴅᴠ'.

(i) a. *hitro   /   hitrega   brus-i-l-c-a
         fast.ᴀᴅᴠ   fast.ɢᴇɴ.sɢ.ᴍ sharpen-ᴛᴠ-ʟ-*er*-ɢᴇɴ.sɢ
         'a fast   sharpener.ᴍ'

     b. hitre   brus-i-l-k-e
         fast.ɢᴇɴ.sɢ.ꜰ sharpen-ᴛᴠ-ʟ-*er*-ɢᴇɴ.sɢ
         'a fast   sharpener.ꜰ'

     c. hitrega   brus-i-l-∅-a
         fast.ɢᴇɴ.sɢ.ɴ sharpen-ᴛᴠ-ʟ-*er*-ɢᴇɴ.sɢ
         'a fast   sharpening device'

Still, as pointed out in Marvin (2002: 101), such nominalisations can be modified by, for example, manner adverbials, which in fact modify the event included in the nominalisations, in turn implying an event component in these nouns.

The existence of such families of related derivations is a strong argument for the decomposition of *-lec* into ʟ and *ec*, but also an argument for recognising ʟ as a morpheme that is required for the verbal base to combine with derivational affixes, especially given the observation that there are no nominalisations in which the nominalising affix simply combines with the verb stem (i.e. minimally, root + theme vowel), as shown in (6). In other words, as soon as the base is verbal, and marked as such by the presence of the theme vowel, another item needs to 'mediate' in the attachment of the nominaliser.

A decomposition of *-lec* is further supported by the fact that the affixes added to the ʟ-morpheme, as in the examples (7) and (8), also show up in other environments. For example, Marvin (2002) shows that the suffix *-ec* can also be found in various non-verbal environments, i.e., with adjectives, roots, and nouns, as illustrated by (9).[7] An analogous argument can be made for its feminine counterpart *-k*, as shown in (10).[8]

(9) a. bakr-en | bakr-en-ec-∅
         copper-ADJ   copper-ADJ-*er*-NOM.SG
         'made of copper' | 'copper coin'

     b. hod- | hod-ec-∅
         √WALK   walk-*er*-NOM.SG
                 | 'walker'

     c. krog | krog-ec-∅
         circle   circle-*er*-NOM.SG
         'circle' | 'small circle'

(10) a. jekl-en | jekl-en-k-a
         steel-ADJ   steel-ADJ-*er*-NOM.SG
         'made of steel' | 'gas cylinder'

     b. hod- | hod-k-a
         √WALK   walk-*er*-NOM.SG
                 | 'walker'

     c. adidas | adidas-k-a
         Adidas   adidas-*er*-NOM.SG
                 | 'adidas-shoe'

---

[7]The relevant nominalising suffix in Slovenian is really just *-c*, and the vowel in *-ec* [əts] is an epenthetic vowel inserted to avoid a complex coda. As such, the vowel is absent in many forms of each paradigm, such as the dual *bakr-en-c-a* 'copper coin.DU' for (9a), *hod-c-a* 'walker.DU' for (9b), *brusil-c-a* 'sharpener.DU' for (7) etc. We continue to use *-ec* in the text for simplicity.

[8]The situation is somewhat more complicated with nouns. When merging with a masculine nP, as in (9c), the noun with *-ec* will get a diminutive reading. We leave the diminutive interpretation aside at this point.

These facts unequivocally demonstrate that the item *-lec* is complex, comprising two distinct items, namely ʟ and *-ec*. And while the morpheme *-ec* may be treated as invariant between deverbal and other nominalisations in which it occurs (it consistently restricts the denotation to count objects), the status of the ʟ-morpheme in these nominalisations and its relation to the ʟ-morpheme that surfaces in the ʟ-participle is more complex. We address the issue in what follows and argue that this morpheme universally stativises event predicates of various sizes in order to license the derivation of words that are not verbs and that denote the event described by the verbal expression.

3 What is ʟ?

The question of the status of the ʟ-morpheme in *lec*-nominalisations is not a new one. As we show in what follows, both traditional descriptive sources and formal accounts offer a variety of solutions. We start with an overview of traditional accounts.

3.1 Traditional accounts of ʟ in nominalisations

For Slovenian, some authors take *-lec* to be a single morpheme; for example, Toporišič (2000) treats *-lec*, as in example (1), and *-ec*, as in (9), as two separate items. The issue of ʟ in *lec*-nominalisatons is considered in Stramljič Breznik (1999), who mentions as a possible answer Bajec's (1950) proposal that ʟ essentially generalises from neuter-gender nominalisations such as *zija-lo* 'gawker' (form *zijati* 'to gawk'). On the other hand, Bajec et al. (1956) argue that either the neuter-gender nouns with *-lo* or ʟ-participles can serve as the derivational base for *lec*-nominalisations.

   Similar proposals also exist for BCMS, where a common denominator of the accounts which propose a single suffix *-lac* (Maretić 1963, Babić 2002, Klajn 2003) is the assumption that the ʟ-participle encodes past. Given that ʟ-participles are also used in past tense in Slovenian, such an assumption could easily be extended to Slovenian. However, this assumption has little empirical ground, since the ʟ-participle is used in a variety of syntactic contexts in Slovenian, e.g., with the conditional or the future tense, where it does not receive 'past' interpretation.[9]

---

[9]In fact, even as a part of the perfect form, traditionally analysed as past tense in BCMS, its meaning varies between the 'past' interpretation and the present perfect, as extensively argued in Todorović (2016). This is exemplified in (i) and also holds for Slovenian.

(i) Jeo   sam.   (BCMS)
     eaten.M.SG AUX.1.SG
     'I ate/I've eaten. (i.e., I'm not hungry.)'

(11) je   hodil | bo   hodil
     AUX walk.PTCP AUX walk.PTCP
     'has walked' | 'will walk'

We therefore do not take this as an argument against the decomposition of *-lec*.

Summing up, three options seem to emerge in the traditional literature: (i) *-lec* is a single suffix, (ii) *lec*-nominalisations are derived from L-participles and (iii) L in *lec*-nominalisations spreads from *lo*-nominalisations (where it is unclear what L in *-lo* is).

3.2 Formal accounts of L and theme vowels in nominalisations (and beyond)

3.2.1 There is no single L (Marvin 2002)

The account, which will in many ways serve as the starting point of our analysis, is the account of Slovenian *lec*-nominalisations in Marvin (2002), couched in Distributed Morphology (Halle & Marantz 1993, 1994). Considering the identity of the L-morpheme, Marvin (2002) proposes that L in these nominalisations is the participial L. This L in turn corresponds to the featureless Elsewhere Vocabulary Item that gets inserted in the T_2/Participle head.[10] And since in Slovenian *-lec* denotes an external argument, be it an agent or an instrument, Marvin (2002) proposes that *-ec* is merged in the agentive position (Spec of *v*P) and undergoes subject movement to the assumed SpecTP, resulting in the correct order in the linearised structure. Figure 1 provides the relevant structure before the movement of the nominal *-ec* (as given in Marvin 2002, 99, (25)).[11] Placing *-ec* in the external argument position in the described structure can be seen as predicting that only unergative and transitive verbs are able to form these nominalisations, while unaccusatives will not be able to do so (since, as stated in Marvin 2002: 99, unaccusatives do not have an external argument position). This tentative prediction is confirmed by the WeSoSlaV database (Arsenijević et al. 2024), but see also Marjanovič et al. (2013). Out of 728 *lec*-nominalisations in the Slovenian subbase of WeSoSlaV, only 3 can be taken to be derived from unaccusatives, and the majority of unaccusative verbs, such as *porumeneti* 'to become yellow', do

---

[10]Marvin (2002) distinguishes between two T(ense) heads, T_1 and T_2, whereby the latter corresponds to participles.

[11]Following Marvin (2002: 105-107), the inflectional ending carrying number and case agreement is inserted in the Number head when the nominalisation is used in a sentence, and this head nominalises the structure. The NumberP is, on her account, embedded under a DP.

not derive *lec*-nominalisations (**porumenelec* '(intended) someone who becomes yellow').[12]
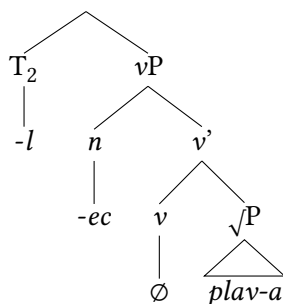
Figure 1: Structure proposed by Marvin (2002: 99, (25)).

The analysis just outlined, under which these nominalisations are derived from ʟ-participles, finds its support in the fact that virtually all *lec/lac*-nominalisations can be derived from an ʟ-participle of an existing verb. However, if we take as a starting point existing *lec/lka/lo*-nominalisations and work our way back towards a participial base, we soon find nominalisations which contain ʟ preceded by a combination of a root and a theme vowel that cannot be found in an attested verb. Such cases, exemplified by the last example in each row in (12), are considered in Marvin (2002), who treats them as (non-compositional) "root *l*-participle nominalisations". In these nominalisations the root together with the theme vowel is the complement of a Part(iciple)P (headed by the ʟ-morpheme), which gets nominalised by *-ec/k/∅*. Crucially, Marvin argues that in these root nominalisations (unlike the de*verbal lec*-nominalisation) the nominalised structure does not include a *v*-head. Consequently, these nominalisations (e.g. *rezilo*, unlike *rezalo*) are argued to exhibit a lack of an event component (cf. fn. 6), and of an external agent position (Spec*v*P).

---

[12]In fact, even these three examples can be successfully accounted for under an alternative analysis. That is, the set of *lec*-nominalisations that *prima facie* seem to be derived from unaccusatives consists of *pogorelec* 'victim of a fire' (from *pogoreti* 'burn down'), *otrdelec* 'something hardened (usually penis)' (from *otrdeti* 'harden') and *osamelec* 'something isolated (usually tree or hill)' (from *osameti* 'become alone'). As is clear from the translations, all of these have a very specific interpretation which is never agentive. Furthermore, all of these items can be argued to be deadjectival. As shown by Aljović (2000) for BCMS and Simonović & Mišmaš (2022) for Slovenian, unaccusatives can derive adjectival ʟ-participles which have full adjectival paradigms and can serve as bases for further derivation (e.g., *-ost*-nominalisation, in *osamelost* 'the property of being left alone'). Taking this into consideration, the three *lec*-nominalisations that seem to be derived from unaccusative verbs may well be derived from adjectives and therefore lack the agentive interpretation.

(12) a. rez- | rez-a-ti | rež-e-mo | rez-a-l-∅-o | rez-i-l-∅-o
      $\sqrt{\text{CUT}}$ | 'cut.INF' | 'cut.PRS.1PL' | 'cutter' | 'blade'

    b. barv- | barv-a-ti | barv-a-mo | barv-a-l-∅-o |
      $\sqrt{\text{COLOUR}}$ | 'colour.INF' | 'colour.PRS.1PL' | 'colouring device' |
      barv-i-l-∅-o
      'pigment'

    c. god- | gos-∅-ti /god-∅-ti/ | god-e-mo | god-a-l-∅-o
      $\sqrt{\text{PLAY}}$ | 'play.INF' | 'play.PRS.1PL' | 'string instrument'

The fact that we can still observe an L item and a theme vowel in examples like *rezilo, barvilo, godalo*, despite the lack of a v^0, is important.

On Marvin's analysis, neither the theme vowel nor the Part(iciple) projection are inherently linked to the verbal domain. Marvin (2002: 110) takes the L (and also the theme vowel) in these nominalisations to be a part of an extended root and proposes that "its meaning is non-compositional (encyclopedic) as if it were a regular bare root with some extra pieces of morphology, to which then a nominaliser is added in root nominals in general." She further states that "it appears that the language is making use of the process of root extension to introduce new non-compositional meaning that for some reason could not be introduced by nominalising just a bare root" (Marvin 2002: 110, 111). Finally, as for the theme vowels, which in these nominalisations are restricted to the set of two (*i* and *a*), Marvin states that they are the default theme vowels in the language, but does not further elaborate on how they are assigned. To sum up the proposal in Marvin (2002), some nominalisations that include L are taken to be derived from L-participles, while others include a "root extending" L.

Extending our empirical base, the small set of *lo*-nominalisations with a theme vowel switch can be complemented by the even smaller set of *lo*-nominalisations for which no corresponding verb can be found. Despite the fact that no independently attested verbal base is available, these nouns are interpreted as instruments and their theme vowels also come from the set of two: *a* and *i*. Examples of these nominalisations given in (13) come from Simonović (2020).

(13) a. glasb-a | glasb-i-l-∅-o
      music-NOM.SG music-TV-L-*er*-NOM.SG
      'music' | 'musical instrument'

    b. / | zrc-a-l-∅-o
      mirror-TV-L-*er*-NOM.SG
      | 'mirror'

Crucially for the analysis of ʟ, the fact that these examples are not derived from a verb clearly implies that they are also not derived from ʟ-participles.

3.2.2 ʟ is a root, and the importance of the theme vowel set

An account of the ʟ that participates in deverbal derivations is given in Simonović & Mišmaš (2022), where the focus is on adjectival ʟ-participles. This work endorses a DM framework, but the authors assume a specific approach to derivational affixes proposed by Lowenstamm (2014). Under this approach, all derivational affixes are viewed as transitive (or "bound") roots. This means that, on the one hand, these roots require a complement (either a phrase or a root), but they also project and can be embedded under a categorial head or selected by another root. This approach then crucially separates typical traditional derivational affixes into roots (which are acategorial, as are, in accordance with Marantz 2001, all "free" roots, i.e. roots that do not require a complement, e.g., $\sqrt{\text{DOG}}$) and categorial heads.[13] In this pairing, categorial heads are typically phonologically empty and roots have semantic and/or phonological content.

Assuming this approach, Simonović & Mišmaš (2022) discuss two types of participles – verbal ʟ-participles that we can find in complex tenses, and adjectival ʟ-participles, which only derive from unaccusative verbs (see fn. 12), arguing that ʟ is a root in both. This root can merge with either a root or a phrase. In adjectival participles, ʟ is merged with a root, whereas in past participles it is merged with a verbalised structure. In addition, taking ʟ to be a root then allows Simonović & Mišmaš (2022) to offer a unified account even for ʟ beyond the verbal domain, e.g., in the noun *krog-l-a* 'sphere', related more directly to the noun *krog* 'circle' than to the verb *krož-i-ti* 'to circle'.

Interestingly, if we zoom in on allowed theme vowel classes, there is a discrepancy between the adjectival ʟ-participles in Simonović & Mišmaš (2022) and the derivations discussed in this paper. The set of morphemes that can precede the ʟ-morpheme in the adjectival participles in Simonović & Mišmaš (2022) prominently *excludes* the theme vowels *a* and *i*, which are by far the most common theme vowels in Western South Slavic (in Marušič et al. 2022, 1504 out of 3000 verbs in the Slovenian part of the database have the theme vowel *a* in the nonfinite forms, which are relevant here, and 863 have *i*; the situation is similar in

---

[13]This division is motivated by derivational affixes that (under a classic DM view) realise different categorial heads. One such example is the English *-an* which can appear in nouns (*librarian*) or adjectives (*reptilian*), examples from Lowenstamm (2014: 233). An alternative view, according to which only affixes that are associated with different categories are roots, is presented in Creemers et al. (2018).

BCMS). On the other hand, in the nominalisations and adjectivisations discussed in this paper, what can precede the ʟ-morpheme is exactly this set of theme vowels. This holds not only for examples such as those in (13), where the base verb is not attested independently. It also holds for agentive *lec*-nominalisations. (14) illustrates *lec*-nominalisations from four transitive verbs where theme vowels ∅ and *e* are replaced by *i*.

(14)   a.   ves-∅-ti /vez-∅-ti/ – vez-e-l /vez-∅-l/ – vez-i-l-ec
           embroider-ᴛᴠ-ɪɴꜰ   embroider-ᴛᴠ-ʟ   embroider-ᴛᴠ-*er*.ɴᴏᴍ.sɢ
           'to embroider'   'embroidered'   'embroiderer'

       b.   ples-∅-ti /plet-∅-ti/ – plet-e-l /plet-∅-l/ – plet-i-l-ec
           knit-ᴛᴠ-ɪɴꜰ   knit-ᴛᴠ-ʟ   knit-ᴛᴠ-*er*.ɴᴏᴍ.sɢ
           'to knit'   'knitted'   'knitter'

       c.   gnes-∅-ti /gnet-∅-ti/ – gnet-e-l /gnet-∅-l/ – gnet-i-l-ec
           knead-ᴛᴠ-ɪɴꜰ   knead-ᴛᴠ-ʟ   knead-ᴛᴠ-*er*.ɴᴏᴍ.sɢ
           'to knead'   'kneaded'   'kneader'

       d.   vrt-e-ti   – vrt-e-l   – vrt-i-l-ec
           spin-ᴛᴠ-ɪɴꜰ   spun-ᴛᴠ-ʟ   spin-ᴛᴠ-ʟ-*er*.ɴᴏᴍ.sɢ
           'to spin'   'spun'   'spinner'

This change in the theme vowel is noted also in Toporišič (2000: 163–164), who states that perhaps the affix is not *-lec* but is rather *V-lec*, where V is the final vowel of the stem (e.g., *a* or *i*), but if the stem is consonantal (i.e., the verb has a ∅ theme vowel), the vowel is realised as *i*. Based on the observation that verbs with a ∅ theme vowel are nominalised with *-ilec*, Stramljič Breznik (1999) concludes that this form is the least marked option.

In the context of formal accounts, on the other hand, the discrepancy in theme vowels present in adjectival ʟ-participles and *lec*-nominalisations can be taken as a consequence of different structures. While *lec*-nominalisations have convincingly been shown to include a verb phrase, see §3.2.1, in adjectival ʟ-participles the complement of ʟ has been argued to be a root (Simonović & Mišmaš 2022).[14] And yet, if we assume that theme vowels surface as exponents of the verbalising head v^0 (as proposed in Quaglia et al. 2022 for Slavic, Milosavljević & Arsenijević 2022 for Serbo-Croatian; see also Svenonius 2004 for Russian, Biskup 2019 for Czech), the presence of the theme vowels *a* and *i*, the two most productive verbal theme vowels, in all nominalisations with ʟ under discussion, even the ones in

---

[14]Note that the same restriction to the theme vowels *a* and *i* is attested in much less productive nominalisations derived from short infinitives.

(12), implies that all these nominalisations are in fact deverbal. This means that both ʟ-participles and *lec*-nominalisations have a verbal structure, but also need to be different in some way. Put differently, if the set of allowed theme vowels is a reliable diagnostic for differentiating between different structural environments, then *lec*-nominalisations (where, again, only *a* and *i* are allowed) is a different environment from the ʟ-participle, where all theme vowels are allowed.

We propose a specific solution in §4.

3.2.3 Change in the secondary imperfective

Another argument that the base onto which *-ec* is added is not the verbal participle is offered by the fact that in some cases the secondary imperfective morpheme also does not match between the verbal "base" and the *-lec* nominalisation. (15) gives two such nominalisations.

(15) a. obračun-a-ti   – obračun-av-a-ti   – [??]obračun-ov-a-ti   –
          calculate-TV-INF   calculate-SI-TV-INF   calculate-SI-TV-INF
          'calculate.PFV'   'calculate.IPFV'   'calculate.IPFV'
          obračun-ov-a-l-ec
          calculate-SI-TV-L-*er*
          'calculator'

      b. prikim-a-ti – prikim-av-a-ti – [??]prikim-ov-a-ti – prikim-ov-a-l-ec
          nod-TV-INF   nod-SI-TV-INF   nod-SI-TV-INF   nod-SI-TV-L-*er*
          'nod.PFV'   'nod.IPFV'   'nod.IPFV'   'nodder'

We do not provide a full analysis of these examples here, but rather leave this for future work.

3.2.4 Prosody and ʟ

The final type of evidence featuring in DM approaches to *lec*-nominalisation is their prosodic behaviour. All nominalisations and adjectivisations containing the ʟ-morpheme share the same prosodic pattern, i.e., stress on the theme vowel, which overrides the lexical prosody of the base verb (if available). This is illustrated in (16), where the verbal bases do not all have the same stress pattern, as can be seen in the ʟ-participles, but these differences get neutralised in all other cases.[15]

---

[15]A comparable pattern is observed in Caha & Ziková (2022) for Czech, where, in terms of vowel length, all verbal forms have the same allomorph, but the nominalisation has a different one.

(16) a. moˈr-i-l   vs. | moˈr-i-l-ec   –
     murder-TV-L.M.SG   murderer-TV-L-*er*.NOM.SG
     'murdered' | 'murderer.M'
     moˈr-i-l-k-a   – moˈr-i-l-en
     murderer-TV-L-*er*.NOM.SG   murder-TV-L-ADJ.M.SG'
     'murderer.F'   'related to murder'

     b. ˈmer-i-l   vs. | meˈr-i-l-ec   –
     measure-TV-L.M.SG   measure-TV-L-*er*.NOM.SG
     'measured' | 'measurer.M'
     meˈr-i-l-k-a   – meˈr-i-l-en
     measure-TV-L-*er*.NOM.SG   measure-TV-L-ADJ.M.SG
     'measurer.F'   'related to measuring'

     c. ˈrez-a-l   vs. | reˈz-a-l-ec   – reˈz-a-l-∅-o   –
     cut-TV-L.M.SG   cut-TV-L-*er*.NOM.SG   cut-TV-L-*er*.NOM.SG
     'cut' | 'cutting person.M'   'cutter'
     reˈz-i-l-∅-o
     cut-TV-L-*er*.NOM.SG
     'blade'

     d. iˈgr-a-l   vs. | iˈgr-a-l-ec | iˈgr-a-l-k-a |
     play-TV-L.M.SG   player-TV-L-*er*.NOM.SG   player-TV-L-*er*.NOM.SG
     'played' | 'player.M' | 'player.F' |
     iˈgr-a-l-∅-o
     play-TV-L-*er*.NOM.SG
     'playground equipment'

In stark contrast to the nominalisations that contain the L-morpheme, those that contain the passive participle (the N/T-participle) behave as stress-preserving, as illustrated in (17). Here in each case the prosodic pattern of the passive participle is preserved in all further derivations.

(17) a. ˈmerjen   vs. | ˈmerjen-ec   – ˈmerjen-ka   –
     measure.PASS.PTCP   measured-*er*.M   measured-*er*.F
     'measured' | 'measured person.M'   'measured person.F'
     ˈmerjen-je
     measured-*ing*
     'measuring'

     b. umorˈjen   vs. | umorˈjen-ec   umorˈjen-ka
     murder.PASS.PTCP   murdered-*er*.M   murdered-*er*.F
     'murdered' | 'murdered person.M'   'murdered person.F'

c. 'pitan   vs. | 'pitan-ec   – 'pitan-ka   – 'pitan-je
    fatten.PASS.PTCP   fattened-*er*.M   fattened-*er*.F'   fattened-*ing*
    'fattened' | 'fatling.M'   'fatling.F'   'fattening'

d. zga'ran   vs. | zga'ran -ec   zga'ran -ka
    exhaust.PASS.PTCP   exhausted-*er*.M   exhausted-*er*.F
    'exhausted' | 'exhausted person.M'   'exhausted person.F'

Marvin's (2002) account of these facts makes crucial use of phasal spellout. While the L-morpheme is in $\text{Part}^0/\text{T}_2{}^0$, which is not a phasal head, the PASS.PART morpheme is in Pass^0, which is an adjectival head. Since categorial heads trigger spellout, the prosody of the passive participle is computed and shipped off to PF, so it cannot be altered by morphemes that get merged later. L-participles, on the other hand, do not constitute phases and therefore allow morphemes like *-ec* to interfere with the prosody of the whole.

However, while Marvin's account correctly predicts prosodic faithfulness in derivations from passive participles, it does not predict total neutralisation of lexical prosody in derivations from L-participles (including also adjectives in *-n* and *-sk* and others, see Simonović 2020). Rather, what we would expect is that some of the further affixes are stress-affecting, whereas others are stress-neutral and allow for the preservation of lexical prosody.

In order to resolve the problem of obligatory stress-shifting behaviour in derivations from L-participles, Simonović (2020), who also follows Lowenstamm (2014) in assuming that derivational affixes are roots, generalises Marvin's idea of extended roots to all nominalisations that contain the L-morpheme. On this analysis L is a root-selecting root, which appears in a structure that Lowenstamm (2014) terms a "radical core", i.e. a sequence of roots with no intervening categorial heads. In radical cores, default prosody of the language is assigned. Simonović (2020) argues more generally that all cases where affixal prosody overrides lexical verbal stress should be analysed as cases of radical cores. For [me'r-i-l-əts] and [me'r-i-l-ən] from (16), the relevant radical cores would be:

(18) $\sqrt{\text{MER(I)}}+\sqrt{\text{L}}+\sqrt{\text{C}}$ $-$ $\sqrt{\text{MER(I)}}+\sqrt{\text{L}}+\sqrt{\text{N}}$
     'measurer'   'measuring.ADJ'

In both examples the radical cores span over all the morphemes that have phonological content and are embedded under a silent nominaliser and adjectiviser, respectively. Default stress in Slovenian is final, but schwa is avoided by stress, which is why we get [me'r-i-l-əts] and [me'r-i-l-ən] rather than *[mer-i-'l-əts] and

*[mer-i-ˈl-ən] (for a full analysis, see Simonović 2022).[16]

In light of the previous discussion, our approach in this paper departs from Simonović (2020), in that we argue that the structure below ʟ is not itself a root, but minimally a *v*P, since it contains a theme vowel. The question, now, is whether we can still account for the uniform prosodic behavior of all *lec-* and related derivations. The answer is that this uniform prosody is predicted as long as we maintain that ʟ is a transitive root which is required to be selected by a root. The presence of a root selected by a root in the structure will always result in a radical core and impose default prosody whenever the radical core is spelled out. As in Simonović (2020), the lexical prosody of the ʟ-participle is then a consequence of the fact that participles do not contain any radical cores.

And finally, consider the following example as an illustration of the assignment of default prosody. The dual form of *meˈr-i-l-ec*, given in (16b), is *meˈr-i-l-c-a* (we use it because it has an overt case/number ending). This word is spelled out in three cycles. First the *v*P *mer-i-* gets spelled out, then the *n*P -ʟ-*c*- and then the case/number ending -*a* follows.[17] Now, in the first cycle the faithful prosodic pattern wins: ˈmer-i-. This output serves as the input to the second cycle, where there is a clash between the lexical prosody (ˈmerilc) and the pattern imposed by the radical core (meˈrilc). In such cases, the rightmost accent mark wins, so the theme vowel ends up stressed. Finally, the case/number ending -*a* is stress-neutral and does not contain a radical core, so the whole word is realised as *meˈrilca*.

3.3 A summary

In this section, we discussed previous approaches to *lec-* and related nominalisations, while also articulating our own approach. We follow a host of previous formal approaches in severing the ʟ morpheme from -*ec* and all the other morphemes which it gets combined with. We however depart from the previous analyses in that we assume that all *lec-* and related nominalisations contain verbal structure, while at the same time not containing full ʟ-participles. The exact way in which these two verbal structures differ is the main focus of the following section.

---

[16]A question that Simonović (2020) leaves open is the status of theme vowels (e.g., *mer-i̱* in (18)). If the whole structure is a radical core, the theme vowel has to be part of the root, just as in Marvin (2002). Then the problem remains why the same root can appear without the theme vowel, e.g., in the noun *mer-a* 'measure' and in the adjective *mer-en* 'measuring'. Our analysis in §4 addresses this issue.

[17]See fn. 7 for the omission of *e*.

4 An analysis

We pursue a unified analysis of the ʟ morpheme in (i) ʟ-participles, (ii) deverbal nominalisations with ʟ, such as *lec*-nominalisations, and (iii) non-deverbal items such as *krog-l-a* 'sphere' (see §3.2.2). As indicated in §3, we will pursue the idea that ʟ is always a root.

In this section we will tackle the task of explaining the two main issues that emerged throughout the paper. First, all deverbal nominalisations require some extension of the verbal base, be it with ʟ, the passive ɴ/ᴛ or the ᴛ of the short infinitive (see §1). The question, then, is why such extension is required and what the difference between the specific extensions is. Here we limit ourselves to ɴ/ᴛ and ʟ and leave short infinitives for future work. Second, prosodic patterns and the set of allowed theme vowels distinguish between ʟ-participles on the one hand and *lec*- and other related nominalisations on the other. We have argued in the previous section that both of these environments involve a verbal structure. The question, then, is what the exact structure of nominalisations is and what the structure of participles is. Depending on the answer to this question, an account needs to be formulated of the way in which the inventory of theme vowels is restricted to *a* and *i* in *lec*-nominalisations.

4.1 Why is base extension needed and how it works

Given that *lec*-nominalisations are at the centre of this paper and that we have shown the *-ec* in them to constitute a separate, independent suffix, we will limit the discussion to examples with *-ec*. We will also assume that as a derivational affix *-ec* is a transitive root, as argued (for all derivational affixes) in Lowenstamm (2014), summed up in §3.2.2. This means that $\sqrt{\text{EC}}$ can be categorised (and in fact is categorised by an *n* head) and has the ability to select. It it precisely this ability that leads to the modification of the base.

As shown in §1 and §2, nominalisers like *-ec* appear with different complements, but crucially never select for a *v*P (see §1). This is why there are no examples like **predav-a-p* or **uč-i-p*, cf. (6). Since verbal bases are not acceptable complements for *-ec*, merging *-ec* with a *v*P results in a crash and thus requires some extra operation or additional structure. We argue that insertion of the ʟ-root is such an operation, which makes the modified structure parallel to examples such as (9b). On the other hand, *-ec* can merge with a passive (ɴ/ᴛ-) base, since passives are adjectival (and therefore an acceptable complement), cf. (9a).

In proposing root insertion of $\sqrt{\text{L}}$ in *lec*-nominalisations, we essentially extend Acquaviva's (2009) idea of root extension (and generalise Marvin's 2002 idea of

root extension to all *lec*-nominalisations). That is, Acquaviva (2009) argues that items such as *de-stroy* consist of a "lexical" root $\sqrt{\text{STROY}}$ and a root extension $\sqrt{\text{DE}}$, which attaches to $\sqrt{\text{STROY}}$, modifies it, and in doing so creates a complex root which is only then categorised. Since in Lowenstamm's (2014) approach derivational roots are transitive, they are able to take any kind of complement and thus can extend either other roots or phrases.

Our proposal is given in Figure 2. $\sqrt{\text{L}}$ in nominalisations acts as an extension. However, unlike $\sqrt{\text{DE}}$, it projects over the categorised (functional) structure with which it merges, and thus can be selected by items that select for roots, such as *-ec*. Recall from 3.2.4 that this approach then also solves the issue of prosody. Since there is no categorial head above $\sqrt{\text{L}}$ to trigger spell out, the roots $\sqrt{\text{L}}$ and $\sqrt{\text{EC}}$ form a radical core.
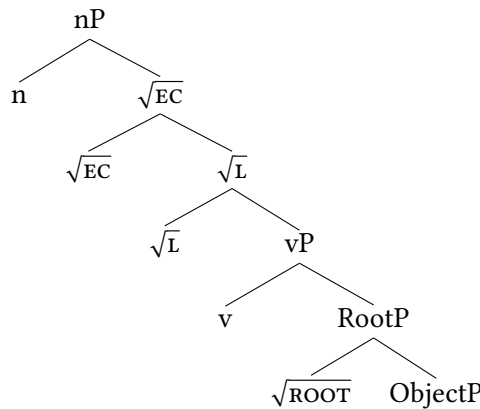
Figure 2: Lec-nominalisations

As we have seen in §2, the morpheme L displays extreme multifunctionality in Slovenian, showing up in non-verbal contexts (*krog-l-a* 'sphere'), as an extension of the domain of deverbal derivation, and as the participial ending. It behaves as the default elsewhere allomorph of the verbal domain (as proposed already in Marvin 2002), which makes it comparable to the morpheme OV, previously described as playing a similar role in the nominal and adjectival domain (Simonović & Mišmaš 2020). In the following subsection, we turn to the nature of L (and its passive counterpart) in participles.

4.2 The analysis of participial roots

We propose that traditional participial morphemes, L and N/T, are conceptually empty roots that are merged (unlike L in *lec*-nominalisations) in the head of AspP.

Our proposal now enables us to take a further step in investigating the nature of derivational roots. That is, Lowenstamm (2014) posits that derivational roots lack semantic content and that their root nature precludes the eventual realisation of syntactic content. The proposal according to which participial morphemes are roots merging in the head of AspP therefore raises two important questions. The first one concerns the possibility of a conceptually empty root to be manifested as either ʟ or ɴ/ᴛ. The second one regards the very possibility of a root being merged in the head of a functional projection.[18]

To address the first problem, we employ allomorphy rules that take the root ʟ to have two phonological realisations (or Vocabulary Items) – one that emerges in specific contexts (i.e., the 'passive' /n/) and the elsewhere form (i.e., /l/).

(19) $\sqrt{\text{ʟ}} \leftrightarrow$ /n/ \   _[ᴘᴀssɪᴠᴇ]
       $\leftrightarrow$ /l/   elsewhere

This then captures Marvin's (2002) observation that ʟ seems to be the elsewhere allomorph of the verbal domain.[19] Still, the two Vocabulary Items are also associated with different interpretations – we return to the issue in what follows.

As for the proposal that the root ʟ is merged in a functional head, we build on the analysis in Cavirani-Pots (2020) (see also Cavirani-Pots et al. 2021), who argues that semi-lexicality of some items emerges when (lexical) roots are merged/incorporated into a functional head. These items have both lexical and functional uses. For instance, the word *bunch* is used in English as both a quantifier (*A bunch of chickens were found on the trail*), and a regular lexical noun (*The flowers were arranged in a beautiful bunch*). In the former, functional use, *bunch* is incorporated into a functional head (Q^0), whereas in its lexical use, it realises a root position (Cavirani-Pots et al. 2021). And while typical instances of semi-lexicality include

---

[18]Another potential issue, peculiar to Slavic languages, concerns a widespread view that secondary-imperfective suffixes are markers of (imperfective) grammatical aspect, and are typically analysed as heads of AspP (e.g. Smith 1997, Ramchand 2004, 2008, Borer 2005, Progovac 2005, Borik 2006, among many others). The compatibility of participial morphemes in Slovenian and BCMS with secondary imperfectives at first glance clashes with our proposal that ʟ and ɴ/ᴛ are merged in Asp^0. However, the problem disappears once we analyse secondary-imperfective suffixes as reverbalisers, i.e. morphemes that combine with perfective verbs, which encode telicity, and return bare *v*Ps (see Arsenijević et al. 2023, Simonović et al. 2021 for detailed argumentation). This means that secondary-imperfective suffixes are merged below grammatical aspect, which is an idea that has also been advocated in Klein (1995), Łazorczyk (2010), Tatevosov (2015, 2017), Mueller-Reichau (2020), Biskup (2023), Milosavljević (2023) – although their exact function varies across approaches.

[19]Given that the passive affixes are referred to as ɴ/ᴛ throughout the paper, we are simplifying the Vocabulary Items and referring the reader to Marvin (2002: 92).

clearly meaningful roots such as *bunch*, in the case of participial morphemes, the underlying procedure is the same; what is merged into a functional head (Asp^0) is a conceptually empty root. Completely parallel to other semi-lexical items, however, L too has lexical and functional uses. When it serves as an extension in deverbal nominalisations it is lexical (except that, here as well, it is devoid of conceptual content), and functional in participles.

The obvious question that emerges is what the motivation is behind merging a conceptually empty root into the Aspectual head. To answer this question, but also to further motivate our analysis of L in Western South Slavic as a functional root, we will extend our discussion to the participial morphology more generally, as it exhibits very similar properties across languages.

Participial morphemes are multifunctional in many languages, i.e., they appear in a variety of contexts. For instance, the same participial forms are found in the verbal/eventive participles, proper adjectives and (present) perfect constructions in languages such as English, German, Italian, Latin (see Borik & Gehrke 2019, Wegner 2019b for overviews). In Romance languages (Italian, Latin), like in Slavic, the "participial" bases are also found in nominalisations (e.g. Calabrese 2020, and references therein). This diversity of contexts is the first property that is common to participial morphemes and roots (both traditional ones and affixes). The multifunctionality of participial morphemes has led many authors to propose that they have either a very light meaning or no meaning whatsoever – which is a property of some affixes as roots in the sense of Creemers et al. (2018) and Simonović & Mišmaš (2020).

There are roughly three families of approaches in the formal literature trying to handle a pure (if any) semantic contribution of participial morphemes. One is to assume that they are exponents of the Asp head, with highly underspecified contexts of application and very abstract semantics (e.g. Embick 2000, 2004, Embick & Halle 2005, Remberger 2012, Wegner 2019a, 2021).[20] An inter-

---

[20]For instance, Embick (2004) postulates different "flavours" of the aspectual head to derive different types of passives in English – eventive, resultative and stative. Notably, Embick also analyses what he calls "stative participles" (but effectively adjectives) like *closed* or *open* as also including a (stative) Asp head that merges directly with the root, i.e. they differ from "true" verbal participles in lacking a verbalising head. Remberger (2012: 286) proposes that participial morphemes in Latin are exponents of the nominal aspect n/Asp that has no specific tense value or temporal semantics, and means something like "concerned/affected". Wegner (2019a,b, 2021) proposes that participial morphemes in English (but possibly also in other languages) are exponents of a single underspecified aspectual head and that specific aspectual values of the given predicate are computed based on its interaction with the telicity properties of the *v*P in the complement of Asp, as well as with the semantics brought about by auxiliaries a particular participle combines with. Wegner's (2019a) approach is reminiscent of more general approaches to grammatical aspect as default aspect such as Bohnemeyer & Swift (2004).

mediate stance is that participial morphemes are meaningless at least in some environments (e.g. in the so-called target and resultant state participles in German or English), as proposed in Kratzer (2000) and Ramchand (2018), but their contribution is not vacuous even in such cases. For Kratzer (2000), the participial morpheme serves only to license the absence of verbal inflection (and consequently the external argument).[21] According to Ramchand (2018), the participial morpheme *-en/-ed* in English is devoid of conceptual content associated with syntactic information. It is a "stunted version of the inflected verbal form" (Ramchand 2018: 127) that can spell out different subparts of the verbal structure up to AspP, i.e. "any non-tense-information-carrying contiguous subset of the root's features" (Ramchand 2018: 81): ResP in the case of the stative passive, InitP for the verbal/eventive passive, and AspP in the present perfect construction. In the spirit of Kratzer (2000), Ramchand (2018: 92) contends that "the effect of participle formation is not vacuous, presumably because it suspends the continuation of the verb to tense inflection and anchoring, and makes adjectivisation possible". Finally, the third, most radical view is that the convergence in "participial" form across different syntactic contexts is a consequence of purely morphological rules rather than a reflex of any common semantic/syntactic core (Calabrese 2020). This last view is based on the assumption that there is a separate morphological module, which can manipulate the output of the narrow syntax. Participial morphemes, alongside theme vowels, constitute a crucial piece of evidence for postulating such a module.[22]

Our approach, on which participial morphemes are conceptually empty roots merging in the Asp head, combines and further elaborates and motivates the first two of the three families of approaches to participial morphemes presented above. We immediately exclude the third option, i.e., the purely morphological analysis, according to which syntax does not play any role, as the roots under discussion clearly have a syntactic role, although they are conceptually empty.[23] Specifically, an analysis in terms of roots explains their multifunctional (and multicategorial) status and their highly abstract/underspecified meaning or lack of meaning. These are, as we have seen, general properties of derivational affixes

---

[21]In Kratzer (2000) the lack of verbal inflection explains why adjectival passives lack an external argument, as in her approach the external argument is introduced by verbal inflection. The stative nature of these participles is brought about either by a zero suffix, or by the adjectivising head itself.

[22]Apart from DM approaches like Calabrese (2020), this stance is at the heart of A-morphous approaches to morphology such as Aronoff (1993), who also analyses participial morphemes, alongside theme vowels, as purely morphological entities.

[23]The reader is also referred to Milosavljević & Arsenijević (2022), Kovačević et al. (2024) for arguments against these being purely ornamental morphemes, much like theme vowels.

reanalysed as roots. It also explains the intuition hinted at in Kratzer (2000) and Ramchand (2018) that they license the absence of anchoring to a specific context or, more generally, the absence of referential properties. This is in full accordance with the view that the meaning of roots is intensional, and that the referential properties are introduced by functional material (see Arsenijević 2022 and the references therein), probably by the inflectional/person morphemes (cf. Ramchand 2018). Additionally, given that adjectives, like roots, do not refer, this aligns with the observed selectional restrictions of the nominalising suffixes discussed in the present paper. That is, suffixes like *-ec* select for either adjectives or root structures where both are basically structures devoid of referential/extensional meanings.

Our analysis also explains the intuition that participles generally have a "stative" or "adjectival" meaning (Kratzer 2000, Ramchand 2018, Tatevosov 2017, Borik & Gehrke 2019). Let us spell this out in more technical terms. The aspectual head normally specifies a temporal relation, which obtains between the eventuality described by its complement (i.e., *v*P) and a temporal pronoun in its specifier (referring to the topical time, i.e., the reference time). When this head is filled with a conceptually empty root (such as $\sqrt{\text{L}}$), the temporal pronoun cannot be merged in the specifier – exactly due to the intensional nature of the root, which licenses the absence of referential properties. Instead, the highest c-commanded argument moves to the specifier of AspP (the external argument in transitives and unergatives, the internal one in unaccusatives), deriving the interpretation that the predicate denoted by the *v*P overlaps with the temporal dimension of this argument. The overlap interpretation is default for viewpoint aspect in the absence of an overt specification of non-overlap. This derives exactly the result state interpretation as in Kratzer (2000): the relevant argument bears the property of having participated, with a particular role, depending on the value of Voice, in events satisfying the description specified by the complement of Asp. This matches exactly the interpretation that verbal participles have.

4.3 The syntax of *lec*-nominalisations and theme vowel realisation

As discussed in §3.2.2, L-nominalisations and L-participles allow different sets of theme vowels. Specifically, L-nominalisations allow only two theme vowels (*i* and *a*), while L-participles allow all theme vowels available in the language. This pattern is crucial evidence that the root L combines with different sizes of the verbal base in nominalisations and participles, i.e. that L-nominalisations do not actually contain L-participles. Specifically, the structure in L-nominalisations is smaller than that in L-participles. As the most conservative implementation, we

assume that ʟ-nominalisations only contain the *v*P (as they always have a theme vowel). Note that this does not hold for ɴ/ᴛ-nominalisations: they are always deadjectival, i.e. contain the passive participle, as also attested in their prosodic behaviour (see §3.2.4). As is fully expected, ɴ/ᴛ-nominalisations display the same set of theme vowels as passive participles. This then also means that the ɴ/ᴛ morpheme never functions as a true *v*P extension, but always as a participial ending.

Our final piece of the puzzle is the exact mechanism behind the variable theme vowel exponence. Recall that in the picture we sketched above (§3.2.2), some verbs have three theme vowel exponents. A verb like *vesti* 'embroider' is a case in point: its theme vowel is realised as ∅ in the ʟ-participle *vez-∅-l-a*, as *e* in the present-tense form *vez-e-mo* 'we embroider' and as *i* in the nominalisation *vez-i-l-ec* 'embroiderer'.

Our general approach is to follow Oltra Massuet (1999) in the assumption that different theme-vowel classes result from root diacritics, such as $[\alpha]$ and $[\beta]$, or lack thereof. The present-tense version of the theme vowel is not in focus here. Suffice it to say that the spellout of the theme vowel is influenced by additional marked features on *v* or an adjacent head (see Oltra Massuet's discussion of the "marked T"). We are then left with the difference between participles and nominalisations/adjectivisations discussed in this paper. We can first define the most general vocabulary item for West South Slavic, which will apply in all cases where no more specific vocabulary item applies, as given in (20).

(20)   TV ↔ /i/

This defines *i* as the elsewhere theme vowel in the whole system. The next is the theme vowel *a*, which is defined by Marvin (2002) as the other default theme vowel. It is the spellout of all theme vowels that have the unmarked negative value of the diacritic feature $[\alpha]$.

(21) $\text{TV}_{[-\alpha]}$ ↔ /a/

Now we arrive at the vocabulary items for the two classes which have a shift to the elsewhere theme vowel in nominalisations.

(22) $\text{TV}_{[+\alpha, \, -\beta]}$ ↔ /e/ \   _ [ᴠᴏɪᴄᴇ]
        $\text{TV}_{[+\alpha, \, +\beta]}$ ↔ /∅/ \   _ [ᴠᴏɪᴄᴇ]
        ...

These theme vowels have vocabulary items which, apart from the diacritic features, also refer to the presence of an adjacent Voice projection. These vocabulary

items are applicable in participles. However, in non-passive nominalisations and adjectivisations, the more general and least specific vocabulary item applies, and *i* gets inserted.

5 Conclusions and further developments

The paper focused on nominalisations seemingly derived from ʟ-participles, exemplified by the *lec*-nominalisations in Slovenian, in order to determine the nature and position of the ʟ-morpheme. One important point of comparison was the passive participles and the ɴ-morpheme in nominalisations that are derived from passive participles.

We argue that the supposed ʟ-participle nominalisations are not derived from participles in that there is no perfect containment relation between the ʟ-participle and the *lec/lac*- nominalisations. Rather, *lec/lac*-nominalisations contain a smaller structure (*v*P) than the corresponding ʟ-participle. This influences the spell out, and consequently, the insertion of theme vowels. In the proposed structure, ʟ (but also ɴ/ᴛ in passive participles and related nominalisations) are realisations of a conceptually empty root. The structure of nominalisations, however, does include the verbaliser *v* realised by a theme vowel, which is, in the contexts without Voice, realised as *a* or *i*, the latter being the most general Vocabulary Item.

It was argued that the ʟ-morpheme is a conceptually empty root that can appear both as an inflectional ending and as a derivational affix. This multifunctionality of roots is an important innovation of our analysis, but, obviously, also the least explored one. While there have been previous analyses of related phenomena for both Slovenian and BCMS (Simonović & Arsenijević 2014, 2020, Simonović & Mišmaš 2020), we hope that future research will bring new insight as well as an integral theory of the phenomenon.

Abbreviations

1	first person	ɪɴꜰ	infinitive
ADJ	adjective	ɪᴘꜰᴠ	imperfective
BCMS	Bosnian/Croatian/	ᴍ	masculine
	Montenegrin/Serbian	ɴ	neuter
AUX	auxiliary	ɴᴏᴍ	nominative
ꜰ	feminine	ᴘᴀss	passive

PFV	perfective	Slo	Slovenian
PL	plural	SI	secondary imperfective
PRS	present tense	S.INF	short infinitive
PTCP	participle	TV	thematic vowel
SG	singular		

Acknowledgments

References

Acquaviva, Paolo. 2009. Roots and lexicality in distributed morphology. In Alexandra Galani, Daniel Redinger & Norman Yeo (eds.), *York-Essex morphology meeting*, vol. 5, 1–21. York: University of York. Department of Language & Linguistic Science. http://hdl.handle.net/10197/4148.

Aljović, Nadira. 2000. Unaccusativity and aspect in SerBoCroatian. In Christine Czinglar, Katharina Köhler, Erica Thrift, Erik Jan van der Torre & Malte Zimmermann (eds.), *Console VIII proceedings: Proceedings of the Eighth Conference of the Student Organisation of Linguistics in Europe*, 1–15. Leiden: SOLE. https://www.researchgate.net/publication/318281300_Unaccusativity_and_aspect_in_SerBoCroatian.

Aronoff, Mark. 1993. *Morphology by itself: Stems and inflectional classes*. Cambridge: MIT Press.

Arsenijević, Boban. 2022. Adjectives as a lexical category: A story of striving for extension. In Phoevos Panagiotidis & Moreno Mitrović (eds.), *A0 – the lexical status of adjectives*, 121–152. Amsterdam: John Benjamins. DOI: 10.1075/lfab.17.04ars.

Arsenijević, Boban, Katarina Gomboc Čeh, Franc Lanko Marušič, Stefan Milosavljević, Petra Mišmaš, Jelena Simić, Marko Simonović & Rok Žaucer. 2024. *Database of the Western South Slavic Verb HyperVerb 2.0 – WeSoSlaV*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1846.

Arsenijević, Boban, Stefan Milosavljević & Marko Simonović. 2023. *Secondary imperfectivisation is reverbalisation is addition of theme vowels*. Presented at the 56th Annual Meeting of the Societas Linguistica Europaea (SLE 56), Workshop Inflection and derivation in the Slavic verb, National and Kapodistrian University of Athens, August 29–September 1, 2023. https://www.academia.edu/108338445/Secondary_imperfectivisation_is_reverbalisation_is_addition_of_theme_vowels.

Babić, Stjepan. 2002. *Tvorba riječi u hrvatskom književnom jeziku* (Velika hrvatska gramatika). Zagreb: Hrvatska akademija znanosti i umjetnosti, Nakladni zavod Globus.

Bajec, Anton. 1950. *Besedotvorje slovenskega jezika*, vol. 1. Ljubljana: Slovenska akademija znanosti in umetnosti.

Bajec, Anton, Rudolf Kolarič & Mirko Rupel. 1956. *Slovenska slovnica*. Ljubljana: Državna založba Slovenije.

Birtić, Matea. 2008. *Unutarnja struktura odglagolskih imenica u hrvatskome jeziku*. Zagreb: Institut za hrvatski jezik i jezikoslovlje.

Biskup, Petr. 2019. *Prepositions, case and verbal prefixes: The case of Slavic*. Amsterdam: John Benjamins.

Biskup, Petr. 2023. Aspect separated from aspectual markers in Russian and Czech. In Petr Biskup, Marcel Börner, Olav Mueller-Reichau & Iuliia Shcherbina (eds.), *Advances in formal Slavic linguistics 2021*, 61–98. Berlin: Language Science Press. DOI: 10.5281/zenodo.10123631.

Bohnemeyer, Jürgen & Mary Swift. 2004. Event realization and default aspect. *Linguistics and Philosophy* 27(3). 263–296. DOI: 10.1023/b:ling.0000023371.15460.43.

Borer, Hagit. 2005. *Structuring sense, vol. 2: The normal course of events*. Oxford: Oxford University Press.

Borik, Olga. 2006. *Aspect and reference time*. Oxford: Oxford University Press.

Borik, Olga & Berit Gehrke. 2019. Participles: Form, use and meaning. *Glossa: a journal of general linguistics* 4(1). DOI: 10.5334/GJGL.1055.

Caha, Pavel & Markéta Ziková. 2022. Prefixes in Czech zero-derived nominalizations and verbs. In Petr Karlík & Lucie Taraldsen Medová (eds.), *Nominalizations and participles in Czech and beyond*, 3–61. München: Lincom.

Calabrese, Andrea. 2020. Remarks on the role of the perfect participle in Italian morphology and on its history. *Probus* 32(2). 209–269. DOI: 10.1515/probus-2020-0006.

Cavirani-Pots, Cora. 2020. *Roots in progress: Semi-lexicality in the Dutch and Afrikaans verbal domain*. Utrecht: LOT.

Cavirani-Pots, Cora, Marijke De Belder & Heidi Klockmann. 2021. *Semi-lexicality: Syntax or lexicon?* Presented at GLOW 43, HU Berlin (online), April 15, 2021. https://www.crissp.be/wp-content/uploads/2021/04/GLOW_Cavirani-Pots_De-Belder_Klockmann_FINAL.pdf.

Creemers, Ava, Jan Don & Paula Fenger. 2018. Some affixes are roots, others are heads. *Natural Language & Linguistic Theory* 36(1). 45–84. DOI: 10.1007/s11049-017-9372-1.

Embick, David. 2000. Features, syntax, and categories in the Latin perfect. *Linguistic Inquiry* 31(2). 185–230. DOI: 10.1162/002438900554343.

Embick, David. 2004. On the structure of resultative participles in English. *Linguistic Inquiry* 35(3). 355–392. DOI: 10.1162/0024389041402634.

Embick, David & Morris Halle. 2005. On the status of stems in morphological theory. In Twan Geerts, Ivo van Ginneken & Haike Jacobs (eds.), *Romance languages and linguistic theory 2003*, 59–88. Amsterdam: John Benjamins. DOI: 10.1075/cilt.270.03emb.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale & S. Jay Keyser (eds.), *The view from Building 20*, 111–176. Cambridge: MIT Press.

Halle, Morris & Alec Marantz. 1994. Some key features of distributed morphology. In Andrew Carnie & Heidi Harley (eds.), *MITWPL 21: Papers on phonology and morphology*, 275–288. Cambridge: MITWPL.

Klajn, Ivan. 2003. *Tvorba reči u savremenom srpskom jeziku. Drugi deo. Sufiksacija i konverzija* (Prilozi gramatici srpskoga jezika). Beograd: Zavod za udžbenike i nastavna sredstva, Institut za srpski jezik SANU, Matica srpska.

Klein, Wolfgang. 1995. A time-relational analysis of Russian aspect. *Language* 71(4). 669–695. DOI: 10.2307/415740.

Kovačević, Predrag, Stefan Milosavljević & Marko Simonović. 2024. Theme-vowel minimal pairs show argument structure alternations. *Journal of Linguistics.* 1–30. DOI: 10.1017/S0022226723000415.

Kratzer, Angelika. 2000. Building statives. In Lisa J. Conathan, Jeff Good, Darya Kavitskaya, Alyssa B. Wulf & Alan C. L. Yu (eds.), *Proceedings of the Berkeley Linguistic Society 26*, 385–399. Berkeley: University of California, Berkeley Linguistics Society. DOI: https://doi.org/10.3765/bls.v26i1.1131.

Łazorczyk, Agnieszka A. 2010. *Decomposing Slavic aspect: The role of aspectual morphology in Polish and other Slavic languages.* University of Southern California. (Doctoral dissertation).

Lowenstamm, Jean. 2014. Derivational affixes as roots: Phasal spell-out meets English stress shift. In Artemis Alexiadou, Hagit Borer & Florian Schäfer (eds.),

The syntax of roots and the roots of syntax, 230–258. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199665266.003.0011.

Marantz, Alec. 2001. *Words and things.* Ms., MIT. https://babel.ucsc.edu/~hank/mrg.readings/Marantz_words.pdf.

Maretić, Tomislav. 1963. *Gramatika hrvatskoga ili srpskoga književnog jezika.* Zagreb: Matica hrvatska.

Marjanovič, Katarina, Christina Manouilidou & Tatjana Marvin. 2013. Word-formation rules in Slovenian agentive deverbal nominalization. *Slovene Linguistic Studies* 9. 93–109. DOI: 10.17161/SLS.1808.11432.

Marušič, Franc Lanko, Rok Žaucer, Petra Mišmaš, Boban Arsenijević, Marko Simonović, Stefan Milosavljević, Katarina Gomboc Čeh & Jelena Simić. 2022. *Database of the Western South Slavic Verb WeSoSlav Inflection.* Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1683.

Marvin, Tatjana. 2002. *Topics in the stress and syntax of words.* Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Marvin, Tatjana. 2015. Deverbal external argument nominalizations in Slovenian. *Grazer Linguistische Studien* 83. 155–160. https://unipub.uni-graz.at/gls/content/titleinfo/1283377.

Marvin, Tatjana. 2016. On agentive and instrumental deverbal nominalizations in Slovenian. *Jezikoslovlje* 17(1-2). 321–337. https://hrcak.srce.hr/167809.

Milosavljević, Stefan. 2023. *Specification of event duration and aspectual composition in Slavic.* Graz: University of Graz. (Doctoral dissertation).

Milosavljević, Stefan & Boban Arsenijević. 2022. What differentiates Serbo-Croatian verbal theme vowels: Content or markedness? *Glossa: a journal of general linguistics* 7(1). DOI: 10.16995/glossa.8535.

Mueller-Reichau, Olav. 2020. Default aspect based on state change. *Rhema* (1). 90–105. DOI: 10.31862/2500-2953-2020-1-90-105.

Oltra Massuet, Maria Isabel. 1999. *On the notion of theme vowel: A new approach to Catalan verbal morphology.* Cambridge: Massachusetts Institute of Technology. (MA thesis).

Progovac, Ljiljana. 2005. *A syntax of Serbian: Clausal architecture.* Bloomington: Slavica Publishers.

Quaglia, Stefano, Marko Simonović, Svitlana Antonyuk Yudina & Boban Arsenijević. 2022. Allomorphy, morphological operations and the order of Slavic verb-prefixes. *Journal of Slavic Linguistics* 30(FASL 29 issue). 1–15. DOI: 10.1353/jsl.2022.a923080.

Ramchand, Gillian Catriona. 2004. Time and the event: The semantics of Russian prefixes. *Nordlyd* 32(2). 323–361. DOI: 10.7557/12.72.

Ramchand, Gillian Catriona. 2008. Perfectivity as aspectual definiteness: Time and the event in Russian. *Lingua* 118(11). 1690–1715. DOI: 10.1016/j.lingua.2007.03.008.

Ramchand, Gillian Catriona. 2018. *Situations and syntactic structures: Rethinking auxiliaries and order in English.* Cambridge: MIT Press.

Remberger, Eva-Maria. 2012. Participles and nominal aspect. In Sascha Gaglia & Marc-Olivier Hinzelin (eds.), *Inflection and word formation in Romance languages*, 271–294. Amsterdam: John Benjamins. DOI: 10.1075/la.186.10rem.

Simonović, Marko. 2020. Categories, root complexes and default stress: Slovenian nominalizations revisited. *Linguistica* 60(1). 103–117. DOI: 10.4312/linguistica.60.1.103-117.

Simonović, Marko. 2022. Derivational affixes as roots across categories. *Journal of Slavic Linguistics* 30(2). 195–233. DOI: 10.1353/jsl.2022.a909904.

Simonović, Marko & Boban Arsenijević. 2014. Regular and honorary membership: On two kinds of deverbal nouns in Serbo-Croatian. *Lingue e Linguaggio* 13(2). 185–210. DOI: 10.1418/78407.

Simonović, Marko & Boban Arsenijević. 2020. Syntax predicts prosody: Multipurpose morphemes in Serbo-Croatian. In Franc Marušič, Petra Mišmaš & Rok Žaucer (eds.), *Advances in formal Slavic linguistics 2017*, 277–304. Berlin: Language Science Press. DOI: 10.5281/zenodo.3764865.

Simonović, Marko, Stefan Milosavljević & Boban Arsenijević. 2021. Serbo-Croatian secondary imperfectivisers consist of theme vowels. *Journal of Slavic linguistics* 31(FASL 30 issue). 1–27. DOI: https://ojs.ung.si/index.php/JSL/article/view/178..

Simonović, Marko & Petra Mišmaš. 2020. √ov is in the air: The extreme multifunctionality of the Slovenian affix ov. *Linguistica* 60(1). 83–102. DOI: 10.4312/linguistica.60.1.83-102.

Simonović, Marko & Petra Mišmaš. 2022. Lowest theme vowels or highest roots? An 'unaccusative' theme-vowel class in Slovenian. *Glossa: a journal of general linguistics* 7(1). 1–31. DOI: 10.16995/glossa.5809.

Smith, Carlota S. 1997. *The parameter of aspect.* Dordrecht: Kluwer.

Stramljič Breznik, Irena. 1995. Specializiranost obrazil za izpeljanke s pomenom vršilca dejanja, nosilca lastnosti ali stanja in opravkarja. *Jezik in slovstvo* 40(8). 285–291. http://www.dlib.si/?URN=URN:NBN:SI:DOC-MHEB8X81.

Stramljič Breznik, Irena. 1999. *Prispevki iz slovenskega besedoslovja* (Zora). Maribor: Slavistično društvo Maribor.

Svenonius, Peter. 2004. Slavic prefixes and morphology: An introduction to the Nordlyd volume. *Nordlyd* 32(2). 177–204. DOI: 10.7557/12.67.

Tatevosov, Sergei. 2015. Severing imperfectivity from the verb. In Gerhild Zybatow, Petr Biskup, Marcel Guhl, Claudia Hurtig, Olav Mueller-Reichau & Maria Yastrebova (eds.), *Slavic grammar from a formal perspective: The 10th anniversary FDSL conference*, 465–494. Frankfurt am Main: Peter Lang.

Tatevosov, Sergei. 2017. *On the aspectual architecture of Russian*. Ms.HSE University. http://otipl.philol.msu.ru/staff/people/tatevosov/aspectless_verb_3.1.pdf.

Todorović, Neda. 2016. *On the presence/absence of TP: Syntactic properties and temporal interpretation*. Storrs: University of Connecticut. (Doctoral dissertation).

Toporišič, Jože. 2000. *Slovenska slovnica [Slovenian Grammar]*. 4th edition. Maribor: Obzorja.

Wegner, Dennis. 2019a. The properties of perfect(ive) and (eventive) passive participles: An identity approach. *Glossa: a journal of general linguistics* 4(1). 1–33. DOI: 10.5334/GJGL.751.

Wegner, Dennis. 2019b. *The underspecification of past participles: On the identity of passive and perfect(ive) participles*. Berlin: De Gruyter.

Wegner, Dennis. 2021. The categorial, argument structural and aspectual indeterminacy of past participles: A holistic approach. *Zeitschrift fur Sprachwissenschaft* 40(2). 199–249. DOI: 10.1515/zfs-2021-2027.

Chapter 20

Focus-sensitive particles in Bulgarian: Towards an adverbial-only analysis

Carla Spellerberg

University of Massachusetts, Amherst

This paper is an investigation of the placement and potential adjunction sites of focus-sensitive particles (FSPs) in Bulgarian. In contrast to well-researched languages such as English and German, there is currently no full analysis of FSP-placement in Bulgarian. I propose an analysis here based partly on results of previous analyses by Büring & Hartmann (2001) for German and Zanon (2023) for Russian, arguing that Bulgarian FSPs adjoin to projections belonging to the extended verbal projection (EVP) as well as a functional projection FP in the nominal domain. In addition, I discuss the implications that right-adjunction of FSPs to F-marked constituents in Bulgarian has for the Particle Theory as proposed here. Future research in this direction could focus on the connection between adjacency of the FSP and overt focus movement as well as semantic restrictions that individual modifiers and particles impose upon the possibility of adjunction of FSPs.

1 Introduction

Focus-sensitive particles (FSPs), particles such as English *only*, *even*, and *also*, have received attention to varying degrees depending on the language studied. While association with focus in English is by now a well-studied phenomenon, especially from a semantic perspective (see Rooth 1985, for example), and the syntactic properties of FSPs in Germanic languages such as English and German have been extensively researched (Jacobs 1983, Büring & Hartmann 2001, von Stechow 2008, Mursell 2021), the same cannot be said of many other language families. In Slavic, for example, information structure-sensitive particles, including FSPs, remain understudied in comparison to other phenomena in information

structure (IS), such as the interaction of IS and free word order (Jasinskaja 2016: 731). Additionally, there is generally a strong bias towards Russian data, with other Slavic languages being either less studied or even understudied in comparison to that, as Jasinskaja (2016) notes.

In the following, I provide a first analysis of Bulgarian focus-sensitive particles with an emphasis on *samo* 'only', *săšto* 'also', and *dori* 'even'. (1) shows how these three particles associate with a f(ocus)-marked constituent in Bulgarian.[1,2]

(1) a. Obadi-h   se   samo [na IVAN]$_F$.
         call-PST.1SG REFL only   to Ivan
         'I only called Ivan.'   (Tisheva & Dzhonova 2003: ex. 8a)

     b. Včera   săšto [četo-h   ROMAN]$_F$.
         yesterday also   read-PST.1SG novel
         'I also read a novel yesterday.'

     c. Včera   dori [AZ]$_F$ pročeto-h   edin roman.
         yesterday even   I   read-PST.1SG one novel
         'Even I read a novel yesterday.'

In addition to pre-focal association with focus, Bulgarian FSPs are also able to follow the focused constituent they are associated with, as (2) demonstrates. Post-focal association is attested for other Slavic languages such as Russian as well and distinguishes Slavic in this respect from German and other Germanic languages.[3]

(2) a. Obadi-h   se   [na IVAN]$_F$ samo.
         call-PST.1SG REFL to Ivan   only
         'I only called Ivan.'   (Tisheva & Dzhonova 2003: ex. 8a)

     b. [PETĂR]$_F$ săšto mož-e   da gotvi.
          Petăr   also able-PRS.3SG to cook
         'Petăr is able to cook as well.'

---

[1]All non-English examples in this paper are from Bulgarian, unless marked otherwise next to the example.

[2]In the basic cases shown in this paper, the three particles generally behave the same way with respect to their syntactic behavior. However, once they are investigated in further detail, their behavior (unsurprisingly) diverges. I cannot provide a detailed investigation of this in this paper as the purpose is to provide a first analysis of Bulgarian FSPs, and this will be addressed in future research. Throughout the paper, I indicate relevant differences between the particles with respect to their placement when needed.

[3]While the differences in interpretation between the two word orders shown here deserve an investigation of their own, I provide some preliminary results concerning these differences in Section 4 of this paper.

    c.   [ANA]$_F$ dori šte   trjabva da dojd-e.
        Ana   even AUX.FUT have   to come-PRS.3SG
        'Even Ana will have to come.'

So far, only few studies have focused on the syntax and semantics of Bulgarian FSPs. For instance, a semantic study of *samo* 'only' is given in Nicolova (2000). The semantic properties of the FSP *aż/čak*, the scalar opposite of scalar 'only' which is present in several Slavic languages, including Bulgarian, is extensively studied in Tomaszewicz (2013). The syntactic distribution of *samo* 'only' is described in Tisheva & Dzhonova (2003). Their corpus study provides a detailed description of the adjunction sites of *samo*. However, the study only considers surface word order and does not contain an analysis that goes beyond surface level. In the present study, I close a research gap in this respect and argue that a so-called ADVERBIAL-ONLY ANALYSIS correctly predicts the possible adjunction sites of Bulgarian FSPs along the lines of what is argued for German in Büring & Hartmann (2001) and for Russian in Zanon (2023).

Adverbial-only analyses of FSPs predict that FSPs are only able to adjoin to projections belonging to the Extended Verbal Projection (EVP), although individual analyses of this kind may differ with respect to the projections which they allow adjunction to. The logical alternative to this type of analysis is the so-called ADNOMINAL or MIXED analysis, which predicts that FSPs can adjoin to any type of phrase, and especially also to argument DPs. For German, it has been argued extensively in Büring & Hartmann (2001) that an adverbial-only analysis successfully captures the syntactic properties of the language's FSPs (see Mursell 2021 for an extension and discussion of this proposal, and Sudhoff 2010 for criticism of this line of analysis).

In addition to arguing for an adverbial-only analysis of Bulgarian FSPs, I show that a combination of Büring & Hartmann's (2001) Particle Theory on the one hand, and Zanon's (2023) Particle Theory on the other hand can be adapted to account for the placement options of FSPs in Bulgarian. Since Bulgarian is considered to be a language with relatively "free'' word order, the successful extension of Büring and Hartmann's adverbial-only analysis is relevant insofar as it shows that languages with a more flexible word order than German can also impose heavy restrictions on the distribution of FSPs. This indicates that 'free' word order does not have to mean that FSP-adjunction is necessarily free as well.

This article is structured as follows. In Section 2, I briefly present the few studies that have already been conducted on Bulgarian FSPs, give an overview of the placement options of the three particles studied here on the clausal level and in the nominal domain, and summarize the most important aspects of Büring &

Hartmann's (2001) and Zanon's (2023) Particle Theories. In Section 3, I present syntactic arguments for an adverbial-only analysis of FSP adjunction in Bulgarian and discuss challenges to conventional adverbial-only analyses caused by the particle placement options available in the nominal domain, before moving on to presenting a Particle Theory for Bulgarian. Since Bulgarian FSPs can also appear post-focally, in contrast to German FSPs, I discuss post-focal FSPs in Section 4. Section 5 concludes the paper.

2 Focus-sensitive particles in Bulgarian and beyond

The aim of this section is twofold. First, I offer a basic description of the placement options of Bulgarian FSPs based on previous work on Bulgarian as well as novel data. This is followed by an overview of the basics of Büring & Hartmann's (2001) Particle Theory as well as an analysis of the syntactic behavior of Russian *tol'ko* 'only' in Zanon (2023), both of which I am going to extend to Bulgarian in Section 3.

2.1 Previous research on (Bulgarian) FSPs

2.1.1 Syntactic aspects

In this section, I provide a basic description of the placement of FSPs in Bulgarian based on Tisheva & Dzhonova (2003), a descriptive corpus study of Bulgarian *samo* 'only', and also provide novel data. *Only* and its approximate equivalents are the particles that have received most attention in the literature on FSPs. This is also the case for the corpus study by Tisheva & Dzhonova (2003) already mentioned. The authors argue that *samo* 'only' in Bulgarian "can have scope over NP, PP, AdvP, VP, or part of XP" (Tisheva & Dzhonova 2003: 65).[4] Their data shows how flexible *samo* seems to be when it comes to the potential adjunction sites of the particle. Nicolova (2000: 109) also acknowledges the flexibility of the placement of FSPs in Bulgarian, remarking that NPs, PPs, VP, verbs, AdvPs, or whole subordinate clauses can associate with an FSP. (3) demonstrates that the FSP can adjoin to NPs and Vs (in addition to the apparent PP adjunction demonstrated in (1)).

---

[4]As pointed out by a reviewer, this usage of the term "scope" is potentially misleading as what the authors describe in their paper is actually possible adjunction sites of *samo* 'only' in Bulgarian. I therefore follow the reviewer's suggestion and avoid this usage of the term "scope". While I cannot discuss these matters further here due to reasons of space, more in-depth discussion of different usages of the notion of scope can be found in Branan & Erlewine (2023).

(3) a. Kupi-h   samo [KRASTAVIC-I]$_F$ za salata-ta.
        buy-PST.1SG only   cucumber-PL   for salad-DEF
        'I only bought cucumbers for the salad.'

    b. Samo [ČET-A]$_F$ roman-i.
        only   read-PRS.1SG novel-PL
        'I only read novels.'   (Tisheva & Dzhonova 2003: exx. 7a, 6a)

These examples also show that *samo* generally marks narrow focus (Tisheva & Dzhonova 2003). The authors also note that *samo* can associate with the phrase preceding it when it is placed at the end of the clause, as in (2) above. (2) also shows that the FSPs under discussion here can also follow a subject that they associate with, an option that is not discussed by Tisheva & Dzhonova (2003). In other positions, *samo* is much more likely to associate with the phrase following it (Tisheva & Dzhonova 2003: 6–7). In general, there are only two cases in which the focused constituent is not right-adjacent to the FSP associated with it. Apart from the case of (apparent) right-adjunction already shown, it is also possible for the focused constituent to move to a position at the left edge of the clause, stranding the FSP that is associated with it. This is shown in (4).

(4) a. Čet-a   samo [ROMAN-I]$_F$.
        read-1SG only   novel-PL
        'I read only [novels]$_F$.'

    b. [ROMAN-I]$_F$ čet-a   samo
         novel-PL   read-1SG only
        '[Novels]$_F$, I read only.'

Tisheva & Dzhonova (2003) note that there are two restrictions that the placement of *samo* must adhere to: the prohibition against insertion into PPs and the prohibition against insertion into complex verbal complexes. Examples of both can be seen in (5).

(5) a. * Obadi-h   se   na samo [IVAN]$_F$.
          call-PST.1SG REFL to only   Ivan
          Intended: 'I called only Ivan.'

    b. * Ti   šte   samo [SEDI-Š]$_F$.
          you AUX.FUT only   sit-PRS.2SG
          Intended: 'You will just sit.' (Tisheva & Dzhonova 2003: exx. 8d, 11b)

Apart from these prohibitions, the authors argue that *samo* can be placed relatively freely within the Bulgarian clause. However, a crucial restriction of the

scope of their investigation is that their study is mostly descriptive and only takes surface word order into consideration. Once the aim is to identify why it should be the case that the restrictions in place in Bulgarian exist and what they reveal about the underlying adjunction sites of Bulgarian FSPs, it becomes evident that the potential adjunction sites for FSPs in Bulgarian are much more restricted than can be seen at the level of surface word order.

In addition to the data discussed by Tisheva & Dzhonova (2003), several more potential adjunction sites of Bulgarian FSPs can be discovered in the nominal domain. (6) demonstrates that FSPs such as *samo* can adjoin to PPs within NPs:

(6) a. samo kotka-ta [na SĂSEDKA-TA]$_F$
         only cat-DEF   of neighbor-DEF
         'only the cat [of the neighbor]$_F$'

     b. kotka-ta samo [na SĂSEDKA-TA]$_F$
         cat-DEF only   of neighbor-DEF
         'only the cat [of the neighbor]$_F$'

     c. kotka-ta [na SĂSEDKA-TA]$_F$ samo
         cat-DEF   of neighbor-DEF   only
         'only the cat [of the neighbor]$_F$'

Adjunction to nominal modifiers within NPs sometimes even circumvents the "no PP-insertion"-prohibition discussed by Tisheva & Dzhonova (2003), irrespective of the FSP that is being inserted. However, the acceptability of these examples depends on the preposition and modifier involved, as (7) shows (further discussion can be found in Section 3.2 of this paper).

(7) a. * săs   samo [EDNA]$_F$ kola
            with only   one   car
            Intended: 'with only one car'

     b. $^?$ sled   samo [NJAKOLKO]$_F$ sekund-i
            within only   few   second-PL
            'within only a few seconds'

     c. meždu   samo [DVE]$_F$ opci-i
         between only   two   option-PL
         'between only two options'

     d. meždu   dori [DVE]$_F$ opci-i
         between even two   option-PL
         'between even two options'

After briefly discussing semantic research on Bulgarian FSPs, I turn to theoretical approaches that can be adapted to analyze particle placement in Bulgarian in the remainder of the section. (both are more about syntax though)

2.1.2 Semantic aspects

With respect to the semantics of Bulgarian FSPs, three particles have been studied in the literature in more depth, namely *samo* 'only', *dori* 'even', and *čak*.[5]

According to Nicolova (2000), both the exclusive particle *samo* and the additive particle *dori* mark contrastive focus in Bulgarian.[6,7] The two particles can associate with different types of phrases such as NPs, PPs, VPs, or AdvPs (Nicolova 2000: 109). In addition to that, Nicolova (2000) notes that the contrastively focused constituent can be placed everywhere in the clause and is not restricted to a designated position while the most prominent sentence accent is placed in the domain of the FSP and its adjacent focused constituent.

Tomaszewicz (2013) provides a semantic study of the Slavic FSP *aż/čak*.[8] According to the author, *aż/čak* makes three basic contributions, namely the assertion that lower alternatives than the one presented in the clause are excluded as well as the presuppositions that "the prejacent is high on the scale" and that "the prejacent or an alternative at most as strong is true" (Tomaszewicz 2013: 321). A Bulgarian example is shown in (8) (from Tomaszewicz 2013: 302).[9]

(8) Govori-h   čak s   [MARY]$_F$.
     talk-PST.1SG ČAK with   Mary
     'I talked to somebody so important as [Mary]$_F$.'

In (8), *čak* makes a contribution similar to English *even* in that it singles out Mary as a very important person to talk to. However, the particle is not merely

---

[5]No translation of *čak* is provided here due to its intriguing semantic properties that impede a direct translation into English.

[6]Nicolova (2000: 108) argues that additive as well as exclusive FSPs induce contrastive focus as both particles express a difference between a predicted and a real sum, a position that can be (and has been) debated.

[7]Nicolova (2000) labels *even*-type as well as *also*-type particles as additive particles, a (terminological) decision that does not seem intuitive to readers nowadays, as a reviewer notes. In Nicolova's system, both particles are additive, but *even* is scalar while *also* is non-scalar (which also holds for *only* in her classification). I follow Nicolova's terminology here for the sake of correctly presenting her proposal.

[8]*Aż* is found in Czech, Polish, Slovak, and Russian (with different spellings), and *čak* is found in Bulgarian and other South Slavic languages. Both particles have similar properties and can be treated as two forms of the same particle.

[9]Transliteration changed to scientific transliteration.

presuppositional and can also be the direct opposite of *only*, in contrast to *even* (Tomaszewicz 2013). In contrast to *aż*/*čak*, *only* would assert that there is no higher, true alternative and presuppose that "the prejacent is low on the scale" and that "the prejacent or an alternative at least as strong is true" (Tomaszewicz 2013: 321). Particles such as *čak* are heavily restricted in their usage due to their particular semantic properties. While *only*, *also*, and *even* can be used interchangeably in most examples discussed here, this is not the case for *čak*. In the remainder of this paper, I am going to focus on the less semantically restricted FSPs in Bulgarian. In the next section, I turn to Büring & Hartmann's (2001) theory of FSPs in German and Zanon's (2023) account of the syntax of *tol'ko* 'only' in Russian, which will provide the basis for the proposed analysis in Section 3.

2.2 The Particle Theories of Büring & Hartmann (2001) and Zanon (2023)

2.2.1 Büring & Hartmann's (2001) Particle Theory for German

Büring & Hartmann (2001) propose an adverbial-only analysis of German focus-sensitive particles. One of their many arguments is that this kind of analysis naturally excludes the adjunction of FSPs to DPs within PPs or embedded within other DPs, which is ungrammatical in German. (9) shows both ungrammatical cases.

(9)   a. * mit   nur   [HANS]*F*
          with only   Hans
          Intended: 'only with Hans'

      b. * der Bruder nur   [de-s   GRAF-EN]*F*
          the brother only   the-GEN   count-GEN
          Intended: 'only the count's brother'
                                  (German; Büring & Hartmann 2001: exx. 7a, 8a)

As I demonstrate in the next section, the adjunction of FSPs to NPs/DPs within PPs is also ungrammatical in Bulgarian in most cases.

The specific adverbial-only analysis that Büring & Hartmann (2001: ex. 6) propose allows adjunction of FSPs only to projections belonging to the Extended Verbal Projection (EVP). Their Particle Theory (in its preliminary version) consists of four clauses plus an additional clause concerning left-adjunction of FSPs in German and is shown in (10)–(11).

(10)   For any node α marked F in a phrase marker P, let the set of f-nodes of α consist of all nodes β in P such that

    a. β is an EP (extended projection) of some V γ

    b. β is a maximal projection

    c. β dominates α or is identical to α

    d.   there is no EP β' of γ such that β dominates β' and β' meets (10b) and (10c).   (Büring & Hartmann 2001: ex. 11)

(11)   A FSP must be left-adjoined to an f-node of its focus.
                                                    (Büring & Hartmann 2001: ex. 12)

Apart from adjunction to EVP, the Particle Theory predicts that FSPs only adjoin to maximal projections (10b), that the FSP has to dominate the F-marked constituent (10c) ("dominate" means "c-command" for the purpose of the discussion here), that FSPs adjoin to the focus as closely as possible (10d), and that FSPs can only be left-adjoined in German.[10] In Section 3.3, I discuss how this Particle Theory could be adapted to Bulgarian, after arguing that an adverbial-only analysis should, in fact, be pursued for this language.

   Büring & Hartmann's (2001) proposal has been met with criticism in the literature, much of which has implications for a Particle Theory for German, but not necessarily for the Particle Theory for Bulgarian developed here. Reis (2005) remarks that the PT for German sometimes predicts V3 structures that should be ungrammatical according to the strict V2 requirement in German. This is mostly irrelevant for Bulgarian, but see Mursell (2021) for a defense of this aspect of Büring & Hartmann (2001). The reconstruction-based arguments employed by Büring & Hartmann (2001: Section 5) have also been an object of debate, for example in Meyer & Sauerland (2009) and Smeets & Wagner (2018). I will leave this aspect of the debate aside for now as I will not employ reconstruction-based arguments for developing my Particle Theory for Bulgarian (but see, again, Mursell 2021 for an extensive and recent discussion of the reconstruction facts in German). Additionally, Reis (2005) discusses the adjacency requirement already mentioned (which is termed the "closeness condition" in Reis 2005). The following example is given by her to show that the adjacency requirement can be violated in German:

---

[10]See Büring & Hartmann (2001) for a detailed discussion of and argumentation for the individual clauses, and Sudhoff (2010) as well as Mursell (2021) for discussion and criticism of individual aspects of their proposal.

(12) a. Ich hab nur {darin   / in dem Buch} [geLESen]]$_F$.
     I   have only therein in the book   read
     'I have only read it/the book.'

   b. Ich hab {darin   / in dem Buch} nur [geLESen]$_F$.
     I   have therein in the book only read
     'I have only read it/the book.'   (German; Reis 2005: ex. 23a)

Regarding this example, I agree with Mursell (2021: 230) with respect to the questionable grammaticality of (12a). (12a) is labelled as grammatical in Reis (2005), a fact with which many German speakers do not agree. More important, however, is that (12a) does not violate the adjacency requirement, as *nur* 'only' still adjoins to the projection of the EVP most immediately dominating the F-marked constituent. The only difference between (12a) and (12b) is that there is no scrambling of the argument out of the VP in (12a), an option which is permitted under Büring & Hartmann's (2001) theory (Mursell 2021: 230). Regarding the closeness condition/adjacency requirement, it is interesting to note that evidence for the validity of such a condition can also be found in languages unrelated to the ones discussed in this paper. Erlewine (2017) discusses focus association with *chỉ* 'only' in Vietnamese and argues that *chỉ* needs to adjoin as early as possible in each phase during the derivation, relating the closeness condition to cyclic structure-building facts. This is in line with findings concerning the exhaustive focus marker *shì* in Mandarin (Erlewine 2022). These findings are important in the context of Büring & Hartmann's (2001) theory as they show that the closeness condition can be motivated independently of the facts found for German, and, crucially, independently of the arguments brought forward by Büring & Hartmann that have been criticized so markedly in the literature.

Later on in their paper, Büring & Hartmann (2001: 265-266) modify their proposal further and argue that FSPs only adjoin to non-arguments (this condition replaces the EVP condition mentioned above). Among tricky CP adjunction data discussed by the authors, this proposal also accounts for cases of adjunction within DPs that would be excluded by the EVP condition, such as the cases in (13).

(13) a. eine nur   an [MUSIK]$_F$ interessierte Student-in
     a   only in music   interested   student-F
     'a student interested only in music'

   b. der sogar mit   [KARL]$_F$ verfeindete Förster
     the even with Karl   quarreling forest_ranger
     'the forest ranger who is quarreling even with Karl'

    c. unser auch von [Origami]$_F$ begeisterter Hausmeister
        our   also of Origami   enthusiastic janitor
        'our janitor who is enthusiastic also about Origami'

                                          (German; Büring & Hartmann 2001: ex. 74)

In these cases, the modified Particle Theory predicts the adjunction of the FSP to the modifier instead of DP, which would then be adjunction to a non-argument. This aspect of Büring & Hartmann's proposal has been met with criticism as well; however, a detailed discussion of this would go beyond the scope of this paper and can be found in Mursell (2021). Additionally, Mursell (2021: 247-248) discusses DP data such as (14), which is not explained by the "adjunction to non-arguments"-condition either.[11]

(14)   a.   ein nur [MIttelmäßiger]$_F$ Student
           an only mediocre   student
           'an only mediocre student'

      b. * eine nur [ROte]$_F$ Tasche
           an only red   bag
           Intended: 'an only red bag'   (German; Mursell 2021: 247, ex. 78)

As discussed by Mursell (2021), the fact that individual modifiers provide different adjunction options for FSPs or even prohibit adjunction points towards the fact that there could be additional semantic reasons that permit or prohibit FSP adjunction. Please note as well that Büring & Hartmann's (2001) proposal excludes data such as the cases in (6) found in Bulgarian, which demonstrate that the adjunction of FSPs to PPs within NPs/DPs is possible in Bulgarian. A possible solution to this is offered by the next proposal, Zanon (2023), to be discussed in the following section.

2.2.2 Zanon's (2023) analyis of Russian *tol'ko*

Zanon (2023) examines the behavior of Russian *tol'ko* 'only', arguing that *tol'ko* is always adjacent to the F-marked constituent due to a strong [Foc] feature of *only* that triggers movement of the F-marked constituent to a position adjacent to it. Along the lines of Rudin's (1988) proposal for Bulgarian multiple wh-questions,

---

[11]While I judge (14b) as degraded, it improves for me as an answer to a question such as *What kind of bag would you like to have?*

Zanon (2023: 420) argues that *tol'ko* and the F-marked constituent form an unsplittable complex, as shown in Figure 1.[12,13]
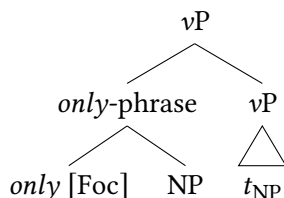
Figure 1: Structure of an *only*-phrase from Zanon (2023: ex. 7)

Similarly to my proposal for Bulgarian, Zanon (2023) provides three arguments against *tol'ko* being an NP-adjunct. Just like Bulgarian and German, Russian does not allow the insertion of *only* into a PP:[14]

(15) a. * dlja tol'ko sestry
          for only   sister
          Intended: 'only for the/a sister'

    b. * ... s   tol'ko krupnymi finansovymi gruppami
            with only   large   financial   groups
            Intended: '... only with the large financial groups'

(Russian; Zanon 2023: exx. 8a, c)

---

[12]As a reviewer notes, Zanon's adaptation of Rudin's analysis presupposes that the phrase that is the sister to *only* can move into a non-c-commanding position, along the lines of Rudin's proposal for the right-adjunction of Bulgarian wh-words in multiple wh-environments. Rudin refers to Chomsky (1986), who argues that this type of movement is an option in these environments. While it would be interesting to investigate the disadvantages of this adaptation and the advantages of other analyses, I must postpone this to future research, as this is not the focus of my paper.

[13]While the adjacency account correctly derives the particle placement facts for Bulgarian, the result cannot always be an unsplittable complex, as I discuss in Section 4 of this paper.

[14]A reviewer notes that the insertion of *tol'ko* 'only' into a PP is sometimes possible in Russian and provides the following example:

(i) vopreki   tol'ko zdravomu   smyslu
     in.spite.of only   common.DAT sense.DAT
     'in spite of only common sense'

This is in line with the occasional circumvention of the "no PP-insertion"-prohibition that can be found in Bulgarian. Since Russian is not the focus of this paper, I cannot discuss these Russian examples further, but suggest that if Russian and Bulgarian pattern similarly here, these cases are rather infrequent in comparison to the general prohibition at work in both languages.

Additionally, *tol'ko* does not pattern with adjectival or adverbial modifiers that would be expected to be NP-adjuncts (Zanon 2023: 422). (16) and (17) show that *tol'ko* neither patterns with adverbial modifiers like *očen'*[15] nor with adjectival modifiers like *svežuju*.[16]

(16)   a. * Tol'ko vy   [SVEŽUJU]$_F$ rybu kupili?
          only   you fresh   fish bought
          Intended: 'Did you only buy the [FRESH]$_F$ fish?'

       b.   Očen ty   bol'šuju cenu za škury zaprosil.
            very you big   price for pelts asked
            'You requested too high a price for the pelts.'

                                              (Russian; Zanon 2023: ex. 11)

(17)   a. * Vy tol'ko kupili [RYBU]$_F$?
            you only   bought fish
            Intended: 'Did you only buy [FISH]$_F$?'[17]

       b.   Vy [SVEŽUJU]$_F$ kupili rybu?
            you fresh   bought fish
            'You bought [FRESH]$_F$ fish?'

                                              (Russian; Zanon 2023: ex. 12)

A last argument in favor of *tol'ko* not being adjoined to NP is of a semantic nature: As observed in Taglicht (1984), NP-adjacent *only* in English causes scope ambiguity. Zanon (2023: 423-424) does not find the analogous ambiguity in Russian, as (18) shows.

(18)   a. Ja žaleju, čto poceloval tol'ko [MAŠU]$_F$.
           I regret that kissed   only   Maša
           'I regret that I only kissed [Maša]$_F$.'

           i.   ...and no one else.
           ii. # ...but I don't regret that I kissed Anastasia.

---

[15]It is unclear to me why *očen'* is analyzed as an adverbial modifier here when it seems to behave more like a degree expression whose purpose is to modify the adjective, as noted by a reviewer. I leave these terminological problems aside for now as this respective example does not affect my analysis of Bulgarian.

[16]Transliteration adapted to scientific transliteration in (16).

[17]A reviewer notes that this example is not fully ungrammatical for some Russian speakers, but only degraded. In Bulgarian, it is generally not possible to separate FSPs from the constituents they associate with, although there are (scarce) examples to be discussed in Section 3. Even if the adjacency facts are not as clear-cut in Russian as presented in Zanon (2023), this does not affect my analysis of Bulgarian.

b. Ja tol'ko [MAŠU]$_F$ žaleju, čto   poceloval.
 I   only   Maša   kissed that regret
 'I only regret that I kissed [Maša]$_F$.'

  i. # ...and no one else.
  ii.   ...but I don't regret that I kissed Anastasia.

(Russian; Zanon 2023: ex. 15)

With respect to the potential adjunction sites of *tol'ko*, Zanon argues that *v*P, CP, and a functional projection in the DP, namely FP, are potential adjunction sites for *tol'ko*. What unites these projections is that they can all be argued to be phases in Russian that are functional projections at the same time. The 'only'-complex can adjoin to *v*P in the verbal domain (with the verb optionally raising above the complex then) (Zanon 2023: 426-427).[18] *Tol'ko* is adjoined to CP in *tol'ko*+subject complexes as in (19) (Zanon 2023: 429).[19]

(19)   Tol'ko [IVAN]$_F$ posmotrel ėtot fil'm.
    only   Ivan   watched   this movie
    'Only [IVAN]$_F$ watched this movie.'   (Russian; Zanon 2023: ex. 25)

In the nominal domain, Zanon (2023: 432-433) notes that *tol'ko* can be adjoined to the genitive complement inside an NP (or to parts of it such as the numeral modifier shown below), as demonstrated in (20). This fact can be accounted for

---

[18]This analysis would run into problems under approaches that assume that the finite verb never moves out of *v*P in Russian, as a reviewer remarks. I refer the reader to the discussion of this issue in Zanon (2023). Since movement of the finite verb to T is assumed for Bulgarian, this debate is less relevant for the analysis that I am pursuing here. See, for example, Harizanov (2019) for a recent discussion and summary of verb position in Bulgarian.

[19]A reviewer provides the following example and remarks that Zanon's analysis would not be able to account for subjects in embedded CPs that follow a complementizer in C:

(i)   Ja znaju čto   tol'ko [IVAN]$_F$ posmotrel ėtot fil'm.
    I   know that only   Ivan   watched   this movie
    'I know that only Ivan watched this movie.'

I leave the solution of this problem under Zanon's account open here but would suggest that examples like this perhaps show that Zanon's restriction of adjunction sites to CP, *v*P, and FP is too restrictive for Russian. Again, this problem does not affect my more permissive analysis of adjunction sites for Bulgarian FSPs.

by the existence of a functional projection FP that is able to host *tol'ko*.[20] As I discuss in Section 3.3, this account can be extended to the nominal domain in Bulgarian, but also runs into problems depending on the modifier studied.

(20)   Ja znaju [studentov tol'ko PERVOGO$_F$ kursa].
        I   know   students   only   first   year
        'I know only the [FIRST]$_F$ year students.'   (Russian; Zanon 2023: ex. 32c)

In a way, Zanon (2023) provides an explanation for the adjunction patterns of *only* in Russian that is the exact opposite of what Tisheva & Dzhonova (2003) assume for Bulgarian: Instead of arguing that *only* can adjoin to any type of syntactic constituent, Zanon limits the number of adjunction sites of *only*. As a next step, the F-marked constituent moves towards the position of *only*, instead of *only* adjoining to the respective F-marked constituent anywhere in the clause. This analysis correctly rules out the restrictions on the placement of FSPs in Bulgarian that Tisheva & Dzhonova's (2003) account leaves unexplained, as I argue in Section 3.

3 An adverbial-only analysis of Bulgarian FSPs

In this section, I present arguments against an adnominal analysis of the adjunction behavior of Bulgarian FSPs as well as arguments in favor of an adverbial-only analysis (Section 3.1). Section 3.2 provides additional discussion of the behavior of Bulgarian FSPs in the nominal domain and the challenges that this poses for Büring & Hartmann's (2001) Particle Theory. Finally, I introduce an adaptation of Büring & Hartmann's Particle Theory to Bulgarian (Section 3.3).

3.1 Arguments against an adnominal analysis

The biggest advantage of pursuing an adverbial-only analysis of Bulgarian FSPs is that an analysis of this type predicts and explains certain distributional facts that an adnominal analysis struggles to account for. One distributional fact that has been observed for German, as already mentioned in Section 2.2, is the impossibility of adjoining FSPs to DPs within PPs, which adnominal analyses would predict to be an option. (21) shows that the prohibition, which is judged to be

---

[20]Examples such as (20) differ in acceptability among Russian speakers, as a reviewer remarks. I must leave open for now why that is the case and how widespread this divergence is in Russian. As examples of this kind are grammatical in all environments in which I have tested them, Zanon's prediction is still borne out in Bulgarian.

sharply ungrammatical by all my consultants, is not limited to a specific FSP or the involvement of a specific preposition. The prohibition carries over to all FSPs and prepositions that I tested.[21]

(21) a. * Kupi-h   krastavic-i   za samo [SALATA-TA]$_F$.
         buy-PST.1SG cucumber-PL for only   salad-DEF
         Intended: 'I bought cucumbers only for the salad.'

      b. * Ana glasuva   sreštu   samo [MARIA]$_F$
         Ana vote.PST.3SG against only   Maria
         Intended: 'Ana voted only against Maria.'

      c. * okolo   dori golemite [GRAD-OVE]$_F$
         around even big   city-PL
         Intended: 'even around big cities'

Adverbial-only analyses neatly predict the PP-insertion prohibition. Furthermore, they also account for the exclusion of adjunction to NPs/DPs in other environments in which the FSP would be forced to adjoin to phrases of this type. An example for this is the coordination test proposed by Jacobs (1983) for German, as shown in (22).

(22)   * dass Peter und {nur / sogar / auch} Luise sich in Straßburg trafen
        that Peter and only   even   also Luise REFL in Straßburg meet.PST
        Intended: 'that Peter and {only / even / also} Luise met in Straßburg'

$$\text{(German; Jacobs 1983: 45, ex. 3.29b)}$$

As noted by a reviewer, a purely syntactic account of these examples is unable to account for their ungrammaticality. While I generally agree with this view (and consider it not to be incompatible with my argumentation), I argue that there is a certain component of the ungrammaticality of these examples that can be explained by particle placement. The reviewer gives two reasons for their scepticism: First, *nur* 'only' should be incompatible with DP coordination irrespective of syntax due to its exhaustive interpretation. This cannot be entirely true since

---

[21]A possible exemption is *vmesto* 'instead'. This preposition is the only one which can be inserted within PPs, examples of which can be found in the *Bulgarian National Corpus* (Koeva et al. 2010). This example is from an excerpt of a (spoken) debate: *Pǎrvo da se glasuva pǎrvata čast na teksta s predloženieto na gospodin Bučkov vmesto samo [LICA]$_F$*, 'First to vote the first part of the text with Mr. Bučkov's proposal instead of only persons'. At this point, it is unclear to me why the PP insertion prohibition does not extend to *vmesto*. However, even with this preposition, examples of FSP-insertion within PPs are scarce and Bulgarian speakers prefer to place the FSP before the preposition.

(22) is also degraded/ungrammatical if other FSPs such as *sogar* 'even' and *auch* 'also' are used. Jacobs (1983: 45) discusses the impact of semantic factors on the ungrammaticality of the example and points out that reversing the order of conjuncts makes the example grammatical:

(23)   dass nur   / sogar / auch [LUISE]_F und Peter sich in Straßburg trafen
         that only   even   also Luise   and Peter REFL in Straßburg meet.PST
         'that only / even / also Luise and Peter met in Straßburg'

Even if only *Luise* is focused in this example, it is still grammatical, which could be explained by the fact that the FSP can adjoin to an EVP-projection in this case. This extends to examples with non-reflexivized verbs such as (24a), versus (24b).

(24)   a.   Nur [MARIE]_F und Luise haben die Klausur bestanden.
              only Marie   and Luise AUX.PL the exam   pass.PTCP
              'Only Mary and Luise passed the exam.'

         b. * Marie und nur   [LUISE]_F haben die Klausur bestanden.
              Marie and only Luise   AUX.PL the exam   pass.PTCP
              Intended: 'Marie and only Luise passed the exam.'

Using a non-reflexivized verb improves the situation in the case of *even* and *also* (I consider the examples presented here as only slightly degraded in German with these two FSPs instead of *only*). Nevertheless, the "semantic explanation" does not fully account for why the reversal of the order of conjuncts should lead to grammaticality here.

   In (22), *nur* 'only' is forced to adjoin to the coordinated DP *Luise*, which results in ungrammaticality. This extends to Bulgarian, as can be seen in (25).

(25)   ??/* Znaj-a,   če   Peter i   samo [ANNA]_F se   sreštna-ha   v
            know-1SG that Peter and only   Anna   REFL meet-PST.3PL in
         Berlin.
         Berlin
         Intended: 'I know that Peter and only Anna met in Berlin.'

Consequently, the adverbial-only analysis predicts that adjunction to VP should not be a problem in a coordinating construction. This is borne out, as demonstrated by (26).[22]

(26) Peter izle-ze   i   samo [PAZAR-UVA]$_F$.
     Peter go.out-3SG.PST and only shop-PST
     'Peter went out and only did his shopping.'

An additional argument against the incorporation of Bulgarian FSPs into the NP/ DP is provided by observable stranding phenomena. Stranding of *nur* 'only' is possible in German in many instances.[23] The FSP can be stranded in Bulgarian, as in (27).[24]

---

[22]While the results of the coordination test fit the predictions made by the adverbial-only analysis, it is important to note that a test of this kind should not be used on its own to make predictions about the correctness of this analysis, since in special constructions such as coordinated structures, there could be other interfering factors at work. Moreover, the intuition of German speakers concerning the German equivalents of examples such as (26) differ, a problem that I must leave for further research for now. A reviewer points out that adjunction of *only* to the second conjunct should also be difficult in cases of VP-adjunction that lack the purpose reading found in (26) such as **John cried and only laughed*. The German equivalent *Jan hat geweint und nur gelacht* is grammatical to me, especially under a temporal interpretation of the conjunction. Again, I conclude that only an analysis that takes semantic and syntactic factors going hand in hand into account can grasp adjunction data of this kind to its full extent. Nevertheless, VP-adjunction seems to be often possible in cases in which DP-adjunction is not, favoring the adverbial-only analysis.

[23]See Mursell (2021) for discussion.

[24]A reviewer notes that this example could also be a case of NP-splitting. NP-splitting is possible in Bulgarian, but conflicting judgments are constantly being reported in the literature on these splits so that it is difficult to determine which splits are accepted by a majority of speakers and which ones are not. While the possibility of NP-splitting should be kept in mind when interpreting my examples here, I argue that what we can observe in (27) is not an NP-split as *samo* does not pattern with, for example, adjectival modifiers here. In many NP-splits, it is possible to strand the noun and front the adjective, as in the following example:

(i) Nova$_1$ e kupil   [t$_1$ kola] (ne   stara).
    new   is bought   car   not old
    'He bought a new car, not an old one.'   (Tasseva-Kurktchieva & Dubinsky 2018: ex. 36a)

This is not possible with *samo* and other FSPs that I have tested, as fronting *samo* and stranding the noun would mean that *samo* is not associated with the stranded noun anymore, but with the constituent to its right. Additionally, an explanation would be needed for why only the lowest NP can be split in this case so that *samo* is moved to the left periphery. I therefore tentatively conclude that my example does not show an NP-split.

(27) a. Čet-a   samo [ROMAN-I]$_F$.
          read-1SG only   novel-PL
          'I read only [novels]$_F$.'

     b. [ROMAN-I]$_F$ čet-a   samo.
           novel-PL   read-1SG only
          '[NOVELS]$_F$, I read only.'

Since *romani* 'novels' is placed above the verb in this example, we can conclude that it moved above TP, suggesting that the F-marked constituent moved to the designated FocP in the left periphery.

    These arguments taken together suggest that there are not as many adjunction sites for Bulgarian FSPs as descriptive analyses such as Tisheva & Dzhonova (2003) suggest. In fact, adjunction seems to be restricted to projections belonging to the EVP, a proposal which is in line with Büring & Hartmann (2001) and less restrictive than Zanon's (2023) analysis of Russian *tol'ko*. However, an additional adjunction site in the nominal domain is needed to account for the adjunction options of Bulgarian FSPs there, as I show in the next section.

3.2 Bulgarian FSPs in the nominal domain

Bulgarian FSPs in the nominal domain show the importance of not only taking syntactic but also semantic factors into account when determining possible adjunction sites for FSPs. As discussed by Büring & Hartmann (2001), German FSPs can circumvent the prohibition against adjunction to DPs inside PPs if they are adjoined to an adjectival or numeral modifier, as demonstrated in (28).

(28) a. mit nur [EINEM]$_F$ Wagen.
          with only one   car
          'with only [ONE]$_F$ car.'

     b. in nur [WENIGEN]$_F$ Sekunden
          in only few   seconds
          'within only [A FEW]$_F$ seconds.'

<div align="right">(German; Büring & Hartmann 2001: exx. 82a, c)</div>

Büring & Hartmann (2001) account for this by further generalizing from "adjunction to EVP" to "adjunction to non-arguments" as the principle governing particle placement in German. However, "adjunction to non-arguments" does not explain the fact that German FSPs are unable to adjoin to some modifiers, as discussed in Section 2.2. Mursell (2021: 247) discusses the possibility that there

could be a bigger reason explaining adjunction possibilities in general, such as that FSPs only adjoin to elements that introduce a scale (an observation that he attributes to Karen De Clercq), which could be argued for some of the modifiers discussed by him as well as for verbs, accounting for the facts captured by the adverbial-only analysis as well. While it is definitely necessary to consider this bigger reason behind adjunction possibilities that goes beyond a syntactic treatment of the problem, a first step is to successfully capture the adjunction options of FSPs from a syntactic perspective. If we consider Bulgarian data equivalent to the German data discussed above, it becomes evident that Bulgarian does not pattern with German here but shares many characteristics with Russian in this respect. Most importantly, Bulgarian FSPs are able to adjoin to PPs and DPs embedded within DPs, which is impossible in German, as a direct comparison shows (in this example, Bulgarian *samo* adjoined to a PP within a DP while we can observe the impossibility of adjunction to DP within a DP in the German example):

(29)   a.   * die   Katze nur   [de-s   NACHBAR-N]$_F$ (German)
            the.F cat   only   the-M.GEN neighbor-M.GEN
            Intended: 'only the cat [of the neighbor]$_F$'

       b.   kotka-ta samo [na SĂSEDKA-TA]$_F$
            cat-DEF only   of neighbor-DEF
            'only the cat [of the neighbor]$_F$'

This adjunction behavior is a major obstacle for a Particle Theory in the style of Büring & Hartmann (2001) as data points such as (29) are one major argument for excluding adnominal adjunction in German. However, these examples can be reconciled with the help of Zanon's (2023) proposal for Russian. Under the assumption that Bulgarian FSPs adjoin to a functional projection in the nominal domain, FP, as she proposes for Russian, the examples can be captured by her analysis. Russian and Bulgarian pattern similarly here:[25]

(30)   a.   Ja znaju tol'ko [studentov   PERVOGO$_F$ kursa].
            I   know only   students.ACC first.GEN   year.GEN
            'I only know the [FIRST]$_F$ year students.'

---

[25]I prefer Zanon's approach over Büring & Hartmann's approach here as Zanon's approach does not exclude the adjunction options shown in (29) to be possible in Bulgarian. Adopting the "adjunction to non-arguments"-condition for Bulgarian would mean that the Bulgarian examples in (31) could be explained by arguing for adjunction to the numeral, while (29) would be predicted to be ungrammatical as there is no modifier present.

    b. Ja znaju [studentov   tol'ko PERVOGO$_F$ kursa].
        I know students.ACC only first.GEN   year.GEN
        'I know only the [FIRST]$_F$ year students.'

                                    (Russian; Zanon 2023: ex. 32a, c)

(31) a. Pozna-vam samo [student-i   PĂRVA godina]$_F$.
        know-1SG only student-PL first   year
        'I only know the [FIRST]$_F$ year students.'

      b. Pozna-vam student-i   samo [PĂRVA godina]$_F$.
        know-1SG student-PL only first   year
        'I know only the [FIRST]$_F$ year students.'

While Zanon's analysis captures these facts effortlessly, neither this analysis nor newer proposals made for languages such as German, e.g. Mursell (2021), account for the variation in adjunction behavior to different modifiers at this point. Adjunction to modifiers within PPs in Bulgarian varies depending on the modifier and the preposition involved. (32) can be accounted for with the help of Mursell's proposal for scalar modifiers mentioned above (except for (32a), which possibly requires additional phonological considerations), as a reviewer notes. The ungrammaticality of the last two examples could then be explained by the fact that 'big' is not a scalar modifier.

(32) a. * săs   samo edna kola
         with only one   car
         Intended: 'with only one car'

      b. $^?$ sled   samo njakolko sekund-i
         within only few   second-PL
         'within only a few seconds'

      c. meždu   samo dve optsi-i
         between only two option-PL
         'between only two options'

      d. meždu   dori dve optsi-i
         between even two option-PL
         'between even two options'

      e. * okolo   samo golemite grad-ove$_F$
         around only big   city-PL
         Intended: 'only around big cities'

      f. * okolo   dori golemite grad-ove$_F$
         around even big   city-PL
         Intended: 'even around big cities'

The puzzle involving the restrictions that different prepositions and modifiers impose on FSP placement in Bulgarian cannot be resolved here. However, it allows us to draw a few conclusions for the analysis. First, it points towards the fact that Zanon's (2023) analysis involving adjunction to FP is on the right track for Bulgarian, while Büring & Hartmann's (2001) adjunction to non-arguments would not be able to capture the available adjunction sites of Bulgarian FSPs in the nominal domain. Second, it is evident how challenging nominal data is for adverbial-only approaches. An additional assumption, such as adjunction to FP, is needed in order to capture the empirical facts. There are two possibilities to develop a proper account that is able to capture the difference between individual modifiers. The first would be to aim for a separate treatment of FSP-adjunction in the clausal and nominal domain, which would be the less economic approach. A second option is the one already sketched, namely exploring the relationship between possible adjunction sites of FSPs and elements introducing a scale. Much of the data discussed in this section can be explained by the presence or absence of scalar modifiers, but (29) remains unexplained under this approach. Nevertheless, I consider this a fruitful path for future research in this area that aims to not only account for the Bulgarian facts, but for FSP adjunction in other languages such as German as well.

3.3 A Particle Theory for Bulgarian

Based on the facts already discussed here, I consider it reasonable to develop a Particle Theory for Bulgarian based on Büring & Hartmann's (2001) Particle Theory for German, since both languages show surprisingly similar patterns with respect to the adjunction behavior of their FSPs. At the same time, essential parts of Zanon's (2023) proposal, such as the possibility of adjunction to a functional projection FP in the nominal domain, are needed to account for the Bulgarian data. I therefore argue for a combination of both proposals for Bulgarian. In this section, I discuss the five clauses of Büring & Hartmann's Particle Theory and how these conditions could be adapted to the Bulgarian facts.

3.3.1 Adjunction to EVP

As discussed in the previous subsections, an adverbial-only analysis of the adjunction behavior of Bulgarian FSPs elegantly excludes the insertion prohibitions that can be found in the language. However, not all adverbial-only analyses are alike. Büring & Hartmann (2001) themselves propose two of them: They first argue that German FSPs only adjoin to the EVP and then further generalize to

adjunction to non-arguments, as described in Section 2.2. While this generalization is, as already discussed, not unproblematic for German, the previous section has shown that adjunction to non-arguments would also not be able to capture the adjunction behavior of Bulgarian FSPs in the nominal domain. A related, but distinct option would therefore be Zanon's (2023) proposal that assumes that Russian *tol'ko* only adjoins to *v*P, CP, and FP. Please note that Zanon's analysis excludes adjunction to TP, which Zanon rules out based on examples such as (33).

(33) Ja ne   znaju...
      I NEG know
      'I don't know...'

   a. $^{?/*}$ posmotrel li tol'ko [IVAN]$_F$ ėtot fil'm.
            watched   Q only   Ivan   this movie
         Intended: 'if only [Ivan]$_F$ watched this movie.'

   b. posmotrel li Ivan tol'ko [ĖTOT]$_F$ fil'm.
       watched   Q Ivan only   this   movie
       'if Ivan watched only [THIS]$_F$ movie.'   (Russian; Zanon 2023: ex. 24)

In (33a), *tol'ko* must be adjoined to TP, below *li* in the CP, which is degraded in Russian. The example becomes grammatical once *tol'ko* is adjoined to *v*P, as in (33b) (Zanon 2023: 428-429).[26] Reproducing these examples in Bulgarian shows that Bulgarian is more permissive here and allows adjunction to TP as well. I conclude that the "adjunction to EVP"-condition that is more permissive than Zanon's proposal makes the correct predictions for Bulgarian.

---

[26]A reviewer notes that TP-adjunction could, in fact, be possible in Russian, and provides the following example:

(i) Ja ne   znaju...
     I NEG know
     'I don't know...'

      a. posmotrel li tol'ko [ĖTOT]$_F$ student ėtot fil'm.
          watched   Q only   this   student this movie
          'if only [THIS]$_F$ student watched this movie.'

      b. $^{?/*}$ posmotrel li Ivan tol'ko [FIL'M]$_F$.
               watched   Q Ivan only   movie
            'if Ivan watched only a [MOVIE]$_F$.'

This example potentially shows TP-adjunction, so Russian could be more permissive than assumed by Zanon (2023), and therefore also closer to Bulgarian in this respect. In any case, my proposal for Bulgarian allows for TP-adjunction and is therefore not affected by the pattern found for Russian here.

(34) Az ne znaj-a…
      1SG NEG know-1SG
      'I don't know…'

    a. dali   samo [IVAN]$_F$ e   gleda-l   tozi film.
          whether only   Ivan   AUX.3SG watch-PTCP this movie
          'whether only [Ivan]$_F$ watched this movie.'

    b. dali   Ivan e   gleda-l   samo [TOZI]$_F$ film.
          whether Ivan AUX.3SG watch-PTCP only   this   movie
          'whether Ivan watched only [this]$_F$ movie.'

3.3.2 Adjunction to maximal projections

As Büring & Hartmann (2001: 240-244) discuss, there are theory-internal reasons that make it desirable to uphold the requirement that FSPs adjoin to maximal projections only. I argue that the clause should hold for Bulgarian as well, since it naturally excludes cases such as (35). Here, the FSP can adjoin to the auxiliary in T, but not to the finite verb on its own.

(35) a. Ti samo šte   [SEDI-Š]$_F$
          you only AUX.FUT sit-PRS.2SG
          'You will only sit.'

    b. * Ti   šte   samo [SEDI-Š]$_F$.
          you AUX.FUT only   sit-PRS.2SG
          Intended: 'You will only sit.'

3.3.3 The c-command condition

The c-command criterion can be maintained for Bulgarian if it is adapted in a way that allows for the FSP to not necessarily c-command the F-marked constituent only, but also its trace in cases in which the focused constituent moved above the FSP.[27] In all other cases, the local feature checking relationship proposed by Zanon (2023) for Russian also holds for Bulgarian. In fact, it is reasonable to argue that in the cases of right-adjunction discussed here, the F-marked constituent moved to adjoin to the FSP to locally check the strong [Foc] feature and moved above the FSP in a second step.[28] I discuss further details of this approach in Section 4.

---

[27]Association of focus with traces is extensively discussed in Erlewine (2014), including discussion of previous work on focus association that deemed this to not be a possible operation.

[28]As I will briefly touch upon in Section 4, this optional step would have to be constrained by discourse-level constraints instead of being driven by a syntactic feature.

3.3.4 The adjacency requirement

In contrast to other criteria such as the EVP requirement, the adjacency requirement does not have to be adapted for Bulgarian. There are abundant examples, such as (36), demonstrating that FSPs need to adjoin to their F-marked constituent as closely as possible.[29]

(36) * Kupi-h samo krastavic-i   [ZA SALATA-TA]$_F$.
      buy-PST only cucumber-PL for salad-DEF
      Intended: 'I bought cucumbers only [for the salad].'

At this point, a typological remark is in order. Although both Bulgarian and German adhere to the adjacency requirement, both languages still differ in the adjunction behavior of their FSPs insofar as Bulgarian FSPs do seem to possess a strong [Foc] feature that triggers movement of the F-marked constituent to the position of the FSP. While Zanon (2023: 428)'s typological generalization, namely that "in overt focus movement languages, a focalized XP-associate must be adjacent to the F-licensing element" was made with Russian in mind, the same holds for Bulgarian. Just like Russian, Bulgarian possesses overt focus movement and adheres to the adjacency requirement. Despite the similarities between FSP-adjunction in Bulgarian, Russian, and German, German seems to be located in another place in the typological realm here. German does not possess overt focus movement, but still requires adjacency if the maximal projection requirement is not violated and the syntax of the language permits adjacency. German is taking a middle ground here between the two stricter Slavic languages discussed and languages whose FSPs adjoin more loosely in general, such as English. While further developing this discussion would go beyond the scope of this paper, I want to underline the insights that could result from an investigation of FSP placement from a typological perspective.[30]

---

[29]The adjacency requirement is very strict in Bulgarian, and it is difficult to find examples in which adjacency of an F-marked constituent to its FSP is not required. The following is a puzzling example (adapted from Tisheva & Dzhonova 2003), as adjacency is not required here for at least some of my consultants (although judgments differ). For this group of speakers, all positions of *samo* indicated in the example are possible while the FSP is associated with the F-marked constituent.

(i) Tova (samo) može (samo) da băde (samo) [ofis-ăt   na MICROSOFT]$_F$.
    this only   could only to be   only office-DEF of Microsoft
    'This could only be the office of Microsoft.'

However, these examples are scarce, and it is unclear at this point for how many Bulgarian speakers they are grammatical.

[30]I thank Željko Bošković for discussion of this point.

3.3.5 Left-adjunction

While the left-adjunction criterion is absolutely necessary to derive the correct particle placement for German, Bulgarian FSPs do not have to be left-adjoined to the F-marked constituent. Nevertheless, they usually remain close to their focused constituent, even when they surface to the right of the F-marked constituent, the reasons for which I discuss in Section 4. In order to capture the apparent right-adjunction of Bulgarian FSPs, Büring & Hartmann's (2001) fifth clause has to be adapted so that not only left-adjunction to an f-node of the FSP's focus is allowed, but also left-adjunction to the trace left behind by the F-marked constituent moving above the FSP.

3.3.6 Summary

Summing up, I propose the following Particle Theory for Bulgarian, based on Büring & Hartmann's (2001) proposal combined with the analysis by Zanon (2023), adapted to account for the Bulgarian data discussed here.

(37) THE PARTICLE THEORY FOR BULGARIAN
For any node α marked F in a phrase marker P, let the set of f-nodes of α consist of all nodes β in P such that

   a. β is an EP (extended projection) of some V γ or a functional projection FP within DP

   b. β is a maximal projection

   c. β dominates α or a trace of α or is identical to α

   d. there is no EP β' of γ such that β dominates β' and β' meets (37b) and (37c).

(38) A FSP must be left-adjoined to an f-node of its focus or its trace.

4 Post-focal FSPs in Bulgarian

As already discussed, Bulgarian FSPs must adjoin to the F-marked constituent that they belong to as closely as possible. However, there are two different, but, as I argue, related cases in which the F-marked constituent is able to move out of its position right-adjacent to the FSP. The first case, shown in (39), involves the focused constituent moving above the FSP, but staying immediately above it. The second case, the stranding case shown in (27) (repeated here as (40)), consists

of the F-marked constituent moving to a high position in the clause, presumably FocP in the left periphery.[31]

(39) a. Čet-a   [ROMAN-I]$_F$ samo.
          read-1SG novel-PL   only
          'I read only [novels]$_F$.'

      b. Včera   [az]$_F$ săšto četo-h   roman.
          yesterday I   also read-PST.1SG novel
          'Yesterday, [I]$_F$ also read a novel.'

(40) a. Čet-a   samo [ROMAN-I]$_F$.
          read-1SG only   novel-PL
          'I read only [novels]$_F$.'

      b. [ROMAN-I]$_F$ čet-a   samo
           novel-PL   read-1SG only
          '[NOVELS]$_F$, I read only.'

The first important question that these types of "movement out of focus" (meaning cases in which a focused constituent left its original position right-adjacent to its FSP) raise is what consequences they have for our Particle Theory. Büring & Hartmann (2001) do not assume the possibility of right-adjunction in German, which derives the German adjunction facts correctly. Although the surface word order of the F-marked constituent and the FSP in Bulgarian suggests that right-adjunction is an option in this language, I will not argue for this to be the case for two reasons. First, the F-marked constituent is still interpreted as the constituent associated with the respective FSP, even if it has moved out of its position right-adjacent to the particle. This is surprising given the fact that Bulgarian FSPs are usually interpreted as strictly associating with the constituent following them. This suggests that the FSP associates with the F-marked constituent's trace and that the focused constituent reconstructs at LF when it is interpreted.[32] The (simplified) trees in Figure 2 and 3 show how cases such as (39) and (40), respectively, could be represented.

---

[31]A reviewer asks how optional movement can take place here if a feature-based theory is assumed. While movement of the focused constituent to the position of the FSP seems obligatory, the movement types described here are not. My explanation is that they are, in fact, not feature-driven. The strong [Foc] feature should have already been checked and deleted by the time the focused constituent has moved to the FSP. This optional movement would then be caused by more discourse-based reasons which would have to be explored in the future.

[32]A reviewer asks why it should not be possible for the FSP to operate on the F-marked constituent while being right-adjoined to it. While this is generally an option, I argue that this is
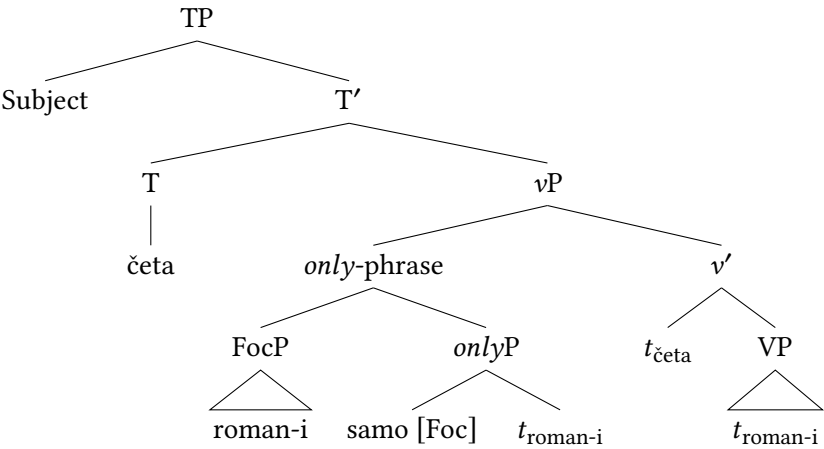
Figure 2: Low movement out of the *only*-phrase
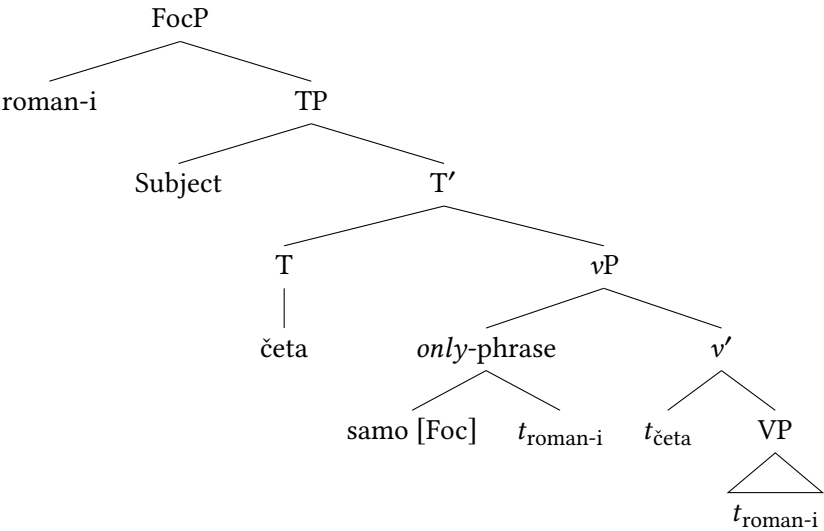
Figure 3: High movement out of the *only*-phrase

As discussed by Erlewine (2014), for example, a similar operation is available in German and Dutch, which are both languages that allow movement of F-marked constituents to clause-initial positions. A second reason for maintaining Büring & Hartmann's (2001) left-adjunction condition for a Particle Theory for Bulgarian is that movement out of focus in Bulgarian is restricted in the sense that only the two types of movement shown in (39) and (40) seem to be allowed. Moreover, movement out of focus in the case in (39) is only permitted if the F-marked constituent is the lowest constituent in the clause, as already noted by Tisheva & Dzhonova (2003), or around the subject position in T. This movement type is not permitted in positions in the middle of the clause since the FSP would then rather be interpreted as being associated with the constituent following it by Bulgarian speakers:[33]

(41)   * Kupi-h   [KRASTAVIC-I]$_F$ samo za salata-ta.
       buy-PST   cucumber-PL   only for salad-DEF
       Intended: 'I bought only cucumbers for the salad.'

Allowing for right-adjunction would overgenerate the options that are there for movement out of focus in Bulgarian and would not predict that there are actually only two positions that the moved F-marked constituent can move to.

At this point, two questions remain open. First, an obvious question is where the F-marked constituent moves to in the low cases of movement out of focus. It is generally noted in the syntactic literature on Bulgarian that while the left periphery of the language is well-researched, much less is known about the verbal domain and the positions that it hosts (Krapova 2002). A possible solution to this problem would be to argue for a low, post-verbal focus position, along the lines of the proposal made by Belletti (2004) for Italian. A second question that I leave open here is whether there are differences in interpretation between the association of an F-marked constituent preceding or following the FSP. As (39) shows, the usage of a post-focal FSP usually requires focal stress on the F-marked constituent, which helps speakers associate and interpret it as belonging to the FSP

---

not the case here. Allowing for general right-adjunction of FSPs in Bulgarian would overgenerate adjunction options in the middle of the clause that are unavailable, as (41) in this section shows. At the same time, Bulgarian FSPs are able to associate with traces of F-marked constituents, which explains cases such as (40). Focus movement to a high and a potentially low focus position is the more economical assumption here.

[33]A reviewer notes that backwards association in the middle of the clause of the kind discussed here is grammatical in Russian once a disambiguating context is employed. To the best of my knowledge, this type of ambiguity does not exist in Bulgarian as all my consultants strongly reject backwards association in the middle of the clause.

following it. This kind of focal stress is not required if the F-marked constituent follows its FSP.[34] Additionally, my consultants (as well as Nicolova 2000) report that there is a register difference between the two low options, with the pre-FSP position that the F-marked constituent can be in being associated with colloquial, informal speech. Future research could focus on further differences between the two positions and what they can tell us about the semantic differences between them.

5 Concluding remarks

In this paper, I argued for an adverbial-only analysis of Bulgarian focus-sensitive particles that combines two proposals, namely Büring & Hartmann's (2001) analysis of German FSPs and Zanon's (2023) analysis of the adjunction behavior of Russian *tol'ko* 'only'. While several arguments point against the feasibility of an adnominal analysis of Bulgarian FSPs, their adjunction options in the nominal domain suggest that Zanon's proposal involving adjunction to FP in the nominal domain is on the right track for Bulgarian, in contrast to Büring & Hartmann's "adjunction to non-arguments"-condition. A gap that I necessarily leave aside in this paper is the question of semantic properties of individual modifiers constraining the adjunction possibilities of Bulgarian FSPs, a question which I argued to be essential for understanding the additional semantic reasons for adjunction, even beyond Bulgarian. Future research could close this gap at this point with a more detailed semantic investigation. Subsequently, I proposed a Particle Theory for Bulgarian based on Büring & Hartmann (2001).

At this point, it becomes evident that Büring & Hartmann's left-adjunction condition that accounts for the rigid exclusion of right-adjunction of FSPs in German cannot be upheld in its original formulation when their analysis is extended to Bulgarian. Bulgarian FSPs are able to move above the FSP dominating them. FSPs adjoined to an F-marked constituent low in the clause can even be stranded while the focused constituent moves to FocP in the left periphery. These two types of movement are, however, highly restricted. F-marked constituents can move above FSPs but have to remain close to them in the first movement type. They are only able to move into the high focus position in the left periphery or

---

[34]A reviewer suggests the possibility that a moved focus is always contrastive. At the same time, in-situ would then be ambiguous between a contrastive and a non-contrastive interpretation. While this is certainly a plausible option, it would not explain what the trigger for this optional movement is. An analysis along the lines of Titov (2020) could solve this problem, but I leave this question open for now.

must remain low, in a position where a second, low focus position could be assumed in Bulgarian. Future research could focus on finding additional evidence for or counterexamples against the existence of such a projection, as well as possible semantic differences between the two available positions for the F-marked constituent that precedes or follows the FSP dominating it. In any case, Bulgarian FSPs can be split from their F-marked constituents after local checking of the strong [Foc] feature, although in a very controlled manner, as argued above.

Finally, the investigation conducted here, on a par with Zanon (2023), suggests fruitful paths for typological research investigating the connection between overt focus movement and the strict adjacency requirement that holds in Russian and Bulgarian. While adjacency is not required in English, it is necessary in both languages. Languages such as German can be placed in the middle ground between these two extremes, with German not requiring overt focus movement, but adjunction as close to the F-marked constituent as German syntax allows. Future investigations into FSP placement could focus on other, also typologically unrelated language families in order to learn more about the connection of adjacency and overt focus movement.

Abbreviations

1	first person	M	masculine
2	second person	PL	plural
3	third person	PRS	present tense
ACC	accusative	PST	past tense
AUX	auxiliary	PTCP	participle
DAT	dative	Q	question
DEF	definite	REFL	reflexive
FUT	future	SG	singular
GEN	genitive		

Acknowledgments

reviewers, whom I would like to thank for their reviews. Additionally, I thank Rajesh Bhatt, Željko Bošković, Katharina Hartmann, and Roumyana Pancheva for discussion of this project. All remaining errors and shortcomings are of course my own.

References

Belletti, Adriana. 2004. Aspects of the low IP area. In Luigi Rizzi (ed.), *The structure of CP and IP* (The cartography of syntactic structures 2), 16–51. Oxford: Oxford University Press.

Branan, Kenyon & Michael Yoshitaka Erlewine. 2023. Anti-pied-piping. *Language* 99(3). 603–653. DOI: 10.1353/lan.2023.a907013.

Büring, Daniel & Katharina Hartmann. 2001. The syntax and semantics of focus-sensitive particles in German. *Natural Language & Linguistic Theory* 19(2). 229–281. DOI: 10.1023/A:1010653115493.

Chomsky, Noam. 1986. *Barriers* (Linguistic Inquiry Monograph 13). Cambridge: MIT Press.

Erlewine, Michael Yoshitaka. 2014. *Movement out of focus*. Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Erlewine, Michael Yoshitaka. 2017. Vietnamese focus particles and derivation by phase. *Journal of East Asian Linguistics* 26. 325–349. DOI: 10.1007/s10831-017-9156-y.

Erlewine, Michael Yoshitaka. 2022. Mandarin exhaustive focus *shì* and the syntax of discourse congruence. In Remus Gergel, Augustin Speyer & Ingo Reich (eds.), *Particles in German, English, and beyond*, 323–354. Amsterdam: John Benjamins. DOI: 10.1075/slcs.224.12erl.

Harizanov, Boris. 2019. Head movement to specifier positions. *Glossa: a journal of general linguistics* 4(1). 140. DOI: 10.5334/gjgl.871.

Jacobs, Joachim. 1983. *Fokus und Skalen: Zur Syntax und Semantik der Gradpartikeln im Deutschen* (Linguistische Arbeiten 138). Tübingen: Max Niemeyer.

Jasinskaja, Katja. 2016. Information structure in Slavic. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure*, 709–732. Oxford: Oxford University Press.

Koeva, Svetla, Diana Blagoeva & Siya Kolkovska. 2010. Bulgarian National Corpus Project. *Politics* 207. 3678–3684.

Krapova, Iliyana. 2002. On the left periphery of the Bulgarian sentence. *University of Venice Working Papers in Linguistics* 12. 107–128.

Meyer, Marie-Christine & Uli Sauerland. 2009. A pragmatic constraint on ambiguity detection: A rejoinder to Büring and Hartmann and to Reis. *Natural Language & Linguistic Theory* 27. 139–150. DOI: 10.1007/s11049-008-9060-2.

Mursell, Johannes. 2021. *The syntax of information-structural agreement* (Linguistik Aktuell/Linguistics Today 268). Amsterdam: John Benjamins.

Nicolova, Ruselina. 2000. Über den Kontrastfokus mit Partikeln im Bulgarischen. In Gerhild Zybatow, Uwe Junghanns, Grit Mehlhorn & Luka Szucsich (eds.), *3. europäische Konferenz "Formale Beschreibung slavischer Sprachen, Leipzig 1999"* (Linguistische Arbeitsberichte 75), 103–115.

Reis, Marga. 2005. On the syntax of so-called focus particles in German: A reply to Büring and Hartmann 2001. *Natural Language & Linguistic Theory* 23(2). 459–483. DOI: 10.1007/s11049-004-0766-5.

Rooth, Mats. 1985. *Association with focus.* University of Massachusetts, Amherst. (Doctoral dissertation).

Rudin, Catherine. 1988. On multiple questions and multiple WH fronting. *Natural Language & Linguistic Theory* 6. 445–501. DOI: 10.1007/BF00134489.

Smeets, Liz & Michael Wagner. 2018. Reconstructing the syntax of focus operators. *Semantics and Pragmatics* 11. 6:1–30. DOI: 10.3765/sp.11.6.

Sudhoff, Stefan. 2010. *Focus particles in German: Syntax, prosody, and information structure* (Linguistik Aktuell/Linguistics Today 151). Amsterdam: John Benjamins.

Taglicht, Josef. 1984. *Message and emphasis: On focus and scope in English* (English language series 15). London & New York: Longman.

Tasseva-Kurktchieva, Mila & Stanley Dubinsky. 2018. On the NP/DP frontier: Bulgarian as a transitional case. In Steven L. Franks, Vrinda Chidambaram, Brian D. Joseph & Iliyana Krapova (eds.), *Katerino mome: Studies in Bulgarian morphosyntax in honor of Catherine Rudin*, 287–312. Bloomington: Slavica.

Tisheva, Yovka & Marina Dzhonova. 2003. Lexical markers on the information structure level. In David Birnbaum (ed.), *Computational approaches to the study of early and modern Slavic languages and texts: Proceeedings of the "Electronic description and edition of Slavic Sources" conference, 24–26 September 2002, Pomorie, Bulgaria*, 1–20. Sofia: Boyan Penev Publishing Center.

Titov, Elena. 2020. Optionality of movement. *Syntax* 23(4). 347–374. DOI: 10.1111/synt.12202.

Tomaszewicz, Barbara. 2013. *Aż/čak*: The scalar opposite of scalar *only*. In Uwe Junghanns, Dorothee Fehrmann, Denisa Lenertová & Hagen Pitsch (eds.), *Formal description of Slavic languages: The ninth conference. Proceedings of FDSL 9, Göttingen 2011*, 301–323. Frankfurt am Main: Peter Lang.

von Stechow, Arnim. 2008. Current issues in the theory of focus. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantik/Semantics* (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science 6), 804–825. Berlin: De Gruyter Mouton. DOI: 10.1515/ 9783110126969.

Zanon, Ksenia. 2023. Focus association with *only* in Russian. In Vera Gribanova, Sabrina Grimberg, Erika Petersen, Eva Portelance & Brandon Waldon (eds.), *Annual workshop on Formal approaches to Slavic linguistics: The Stanford meeting 2018*, 418–437. Ann Arbor: Michigan Slavic Publications. DOI: 10.17863/ CAM.101872.

Chapter 21

The Western South Slavic verbal suffix *-nV/-ne* is a diminutive affix with a theme vowel

Ema Štarkl[a], Marko Simonović[b], Stefan Milosavljević[b] & Boban Arsenijević[b]

[a]University of Nova Gorica, ZRC SAZU, University of Ljubljana [b]University of Graz

The paper proposes a novel analysis of the sequence *-nV/-ne* in Western South Slavic (*-nu/-ne* in BCMS and *-ni/-ne* in Slovenian) as a complex morpheme consisting of the diminutive suffix *-n* and the theme vowel ∅/*e*, whereby the latter realizes the verbal category, like all other verbal themes in Slavic. We argue that the vowel in the suffix *-nV* is a floating vowel that surfaces when it helps optimize the syllable structure. While analyses of *-nV/-ne* as a complex morpheme have been proposed in the literature, the analysis in terms of diminution enables us to account for the peculiar status of the suffix *-nV* among other verbal suffixes, especially its compatibility with other suffixes, including diminutive and secondary imperfectivizing ones, which is either ignored or left unexplained in the previous accounts.

1 Introduction

In this paper, we offer a novel analysis of verbs with the suffix *-nV/-ne* in Western South Slavic, specifically in Bosnian/Croatian/Montenegrin/Serbian (BCMS) and Slovenian, illustrated in (1) and (2), respectively. Our focus is on perfective verbs, as in (1a, 2a), since only they are productive in both languages, although *-nV* is also found in a small number of imperfective degree achievements (DAs), as in (1b, 2b). We propose that *-nV/-ne* in Western South Slavic is complex and

consists of the morpheme $-n^u/-n^i$ (with a floating vowel) and the theme vowel $\emptyset/e$. The morpheme $-n^u/-n^i$ spells out a diminutive feature and the theme vowel spells out the verbal category feature, just like all other theme vowels in Slavic languages (for the latter, see also Svenonius 2004, Biskup 2019, Simonović et al. 2023, Milosavljević & Arsenijević 2022, Kovačević et al. 2024).

(1) a. trep-nu-ti   b. to(n)-nu-ti   (BCMS)
         blink-nV-INF   sink-nV-INF
         'blink'   'sink'

(2) a. mežik-ni-ti   b. to(n)-ni-ti   (Slovenian)
         blink-nV-INF   sink-nV-INF
         'blink'   'sink'

Morphological forms of *-nV/-ne-* and $\emptyset/e$-verbs that will be relevant for our analysis are summarized in Tables 1 and 2, respectively.

Table 1: Conjugation of *-nV/-ne* verbs

Language	INF	PRS.1.PL	PASS.PTCP	Translation
BCMS	gur-nu-ti	gur-ne-mo	gur-nu-t	'push'
Slovenian	pah-ni-ti	pah-ne-mo	pah-nj-en	'push'

Table 2: Conjugation of $\emptyset/e$ verbs

Language	INF	PRS.1.PL	PASS.PTCP	Translation
BCMS	ves-$\emptyset$-ti	vez-e-mo	vez-$\emptyset$-en	'embroider'
Slovenian	ves-$\emptyset$-ti	vez-e-mo	vez-$\emptyset$-en	'embroider'

The paper is organized as follows. In the remainder of this section, we overview different classes of Slavic verbs derived by *-nV/-ne* that are recognized in the literature. In Section 2, we critically assess previous accounts of *-nV/-ne*. Section 3 provides a quantitative description of this sequence in BCMS and Slovenian. In Section 4, we provide morpho-phonological arguments for our segmentation of *-nV/-ne* into the morpheme $-n^u/-n^i$ and the theme vowel $\emptyset/e$.

Our syntactic and semantic analysis of *-n^u/-n^i* as a diminutive suffix that combines with the verbal category (whose exponent is $\emptyset/e$) is provided in Section 5. Section 6 concludes the paper.

Several verb classes derived by *-nV* have been recognized in the literature on Slavic languages. The most typical and the most productive class comprises SEMELFACTIVES, illustrated in (3a) and (4a) for BCMS and Slovenian, respectively. Semelfactives are usually defined as "instantaneous" actions in the classical sense of Smith (1997), and in most formal approaches this is the only identified class of perfective *-nV/-ne* verbs (see e.g. Łazorczyk 2010 and Kwapiszewski 2020, 2022 for Polish, Wiland 2019 for Czech and Polish, Biskup 2023a for Russian and Czech, etc.). In analyses couched in the framework of Cognitive Linguistics, this class of verbs is usually referred to as Single Act Perfectives (see e.g. Janda 2007, Dickey & Janda 2009, Makarova & Janda 2009, Kuznetsova & Makarova 2012, Nesset 2013, Sokolova 2015 for Russian, Nesset 2012 for Old Church Slavonic, Bacz 2012 for Polish).

(3)   a.   mah-nu-ti   b.   mah-a-ti   (BCMS)
            wave-nV-INF   wave-TV-INF
            'wave once'   'wave repeatedly'

(4)   a.   mah-ni-ti   b.   mah-a-ti   (Slovenian)
            wave-nV-INF   wave-TV-INF
            'wave once'   'wave repeatedly'

DEGREE ACHIEVEMENTS, illustrated for BCMS and Slovenian in (1b) and (2b) above, are a small class of imperfective verbs derived by *-nV/-ne* (see e.g. Taraldsen Medová & Wiland 2019 for a formal analysis of this class in Czech and Polish). Degree achievements derived by *-nV/-ne* are no longer a productive class across Slavic languages, which is why they will be set aside in the present paper (though we briefly return to them in Section 3.2).[1] NATURAL PERFECTIVES

---

[1] An anonymous reviewer raises the question of whether imperfective *-nV/-ne* verbs should be analyzed on a par with perfective *-nV/-ne* verbs. Since we focus only on perfective *-nV/-ne* verbs in the paper (as only perfectives are productive in contemporary BCMS and Slovenian), we do not delve deeper into the debate on whether *-nV/-ne* in perfectives and imperfectives should be treated as a unified item. Note, however, that once semelfactives and degree achievements are analyzed as sharing the same semantic core based on atomicity (cf. Rothstein 2008a,b), there might be a semantic justification for treating these two *-nV/-ne* classes as containing the same suffix; see also Taraldsen Medová & Wiland (2019) (presented in Section 2.2.2) for a unified syntactic analysis of this suffix within a Nanosyntactic framework.

are *-nV* verbs that function as lexicalized perfective counterparts of simplex imperfective verbs (e.g. Bacz 2012 for Polish, Sokolova 2015 for Russian). It should be immediately clear that there is no clear-cut boundary between these verbs and "proper" semelfactives, since semelfactives act as aspectual counterparts of iterative verbs, as in (3) and (4). Finally, Sokolova (2015) identifies a class of (PERFECTIVE) DELIMITATIVES in Russian, which are *-nV/-ne* verbs that can combine with durative adverbials indicating a short duration, as in (5). Similar examples are available in BCMS, as evidenced by example (6), whereas in Slovenian this use of *-nV* is not attested.

(5) Ja let-nu-l   2 časa.   (Russian, from Sokolova 2015; our translation)
     I   fly-nV-PST 2 hours
     'I flew for two hours [I took a short two-hour flight].'

(6) Drem-nu-o sam   par   minuta.   (BCMS)
     doze-nV-PST AUX.1.SG couple minutes
     'I dozed for a few minutes.'

2 Previous analyses of *-nV/-ne*

In this section, we overview previous analyses of the sequence *-nV/-ne.* We first briefly comment on traditional approaches to this item in Slavic in Section 2.1, after which we provide a detailed discussion of previous formal analyses of *-nV/-ne* in Section 2.2.

2.1 Traditional approaches to *-nV* verbs in Slavic

In traditional descriptions, *-nV/-ne* is typically analyzed as a monomorphemic theme vowel (TV) defining its own conjugation class (for BCMS, see e.g. Barić et al. 1997: 235, Ivšić et al. 1970: 253, Stevanović 1986: 331, Stanojčić & Popović 2008; for Slovenian, cf. e.g. Breznik 1934: 116, 124, Toporišič 2000: 364, Vidovič Muha 2011: 64; a similar point is made for Russian in Gladney 2013 and references therein). The alternative analysis, whereby *-n* is a separate morpheme and *V/e* is a theme vowel, is usually discarded on the grounds that there is no independently motivated TV class defined by the vowels following *-n* (i.e. *i/e* in Slovenian and *u/e* in BCMS).

2.2 Previous formal approaches to *-nV* verbs in Slavic

In this subsection, we discuss previous formal approaches to *-nV/-ne*, grouping them into those that analyze this segment as a single morpheme – let us label

them MONOMORPHEMIC ANALYSES (Section 2.2.1), and those arguing that *-nV/-ne* is decomposable into a suffix and a theme vowel – BIMORPHEMIC ANALYSES (Section 2.2.2).

2.2.1 Monomorphemic analyses

Schoorlemmer (2004) analyzes *-nu/-ne* in Russian as a lexical marker of perfectivity, which is one of the two basic ways of how perfectivity arises in her approach (the other way being compositionally, through telicity, as in the case of prefixed perfective verbs; see also Schoorlemmer 1997). According to Schoorlemmer, the "lexical" (i.e. non-compositional) status of perfective verbs derived by this suffix is confirmed by the fact that (in Russian) they do not derive secondary imperfectives, unlike (prefixed) telic predicates (accomplishments and achievements). For Borer (2005a,b), *-nu/-ne*, just as prefixes, assigns quantity to a verbal predicate, hence it is generated in the domain of inner aspect (Borer analyzes Slavic perfectivity as quantity, i.e. telicity).

An open question for both Schoorlemmer and Borer is the complementary distribution of *-nV/-ne* and (other) theme vowels. In addition, the complementary distribution with the secondary imperfectivizing suffix argued for in Schoorlemmer (2004) cannot be extended to all Slavic languages, as we show in this paper. This means either that Slavic languages vary in this respect, or that this combination is blocked due to some morphological constraint (as hinted at in Borer 2005b for Russian and Kwapiszewski 2022: 236 for Polish), or some kind of semantic incompatibility of the two suffixes is at stake (e.g. Jabłońska 2007 for Polish, Biskup 2023a for Czech); see Kwapiszewski (2022: 235–236) for a recent critical assessment of both semantic and morphological constraints.

Progovac (2005) also analyzes *-nV/-ne* as an aspectual marker (in BCMS), but she claims that it is generated in the domain of grammatical (outer) aspect. More precisely, she proposes that this suffix denotes existential quantification in the outer AspP, where it encodes "a single event, or, more precisely, at least one event" (Progovac 2005: 109). For instance, according to Progovac, the verb in (7a) has the interpretation as in (7b). She substantiates her analysis of *-nV/-ne* as bearing an existential feature with the fact that verbs with *-nu/-ne* are easily modifiable with the adverbial *jedanput* 'once, one time', which she analyzes as an existential quantifier. For Progovac, further support for the analysis of *-nu/-ne* as a marker of outer aspect comes from its complementary distribution with secondary imperfectivizing suffixes (7c), which in her analysis are markers of grammatical (outer) aspect that bear the feature of universal quantification. The

incompatibility of these two suffixes follows straightforwardly if they check their quantificational features in the same projection.

(7) a. Stefan   je (jedanput) kuc-nu-o   na prozor.
       Stefan.NOM AUX once   knock-nV-PTCP on window.ACC
       'Stefan knocked (at least once) on the window.'
       'There was (at least) one time that Stefan knocked on the window.'

   b. There was some/at least one occasion X for which it is true that
      Stefan knocked on the window on that occasion X.

   c. * kuc-nu-va-ti
        knock-SG-SI-INF

However, the compatibility with existential quantifiers such as *jedanput* 'once' can hardly be taken as evidence that *-nV/-ne* bears the existential feature, since such adverbials are also compatible with imperfective verbs, as well as other types of perfective verbs, and not only with semelfactives (see Milosavljević 2019 for an extensive corpus analysis of these adverbials). When it comes to the compatibility with secondary imperfectivizing suffixes, the reasoning outlined above regarding the proposal in Schoorlemmer (2004) applies to Progovac's analysis as well.

According to Svenonius (2004) and Biskup (2023a,b, 2024), *-nV/-ne* is a verbalizer in Slavic languages (Russian and Czech, respectively). This claim is supported by its complementary distribution with theme vowels, which are analyzed as verbalizers in these works. In addition to its verbalizing role, this suffix also has a perfectivizing effect, i.e. it bears a perfective feature. A question that arises under this family of approaches is why *-nV/-ne* is the only verbalizer with a perfective feature.

Kwapiszewski (2020), working within the framework of Distributed Morphology, analyzes *-nV/-ne* in Polish as an exponent of a complex head realizing (fused) verbal and quantity features. This analysis is based on the complementary distribution of *-nV/-ne* with both theme vowels (as verbalizers) and secondary imperfectivizing suffixes in Polish. In a more recent work, Kwapiszewski (2022: 231–237) refines his proposal of the semelfactive *-nV/-ne* in Polish by arguing that this suffix is an exponent of a complex head comprising the verbal category head (more precisely, v_{DO}, given the unergative or transitive nature of the relevant verbs), the Voice head, and an aspectual perfective head (he maintains the claim that *-nV/-ne* is in complementary distribution with both verbalizing and secondary imperfectivizing suffixes). While Kwapiszewski's approach captures the "dual" behavior of this morpheme (verbalization + quantity/perfectivity) and

explains its complementary distribution with theme vowels, his approach, being based on morphological operations specific to Polish, cannot be generalized to other Slavic languages since, as discussed above, this suffix is not in complementary distribution with secondary imperfectivizing suffixes in at least some languages.

Arsenijević (2006) proposes that *-nV/-ne* in BCMS is a diminutive suffix. Specifically, it introduces some bounded quantity to the interpretation of the eventuality, which is a relatively small part of a larger quantity of the same eventuality. In other words, *-nV/-ne* marks a division into atomic units for the relevant eventuality. Arsenijević provides examples similar in spirit to the delimitative uses of *-nu/-ne* illustrated in (5–6) above, and offers the following explanation:

> The atomic temporal interval appears as the natural interpretation when the description of an eventuality does not provide any unit of division, but division must still be applied. The natural solution is to take the atomic temporal interval as corresponding to the smallest possible quantity of the eventuality. The atomic interval also provides a partitive interpretation, when related to the mass from which it selects a unit. (Arsenijević 2006: 219)

Syntactically, according to Arsenijević, *-nV/-ne* is the head of the VP, and marks the presence of a telic template in cases where the description of the eventuality does not define one. As an argument for this position, Arsenijević lists the incompatibility of *-nV/-ne* with internal prefixes, as these morphemes also license telicity. However, *-nV/-ne* can be combined with (internal) prefixes, as will be shown in Section 3.2 for BCMS and Slovenian (see also Nordrum 2019 for such combinations in Russian, as well as Kwapiszewski 2020 for Polish).

The presented description of the diminutive semantics of the suffix *-nV* closely matches the notion of singularity. In fact, in the semantic approach of Kagan (2008, 2010), both prefixes and the semelfactive suffix *-nV/-ne* in Russian license singularity, but unlike prefixes, which bring additional meaning and/or argument structure effects, "the suffix -nu seems to introduce no further changes except for the singularity restriction. It takes an imperfective activity predicate and renders a perfective predicate whose denotation contains only the smallest instantiation of this activity, each of which has no proper part which instantiates the same type of event" (Kagan 2010: 11); see also Milosavljević (2023b) for a syntactic implementation of this idea.

Relatedly, according to Armoškaitė & Sherkina-Lieber (2008), the semelfactive suffix *-nV/-ne* and the secondary imperfectivizing suffix *-yva* in Russian are markers of number in the verbal domain, licensing singularity and pluractionality, respectively, and thus occupy the same syntactic slot. This is supported by their

complementary distribution in Russian. Armoškaitė & Sherkina-Lieber propose that these suffixes, as markers of verbal number, are modifiers, and not heads, contrary to what we find in the nominal domain. The arguments for a modifier analysis are the following. Heads are obligatory, modifiers are not: e.g., on nouns, the number markers are heads, and since they are obligatory, there are no nouns that are neutral with respect to number. Further, on nouns, number marking, as a head, applies even when the number information is redundant, e.g. in the presence of numerals. Finally, number as a head on nouns triggers agreement on dependent constituents, e.g. Subject-Verb Agreement. However, the status of number in the verbal and nominal domains can be shown not to be as different as proposed in Armoškaitė & Sherkina-Lieber (2008). On the contrary, the motivation for seeing Slavic perfectivity as singularity and imperfectivity as plurality in the verbal domain (proposed in Kagan 2008, 2010; for related approaches, see Arsenijević 2023, Milosavljević 2022, 2023a,b) is argued to rely on compelling parallels between nominal and verbal domains: plural (imperfective) is unspecified for number, while singular (perfective) is the only marked/specified category (in the sense of Sauerland 2003). In that sense, all nouns and all verbs are either unspecified for number (if plural, i.e. imperfective) or specified as singular, i.e. perfective. In other words, the absence of the suffix *-nV/-ne* does not imply the absence of singularity, as there can be another way of realizing it (e.g. via Spec–Head agreement in the case of prefixation, see Milosavljević 2023a,b), licensing a view in which it is not optional. Additionally, verbs suffixed with *-nV/-ne*, just as nouns, appear in the context of numerals, i.e. with the count adverbials like *once* (cf. Progovac 2005).[2]

Markman (2008) analyzes both the semelfactive suffix *-nV/-ne* and the secondary imperfectivizing suffix *-iv* in Russian as exponents of a single *v*P-selecting light verb *v* (in the sense of Diesing 1998), which denotes an atelic event and is merged above lexical prefixes. The light verb is spelled out as *-nV/-ne* when [+Instantaneous] and as a secondary imperfectivizing suffix when [+Progressive] or [+Habitual]. Markman follows Smith (1997) in assuming that semelfactives are perfective atelic predicates. The single-head approach to the two suffixes is based on the claim that they are in complementary distribution in Russian, whereas their status as light verbs is motivated by similar behavior to

---

[2]On a broader scale, there seems to be a tight cross-linguistic connection between diminutives and singulatives (cf. e.g. Rijkhoff 1991; Mathieu 2012: §4, and references therein), and more generally a link between diminutives and atomicity (see also Wiltschko 2006, De Belder 2008, 2011, Ott 2011). We contend that the link between diminution and singularity reflected through the same morphemes cross-linguistically is due to the fact that they share atomicity as a semantic core.

light verbs cross-linguistically. A potential problem for Markman (2008), apart from the issue of complementary distribution with secondary imperfectivizing suffixes discussed above for other approaches, concerns the analysis of semelfactives as atelic predicates. In this paper, we argue that semelfactives are singular telic predicates, like other traditional perfective verbs (see also Rothstein 2008a,b for an analysis of semelfactives as telic predicates in Russian).

In the next subsection, we turn to bimorphemic analyses of *-nV/-ne*.

2.2.2 Bimorphemic analyses

Łazorczyk (2010) treats *-nV/-ne* in Polish as composed of two morphemes: the suffix *-n* as a marker of semelfactivity (deriving also a small number of degree achievements), and a theme vowel, which in her approach is a reflex of verbalization through the structure (in the sense of Borer 2005b), hence inserted once the inner aspect has been projected (since the root is categorized as a verb in the context of inner aspect). Łazorczyk (2010), however, does not elaborate her approach in any detail.

Taraldsen Medová & Wiland (2019) and Wiland (2019), analyzing *-nV/-ne* in Czech within the framework of Nanosyntax (cf. Caha 2009, Starke 2009), propose that *-n* is a light verb, whereas the vocalic segment is a theme vowel. In their approach, roots, *-n* and the theme vowel can all spell out syntactic structures of different sizes (i.e. of varying syntactic complexity), with the relevant containment relations in syntax specified as in (8).

(8) a. containment of the light verbs:
        GIVE > GET
     b. containment of the lexical categories:
        verb > noun > adjective
     c. argument structure hierarchy:
        unergative > accusative > unaccusative

In semelfactives, the root is nominal, *-n* spells out the light verb GIVE, and the theme vowel spells out the accusative or unergative structure. In degree achievements, the root is adjectival, *-n* spells out the light verb GET, and the theme vowel spells out unaccusative syntax. The relation between semelfactives and degree achievements (hence also *-n* and the theme vowel in semelfactives vs. degree achievements) is regulated by the Superset Principle. According to this principle, a phonological exponent of a lexical item is inserted into a syntactic node if its lexical entry has a (sub-)constituent which matches that node. Where several items

meet the conditions for insertion, the item containing fewer features unspecified in the node must be chosen (Starke 2009). Given the containment relations in (8), the light verb component in both semelfactives and degree achievements can be spelled out as *-n*.

One problem with this approach concerns the fact that it is extremely difficult to isolate nominal, adjectival, or verbal roots per se, since the same root may be categorized as a noun, verb or an adjective, depending on the categorizing morpheme and/or syntactic context. Further, this approach does not cover the full range of uses of the suffix *-nV/-ne*, which easily combines also with verbal bases, and even with other suffixes (e.g. *bol-uc-nu-ti* 'hurt a bit', where *-uc* is a diminutive suffix).

3 Quantitative description of *-nV* verbs in BCMS and Slovenian

In this section we first describe our quantitative database in Section 3.1 and then present the quantitative data on the sequence *-nV/-ne* in Section 3.2. In Section 3.3 we summarize the discussion and findings so far to prepare the ground for our morpho-phonological (Section 4) and syntactic/semantic analysis (Section 5).

3.1 Our empirical source: *WeSoSlaV*

Our proposal is informed by quantitative insights from the *Annotated Database of the Western South Slavic Verbal System* (*WeSoSlaV*, Marušič et al. 2022, Milosavljević et al. 2023, Arsenijević et al. 2024). The database consists of 5300 BCMS and 3000 Slovenian verbs retrieved from the *srWaC, hrWaC, bsWaC* and *meWaC* corpora for BCMS (Ljubešić & Klubička 2014) and from *Gigafida*, the Slovenian National Corpus for Slovenian (Logar-Berginc et al. 2012). The verbs are selected based on frequency: the top 3000 highest frequency verbs from each of the corpora are included and annotated. As *srWaC, hrWaC, bsWaC* and *meWaC* are corpora of different BCMS varieties, the BCMS database contains the union of the 3000-verb lists from the four corpora.

Each verb is annotated for a fixed set of over 40 different properties, including grammatical aspect, the characteristic morphemes (the root, prefixes, suffixes), their special properties (e.g. root allomorphy), deverbal nominalizations, prosodic prominence, TVs and others. Our analysis is mainly based on the derivation subpart of WeSoSlaV (Milosavljević et al. 2023) and an additional *-nV*-verb subpart annotated for the purposes of this paper (Štarkl et al. 2024).

3.2 *-nV* verbs: the quantitative data

In this section, we present quantitative data on the aspectual properties of verbs formed with *-nV/-ne* in both BCMS and Slovenian. We start with the correlation between (im)perfectivity and the presence of a prefix, as summarized in Table 3.

Table 3: *-nV* verbs in *WeSoSlav*: prefixation and (im)perfectivity

-nV verbs in WeSoSlav	BCMS (258 in total, 4.87% of all the verbs in WeSoSlav)		Slovenian (143 in total, 4.77% of all the verbs in WeSoSlav)	
	Unprefixed	Prefixed	Unprefixed	Prefixed
All	91/258 (35.27%)	167/258 (64.73%)	24/143 (16.78%)	119/143 (83.22%)
Imperfective	9/258 (3.49%)	0 (0%)	3/143 (2.10%)	0 (0%)
Perfective	82/258 (31.78%)	167/258 (64.73%)	21/143 (14.69%)	119/143 (83.22%)

As is clear from the table, all prefixed verbs are perfective.[3] The very existence of prefixed *-nV/-ne* verbs is theoretically significant since it shows that *-nV/-ne* and prefixes can be combined, contrary to some approaches reviewed in Section 2 above.[4] Another important point that Table 3 makes salient is that the vast majority of unprefixed *-nV/-ne* verbs are perfective. Specifically, out of 91

---

[3]Out of 167 prefixed *-nV* verbs in BCMS, 95 (56.89%) combine with a perfective base, 23 (13.77%) combine with an imperfective base, while in 49 (29.34%) cases there is a bound base (i.e. a base that is not attested without a prefix). Out of the 119 prefixed *-nV* verbs in Slovenian, 43 (36.13%) combine with a perfective base, 13 (10.92%) combine with an imperfective base, while in 62 (52.1%) cases the base is bound.

[4]The majority of such prefixes are lexical/internal prefixes, e.g. in BCMS: *pod-met-nu-ti* [UNDER-put-nV-INF] 'set up, put under', *od-gur-nu-ti* [FROM-push-nV-INF] 'push away', *s-kliz-nu-ti* [OFF-glide-nV-INF] 'slip', *u-tis-nu-ti* [IN-press-nV-INF] 'press in', *iz-tis-nu-ti* [OUT-press-nV-INF] 'press out'. Although in our main database (WeSoSlaV) there are no typical examples with superlexical prefixes, such verbs are possible, especially in the presence of another prefix, which is expected given that the most typical superlexical prefixes stack on top of other prefixes. Some such examples, taken from Stojanović (2016), include: *iz-o-kre-nu-ti* [OUT-ABOUT-start-nV-INF] 'turn over all', *po-o-smeh-nu-ti* [OVER-ABOUT-laugh-nV-INF-REFL] 'laugh a little bit'. However, there are also superlexical-like prefixes, such as the attenuative *pri-*, which combine directly with *-nV* verbs, e.g. *pri-drem-nu-ti* [AT-doze-nV-INF] 'doze a little bit'. A similar picture is observed in Slovenian. An example of LP-prefixed nV-verbs is *iz-tis-ni-ti* [OUT-press-nV-INF] 'press out', whereas *po-na-tis-ni-ti* [OVER-ON-press-nV-INF] 'reprint' illustrates SLPs.

unprefixed verbs in BCMS, 82 (90.11%) are perfective, and only 9 (9.89%) are imperfective. Similarly, out of 23 unprefixed verbs in Slovenian, 18 (78.26%) are perfective, and only 3 (13.04%) are imperfective.[5] These data, together with the fact that new verbs (including the ones with borrowed bases) are always perfective in BCMS and Slovenian (for the former, see also Simonović 2015), strongly indicate that only perfective *-nV/-ne* verbs are productive in the contemporary BCMS and Slovenian. The same has been observed also for other Slavic languages, e.g. Polish (Klimek-Jankowska et al. 2018), Czech (Taraldsen Medová & Wiland 2019, Wiland 2019), Russian (Sokolova 2015). This justifies our choice to focus on perfective verbs in this paper.

We now turn to the quantitative patterns of aspectual pairs *-nV/-ne* verbs participate in. Tables 4 and 5 summarize these patterns separately for prefixed and unprefixed verbs.[6]

We consider prefixed and unprefixed verbs separately to control for the possible influence of prefixation. For instance, on the one hand, Biskup (2023a) refers to Isačenko (1962) and Townsend (1968) for the claim that prefixed semelfactive verbs are not semelfactive anymore, i.e. they behave like any other prefixed perfective verb. On the other hand, Kwapiszewski (2022) indicates that the presence of a prefix in Polish does not change the fact that in that language *-nV/-ne* verbs cannot undergo secondary imperfectivization. For our purposes, two facts evident from Tables 4 and 5 are most significant. First, in the majority of cases, the imperfective aspectual counterpart is either a corresponding unsuffixed verb (i.e. a verb whose root is followed just by a theme vowel), or a verb with some kind of iterative suffix.[7] Second, there are both unprefixed and prefixed verbs that undergo secondary imperfectivization at the same time preserving the morpheme

---

[5]Out of 9 imperfective verbs in BCMS, 7 are degree achievements, and 2 are lexicalized states. Out of the 3 imperfective verbs in Slovenian, 1 is a degree achievement, and 2 are lexicalized states.

[6]The examples of the categories in the first column of this table are in BCMS. There are only 4 and 2 simple perfective verbs with an imperfective secondary imperfective counterpart preserving *-nV* in BCMS and Slovenian, respectively. The remaining three BCMS pairs from WeSoSlav are: *buk-nu-ti − buk-nj-iva-ti* 'erupt', *pla-nu-ti − pla-nj-ava-ti* 'burst into flames', and *ba-nu-ti − ba-nj-ava-ti* 'burst'. The Slovenian verbs and their imperfective counterparts are: *mi-ni-ti − mi-n-eva-ti* 'pass', and *ga-ni-ti − ga-nj-ati* 'move'.

[7]The suffixes *-t* and *-k* that derive diminutive-iterative verbs are traditionally listed as *-ka* and *-ta* in BCMS grammars (e.g., Stanojčić & Popović 2008). However, these suffixes can also be plausibly decomposed into the proper (diminutive-iterative) suffixes and theme vowels, specifically, *k* + TV *a/a* (e.g. *pip-k-a-ti* (INF), *pip-k-a-mo* (PRS.1.PL) 'touch'), and *t* + TV *a/je* (*trep-t-a-ti* (INF), *trep-ć-e-mo* < /trep-t-je-mo/ (PRS.1.PL) 'blink'). These two theme vowels (i.e., *a/a* and *a/je*) are two of the three most productive TVs in BCMS that are also found in secondary imperfectivizing suffixes (Simonović et al. 2023, Arsenijević et al. 2023).

Table 4: Imperfective counterparts of unprefixed perfective verbs

Simple PFV *-nV* verbs in *WeSoSlaV* with …	BCMS (N=82)		Slovenian (N=21)	
an IPFV root-TV counterpart (lup-nu-ti – lup-a-ti 'slap')	43	(52.44%)	10	(47.62%)
an IPFV *-t-* counterpart (trep-nu-ti – trep-ta-ti 'blink')	11	(13.41%)	0	(0%)
an IPFV *-k-* counterpart (tres-nu-ti – tres-ka-ti 'snap')	24	(29.27%)	0	(0%)
IPFV SI counterpart, without preserving *-nV* (crk-nu-ti – crk-ava-ti 'die')	8	(9.76%)	4	(19.05%)
an IPFV apophonical counterpart (mak-nu-ti – mit:c-a-ti 'move')	4	(4.88%)	1	(4.76%)
an IPFV SI counterpart, preserving *-nV* (sva-nu-ti – sva-nj-ava-ti 'dawn')	4	(4.88%)	2	(9.52%)

Table 5: Imperfective counterparts of prefixed perfective verbs

Prefixed PFV *-nV* verbs in *WeSoSlaV* with …	BCMS (N=167)		Slovenian (N=119)	
an IPFV *-t* counterpart	34	(20.36%)	0	(0%)
an IPFV *-k* counterpart	2	(1.20%)	7	(5.88%)
an IPFV *-p* counterpart	0	(0%)	6	(5.04%)
an IPFV SI counterpart, without preserving *-nV*	60	(35.93%)	56	(47.06%)
an IPFV apophonical counterpart	21	(12.57%)	20	(16.81%)
an IPFV SI counterpart, preserving *-nV*	55	(32.93%)	18	(15.13%)

-nV.[8] The first fact is important in the light of our analysis of verbs derived by *-nV/-ne* as diminutive counterparts of the verbal predicates denoted by the corresponding imperfective verbs, as argued in detail in Section 5. The other fact, i.e. the compatibility of *-nV/-ne* with secondary imperfectivizing suffixes in at least some verbs, corroborates our claim that the two suffixes are not in complementary distribution, contrary to much previous work (see Section 2).[9]

3.3 Towards an analysis

So far, we have overviewed previous approaches and presented our quantitative data. We have seen that existing analyses, both monomorphemic and bimorphemic, face both empirical and theoretical issues, at least when applied to Western South Slavic. On the empirical side, it was shown by our quantitative data that some central assumptions in the majority of previous approaches (e.g. complementary distribution of *nV/-ne* and secondary imperfectivizing suffixes) do not hold for all the verbs in Western South Slavic. As for the monomorphemic analyses, apart from the issues discussed in Section 2.2.1, we can add that analyzing *-nV* as a monomorphemic theme vowel leaves open the question of why, unlike all other themes, this theme vowel includes a (non-glide) consonant and is the only theme vowel across Slavic languages that performs a perfectivizing function. An analysis splitting *-nV/-ne* into *-n* as a separate morpheme and *u/ e* and *i/e* as a theme vowel in BCMS and Slovenian respectively lends itself as a solution. While a similar segmentation has already been proposed (see Section 2.2.2), the approaches are either not elaborated (Łazorczyk 2010), or do not cover all the empirical data (Taraldsen Medová & Wiland 2019, Wiland 2019). In the latter case, it is assumed that *-nV* combines with the nominal bases to derive perfective semelfactive (unergative) verbs (in Czech and Polish), but the same suffix, at least in Western South Slavic, also readily combines with verbal bases, and even with other suffixes, e.g. *bol-uc-nu-ti* 'hurt a bit', where *-uc* is a verbal diminutive suffix. In the following sections, we use this compatibility with other (diminutive) suffixes to argue that *-nV* is itself a diminutive suffix *-n^u* (BCMS)/*-n^i* (Slovenian), which selects the theme vowel ∅/*e.*

---

[8]In addition to the examples used as an illustration in Table 4, this pattern can be illustrated by the following prefixed verbs: *na-dah-nu-ti – na-dah-nj-ivati* 'inspire', *za-bezek-nu-ti – za-bezek-nj-iva-ti* 'bewilder' for BCMS; and *s-tr-ni-ti – s-tr-nj-eva-ti* 'sum up', *za-mrz-ni-ti – za-mrz-nj-eva-ti* 'freeze', *u-ki-ni-ti – u-ki-nj-a-ti* 'abolish, cancel', *raz-gr-ni-ti – raz-gri-nj-a-ti* 'unfold, spread out' for Slovenian.

[9]See also Milosavljević (2023b) for the discussion of secondary imperfective forms of semelfactive verbs in South-East Serbo-Croatian, where such forms are much more productive.

4 Morpho(-phono)logical analysis

In this section, we present morpho-phonological arguments for our main claim that the sequence *-nV/-ne* is composed of the suffix *-n^u* (BCMS)/*-n^i* (Slovenian), and the theme vowel ∅/*e*. The theme-vowel class ∅/*e* is independently attested with simple verbs in both BCMS and Slovenian (Arsenijević et al. 2024). The question remains how to treat the vowels *u* and *i* that appear next to the consonant *-n* when it is combined with the ∅-exponent of the theme vowel, but do not appear when it is combined with the *e*-exponent. We propose that the morpheme under consideration has both a consonantal and vocalic part, but that only the consonantal part is lexically affiliated with a timing slot, whereas the vocalic part is floating. This approach has already been applied to the Polish cognate of the same morpheme in Zdziebko (2017). We submit that the realization of the floating vowels is regulated by syllable structure constraints. Floating vowels surface in front of consonant-initial endings (helping to prevent consonant clusters) and they do not surface before vowel-initial endings (because realizing them would create a hiatus). This is illustrated in (9). Specifically, the floating vowel helps avoid the consonant clusters *nt* and *nl* in the infinitive and participle forms (9a, 9b). These clusters do not appear in the verbal systems of BCMS and Slovenian. On the other hand, the floating vowels are not realized before vowels *e* or *i* in the present tense and the imperative forms (9c, 9d) because in this case full (i.e. non-floating) segments already constitute optimal open syllables and the realization of the floating vowels would lead to a hiatus.

(9)   a. max-n^u-∅-ti → maxnuti, *maxnti   (BCMS)
          max-n^i-∅-ti → maxniti, *maxnti   (Slovenian)
          wave-nV-TV-INF

     b. max-n^u-∅-l-a → maxnula, *maxnla   (BCMS)
          max-n^i-∅-l-a → maxnila, *maxnla   (Slovenian)
          wave-nV-TV-PST-F

     c. max-n^u-*e*-mo → maxnemo, *maxnuemo   (BCMS)
          max-n^i-*e*-mo → maxnemo, *maxniemo (Slovenian)
          wave-nV-TV-PRS.1PL

     d. max-n^u-*i*-mo → maxnimo, *maxnuimo   (BCMS)
          max-n^i-*i*-mo → maxnimo, *maxniimo (Slovenian)
          wave-nV-TV.IMP-1.PL

An important argument for adding *-nV/-ne* verbs to the ∅/*e* class lies in the fact that the forms in (9) (as well as the rest of the paradigm) feature the endings

typical of ∅/*e* verbs in general. The only potential exception is constituted by passive participle forms, which we discuss below.

Before turning to the discussion of the passive participle forms, we need to address an alternative to adding floating vowels to the *n*-morpheme. The same surface result could have been achieved by assuming the *n*-morpheme just with a full consonant and adding the floating vowel to the representation of the theme vowel. In this case, the ∅/*e* class would become ^{u}/*e* in BCMS and ^{i}/*e* in Slovenian. This alternative account encounters an empirical problem, as it would predict the floating vowels to surface in all forms where consonant-final bases combine with consonant-initial endings, e.g. in *pad-∅-ti → pasti, *paduti, *paditi* 'fall.ɪɴғ' or *griz-∅-ti → gristi, *grizuti, *griziti* 'bite.ɪɴғ'.

As mentioned above, adding *-nV/-ne* verbs to the ∅/*e* class does appear to face some potential empirical issues. In both languages, the passive participle of *-nV/-ne* verbs diverges from most ∅/*e* verbs. Since BCMS and Slovenian differ at this point, we take a closer look at each language in the following two subsections.

4.1 BCMS

The regular passive participle suffix in the ∅/*e* conjugation in BCMS is *-en*, as illustrated in (10) by the verbs *ukrasti* 'steal' and *ugristi* 'bite'. Given the vowel-initial ending *-en*, for *-nu/-ne* verbs, we would expect the passive participle form ending in *-nen* (with non-realization of the floating vowel, just like in the present tense and in the imperative in 9c and 9d). However, the actual passive participles of these verbs end in *-nut*, as shown in (11) for the verb *dirnuti* 'touch'.

(10) a. ukrad̶-∅-l-a | ukrad-∅-en
          steal-ᴛᴠ-ᴘsᴛ-ғ | steal-ᴛᴠ-ᴘᴀss.ᴘᴛᴄᴘ

      b. ugriz-∅-l-a | ugriz-∅-en
          bite-ᴛᴠ-ᴘsᴛ-ғ | bite-ᴛᴠ-ᴘᴀss.ᴘᴛᴄᴘ

(11) dir-nu-∅-l-a | dir-nu-∅-t,   *dir-nu̶-∅-en
      touch-nV-ᴛᴠ-ᴘsᴛ-ғ | touch-nV-ᴛᴠ-ᴘᴀss.ᴘᴛᴄᴘ   touch-nV-ᴛᴠ-ᴘᴀss.ᴘᴛᴄᴘ

As it turns out, the ∅/*e* class is more heterogeneous than our initial overview reveals. If we zoom into verbs whose infinitival stems end in round vowels, we can find three roots that derive verbs with infinitives in *-uti*. These are illustrated in (12) by the forms of the verbs *obuti* 'put shoes on', *načuti* 'overhear' and *nasuti* 'pour'. As can be observed in (12), the passive participle form in such cases can end in *-t* for the first two verbs, and it obligatorily ends in *-t* for the third listed

verb. This indicates that *-nuti* verbs do not show atypical behavior with respect to other *-uti* verbs in the system. It can thus be submitted that the passive participle allomorph [-t] is conditioned by the adjacent [+round] feature (as one of its contexts of insertion).[10] Once this consonantal allomorph is selected, it comes as no surprise that [nu] surfaces as the exponent of n^u, since, as stated above, the *nt* cluster is blocked in the verbal forms in general.

(12)   a.   obu-∅-l-a | obuʋ-∅-en,
            put.shoes.on-TV-PST-F | put.shoes.on-TV-PASS.PTCP
            ⁇obu-∅-t
             put.shoes.on-TV-PASS.PTCP

       b.   nat͡ʃu-∅-l-a | nat͡ʃu-∅-t,   nat͡ʃuʋ-∅-en
            overhear-TV-PST-F | overhear-TV-PASS.PTCP overhear-TV-PASS.PTCP

       c.   nasu-∅-l-a | nasu-∅-t
            pour-TV-PST-F | pour-TV-PASS.PTCP

Based on the facts above, it is safe to conclude that the allomorph selection in passive participle forms of *-nV/-ne* verbs does not constitute an argument for excluding these verbs from the ∅/*e* theme-vowel class.

4.2 Slovenian

In Slovenian, just like in BCMS, the regular passive participle suffix in the ∅/*e* conjugation is *-en* (pronounced as [-ɛn] when under stress), as illustrated in (13) for the verbs *ukrasti* 'steal' and *gristi* 'bite'.[11] Here again, given the vowel-initial ending, we would expect passive participles derived from *-ni/-ne* verbs to end in *-nen*. However, the actual passive participles of these verbs end in *-njen*, as can be observed from (14).

(13)   a.   uˈkrad-∅-l-a | uˈkrad-∅-en
            steal-TV-PST-F | steal-TV-PASS.PTCP

       b.   ˈgriz-∅-l-a | ˈgriz-∅-en
            bite-TV-PST-F | bite-TV-PASS.PTCP

(14)   napix-n^i-∅-en   → naˈpixnjen, *napixnɨen
       inflate-nV-TV-PASS.PTCP

---

[10]This allomorph shows up in several other environments in the classes ∅/*e* and *a/a*. As shown in Bešlin (2023), its conditioning is at least partially lexical.

[11]The contrast between open-mid vowels [ɛ, ɔ] and close-mid vowels [e, o] can only be observed in stressed syllables. In unstressed syllables, the neutralized mid vowels are traditionally transcribed as close-mid. For clarity, we mark stress in the examples in this subsection.

We suggest that the passive participle morpheme is actually ^{-j}en, with a floating j. This hypothesis is supported by the fact that in the $\emptyset/e$ class there are verbs (beyond $-ni/-ne$ verbs) where the passive participle suffix causes the palatalization of the preceding consonant. Such (admittedly rare) verbs are illustrated in (15). We propose that since both the $-n^i$ morpheme and the passive participle ending ^{-j}en have floating segments (which in addition have the same features), there is a *cumulative faithfulness effect* (Farris-Trimble 2008) strong enough to make the insertion of an additional timing slot and the realization of the [j] obligatory.[12]

(15) a. preˈnɛs-∅-l-a | preneˈʃ-∅-ɛn
          transfer-TV-PST-F | transfer-TV-PASS.PTCP

      b. preˈrast-∅-l-a | preˈraʃt͡ʃ-∅-en
          grow.overTV-PST-F | grow.over-TV-PASS.PTCP

After having provided morpho-phonological evidence for the decomposition of the sequence $-nV/-ne$ into the suffix proper (n^V) and the theme vowel $\emptyset/e$, we are now in a position to turn to our syntactic and semantic analysis of the suffix $-n$ as a diminutive suffix.

5 The syntactic-semantic analysis in terms of diminution

As already previewed, our analysis of the verbal suffix $-nV/-ne$ is bimorphemic. In this section, we focus on the proposed morpheme $-n^u$ (BCMS)/$-n^i$ (Slovenian), which we argue is a diminutive suffix. We start in Section 5.1 by showing the special status of $-nV$ among suffixes: its perfective nature, its possibility to participate in suffix stacking, and the theme vowel it combines with. In Section 5.2 we sketch some similarities in the diminution of verbs and nouns that will be important for our analysis of the suffix $-nV$. Our syntactic modeling and formal semantic description are provided in Section 5.3 and Section 5.4, respectively. Section 5.5 brings a discussion on how the suffix $-nV$ fits the broader picture of suffixes in Western South Slavic. Finally, in Section 5.6 we compare our analysis to the previous approaches to the suffix $-nV$ and outline the advantages of our analysis.

---

[12]The palatalization in passive participles in $^{(-j)}en$ is at least partially lexically determined in Slovenian. This has been discussed for the i/i class in Toporišič (2000). The i/i class features triplets like *ponuditi* 'offer', *začuditi* 'bewilder', *prisoditi* 'attribute', whose passive participles are *ponujen/ponuden*, *začuden* and *prisojen*, respectively ([j] being derived from /dj/). Note that in the i/i class palatalization is much more common than in the $\emptyset/e$ class. This is expected on our account because in the former class both the original theme vowel (i) and the morpheme ^{-j}en favor palatalization.

5.1 Special status of *-nV* among suffixes

The first important property that sets the suffix *-nV* apart from all other verbal suffixes in BCMS and Slovenian concerns its aspectual effects. Specifically, all other verbal suffixes in BCMS and Slovenian derive verbs that pass tests for imperfectivity and atelicity. This is evidenced in (16a) and (17a) by the compatibility of BCMS *ova/uje-* and *ava*-verbs with the phasal verb *početi* 'begin', as well as by their combinability with durative adverbials (16b, 17b). The suffix *-nV*, by contrast, derives verbs that systematically fail both these tests, as illustrated in (18). In other words, the suffix *-nV* derives only perfective/telic verbs.

(16) a. Jan je   počeo da   štrajk-uj-e.   (BCMS)
          Jan AUX begun COMP strike-SUFF-PRS.3.SG
          'Jan began to strike.'

      b. Jan štrajk-uj-e   dva sata.
          Jan strike-SUFF-PRS.3.SG two hours
          'Jan has been striking for two hours.'

(17) a. Ovas je   počeo da   stas-av-a.   (BCMS)
          oat   AUX begun COMP grow-SUFF-PRS.3.SG
          'Oat began to mature.'

      b. Ovas stas-av-a   dva dana.
          oat grow-SUFF-PRS.3.SG two days
          'Oat has been maturing for two days.'

(18) a. *Jan je   počeo da   vik-n-e.   (BCMS)
          Jan AUX begun COMP shout-SUFF-PRS.3.SG
          Intended: 'Jan began to shout.'

      b. *Jan vik-n-e   dva sata.
          Jan shout-SUFF-PRS.3.SG two hours
          Intended: 'Jan has been shouting for two hours.'

Another important property of the suffix *-nV* is its possibility to license the stacking of other verbal suffixes on top of it, unlike most other suffixes, as illustrated by the contrast between (19a) and (19b) on the one hand, and (19c) on the other. Except for *-nV*, the only suffixes that allow stacking of suffixes on top of them are other diminutive suffixes (with which *-nV* forms a natural class), such as *-k* in BCMS (19d), or *-lj* in Slovenian (19e) (as well as some of the suffixes which integrate borrowed verbs).

(19) a. *Jan je   štrajk-ov-av-a-o.   (BCMS)
          Jan AUX strike-SUFF-SUFF-TV-PST.M

 b. * Ovas je stas-av-av-a-o.   (BCMS)
    oat AUX grow-SUFF-SUFF-TV-PST.M

 c. Dan je sva-n^u-av-a-o   [svaɲavao]. (BCMS)
   day AUX dawn-SUFF-SUFF-TV-PST.M
   'The day was dawning.'

 d. Pera je za-pit-k-iv-a-o   Lazu.   (BCMS)
   P AUX PREF-ask-SUFF-SUFF-TV-PST.M L.
   'Pera was asking Laza questions.'

 e. Jan je rez-lj-av-a-l   les.   (Slovenian)
   Jan AUX carve-SUFF-SUFF-TV-PST.M wood
   'Jan was carving out wood.'

The final unique property of the suffix *-nV* concerns theme vowel selection. Specifically, all Western South Slavic verbal suffixes take a theme vowel combination which includes the theme *-a* (i.e. *a/a* or *a/je*), as illustrated in (20a–20d), whereas only *-nV* combines with the theme vowel ∅/*e*, as in (20e, 20f).

(20) a. Marija je gril-ov-a-l-a   povrće.   (BCMS)
    M AUX grill-SUFF-TV-PST-F vegetables
    'Marija was grilling the vegetables.'

  b. Marija je pre-poruč-iv-a-l-a   povrće.   (BCMS)
    M AUX PREF-message-SUFF-TV-PST-F vegetables
    'Marija was recommending the vegetables.'

  c. Marija je gril-uc-k-a-l-a   povrće.   (BCMS)
    M AUX grill-SUFF-SUFF-TV-PST-F vegetables
    'Marija was grilling the vegetables a little bit.'

  d. Marija je marin-ir-a-l-a   povrće.   (BCMS)
    M AUX marinate-SUFF-TV-PST-F vegetables
    'Marija was marinating the vegetables.'

  e. Marija je gril-nu-∅-l-a   povrće.   (BCMS)
    M AUX grill-SUFF-TV-PST-F vegetables
    'Marija grilled the vegetables a little bit.'

  f. Marija je ob(-)r-ni-∅-l-a   kos zelenjave. (Slovenian)
    M AUX (PREF)-turn-SUFF-TV-PST-F piece vegetable
    'Marija turned a piece of vegetables.'

In the following sections, we argue that the special status of *-nV* among other verbal suffixes stems from its diminutive nature.

5.2 Diminution in verbs and nouns, similarities

Diminution is a cross-categorial phenomenon: nouns, verbs and adjectives all undergo this operation, in quite parallel ways. Consider the two structural positions for the diminutive suffix illustrated below for nouns (21a), adjectives (21b) and verbs (21c), respectively.

(21)   a.   i.   lav   (BCMS)
                  lion
                  'lion'

            ii.   lav-ić   lav-č-e   lav-č-ić
                  lion-DIM lion-DIM-INFL lion-DIM-DIM
                  'little lion'

      b.   i.   smeđ-e
                  brown
                  'brown'

            ii.   smeđ-ast-o   smeđ-(i)k-av-o
                  brown-DIM.ADJ-INFL brown-DIM-ADJ-INFL
                  smeđ-(i)k-ast-o
                  brown-DIM-DIM.ADJ-INFL
                  'somewhat brown'

      c.   i.   greb-a-ti
                  scratch-TV-INF
                  'scratch'

            ii.   greb-k-a-ti   greb-uc-a-ti   greb-uc-k-a-ti
                  scratch-DIM-TV-INF scratch-DIM-TV-INF scratch-DIM-DIM-TV-INF
                  'scratch a little'

The illustrated patterns perfectly fit De Belder et al.'s (2014) analysis of diminution, where diminutive suffixes may be base-generated at the level of the root or at the level of the category. This is schematically represented in Figures 1–3, where the maximal structure is given for each of the three categories for the examples in (21). In all three examples, the higher diminutive is fused with the category, i.e. the diminutive suffix in this position realizes both the diminutive and the category, and can be substituted by a suffix realizing only the category. The lower diminutive, by contrast, is merged directly with the root, before the entire (extended root) structure is categorized. Diminution can be realized by either of the two options, or by a combination, without a (necessary) effect of accumulation.

nP
-∅: [n] / -ć: [n][dim]   n′
                          -č : [dim]   √LAV : 'lion'
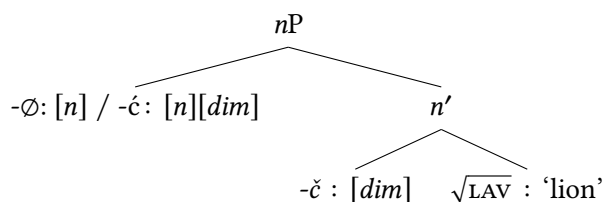
Figure 1: Syntactic representation of (double) diminutive nouns

aP
-av : [a] / -ast : [a][dim]   a′
                          -k : [dim]   √SMEĐ : 'brown'
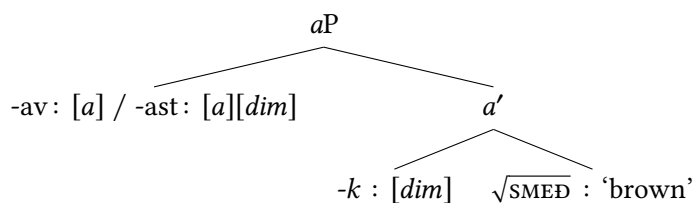
Figure 2: Syntactic representation of (double) diminutive adjectives

The suffix *-nV* is one of the suffixes used for diminution in the verbal domain. Apart from about a dozen exceptions, mostly degree achievements, as in (22), all *-nV* verbs involve the component of a small quantity, as in (23).

(22) to-nu-ti   tru-nu-ti   bri-nu-ti   sva-nu-ti
      √SINK-nV-INF   √ROT-nV-INF   √WORRY-nV-INF   √DAWN-nV-INF
      'sink'   'rot'   'worry'   'dawn'

(23) greb-nu-ti   spav-nu-ti   skok-nu-ti   kuc-nu-ti
      √SCRATCH-nV-INF √SLEEP-nV-INF   √JUMP-nV-INF   √KNOCK-nV-INF
      'scratch a little' 'sleep a little' 'jump a little' 'knock a little'

The suffix *-nV* with the diminutive interpretation normally can be combined with the root-level verbal diminutive suffix *-uc* in BCMS. When this is degraded, there

vP
-∅ : [v] / -k : [v][dim]   v′
                          -uc : [dim]   √GREB : 'scratch'
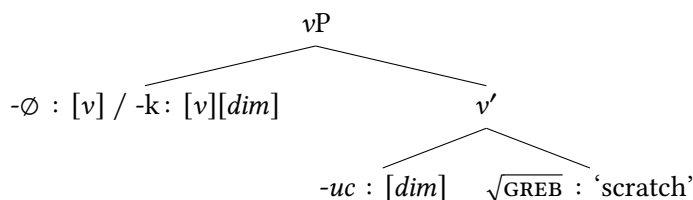
Figure 3: Syntactic representation of (double) diminutive verbs

typically is an independent reason, such as with the verb *kucnuti* in (24), where either the stem already involves the suffix *-uc* (so it is actually impossible to have *-nV* without *-uc*), or some process akin to haplology is at play. With the addition of *-uc*, the meaning is not affected, although sometimes the diminutive semantics feels somewhat stronger (which may be a pragmatic effect).

(24)   greb-uc-nu-ti   spav-uc-nu-ti   (BCMS)
       √SCRATCH-DIM-nV-INF   √SLEEP-DIM-nV-INF
       'scratch a little'   'sleep a little'

       prd-uc-nu-ti   ??kuc-uc-nu-ti
       √FART-DIM-nV-INF   √KNOCK-DIM-nV-INF
       'fart a little'   'knock a little'

All this points in the direction of having *-nV* as a suffix combining the verbal category with the diminutive component in the category head.

   Unlike in BCMS, in Slovenian, the suffix *-nV* does not combine with other diminutive suffixes productively. Judging by the dictionary and corpus data, there is only one verb combining the diminutive suffix *-ic* and *-nV* in Slovenian.

(25)   stop-i-ti   stop-ic-a-ti   stop-ic-ni-∅-ti   (Slovenian)
       √STEP-TV-INF √STEP-DIM-TV-INF √STEP-DIM-nV-TV-INF
       'make a step' 'make little steps/make steps a little' 'make one little step'

However, verb diminution is common in child-directed speech. The examples in (a) in (26–28) below show diminutive verbs derived from simplex verbs with different diminutive suffixes. The examples in (b) show the grammatical combinations of diminutive suffixes in Slovenian verbs and the examples in (c) show the ungrammatical ones. Just like the suffix *-uc* in BCMS, the diminutive suffixes that combine with *-nV* in Slovenian (i.e. *-k* and *-ic*) are instances of lower diminutives and are merged with the root, i.e. before the categorizing head exponed by a theme vowel.

(26)   a.   čič-a-ti   čič-k-a-ti   čič-ni-∅-ti   (se)   (Slovenian)
            √SIT-TV-INF √SIT-DIM-TV-INF √SIT-nV-TV-INF REFL
            'sit' 'sit in a small way' 'sit down'

       b.   čič-k-ni-∅-ti   (se)
            √SIT-DIM-nV-TV-INF REFL
            'sit in a small way'

       c.   * čič-n(i)-k-a-ti   (se)
            √SIT-nV-DIM-TV-INF REFL

(27)   a.   cap-a-ti   cap-k-a-ti   cap-lj-a-ti   (Slovenian)
          $\sqrt{\text{DRIP}}$-TV-INF $\sqrt{\text{DRIP}}$-DIM-TV-INF $\sqrt{\text{DRIP}}$-DIM-TV-INF
          'take steps' 'take little steps/step a little' 'take little steps/step a little'

       b.   cap-k-lj-a-ti
          $\sqrt{\text{DRIP}}$-DIM-DIM-TV-INF
          'take little steps/step a little'

       c.   * cap-lj-k-a-ti
          $\sqrt{\text{DRIP}}$-DIM-DIM-TV-INF

(28)   a.   hop-a-ti   (Slovenian)
          $\sqrt{\text{HOP}}$-TV-INF
          'hop'
          hop-k-a-ti   hop-lj-a-ti   hop-ni-∅-ti
          $\sqrt{\text{HOP}}$-DIM-TV-INF $\sqrt{\text{HOP}}$-DIM-TV-INF $\sqrt{\text{HOP}}$-nV-TV-INF
          'take little hops/hop a little' 'take little hops/hop a little' 'hop once'

       b.   hop-k-lj-a-ti   hop-k-ni-∅-ti
          $\sqrt{\text{HOP}}$-DIM-DIM-TV-INF $\sqrt{\text{HOP}}$-DIM-nV-TV-INF
          'take little hops/hop a little' 'take little hops/hop a little'

       c.   * hop-lj-ni-∅-ti   *hop-n(i)-lj-(a)-ti
          $\sqrt{\text{HOP}}$-nV-DIM-TV-INF $\sqrt{\text{HOP}}$-nV-DIM-TV-INF

We take the similarity of the position in the words between the diminutive suffix *-lj* and the suffix *-nV* (i.e. the fact that they both can precede another verbal suffix or follow another diminutive suffix) and their complementary distribution in Slovenian as additional evidence for *-nV* combining a diminutive and verbal component in the category head.

5.3 Syntactic modeling

We can now lay out our full structural analysis of the sequence *-nV/-ne*. It is decomposed into two morphemes whose insertion is triggered by two features standing in the head–adjunct configuration: the diminutive feature and the verbal category feature. This is illustrated in (29) and the respective structures in Figures 4–5 on two BCMS verbs, one without and another with the additional diminutive suffix *-uc*.

(29)   zev-nu-∅-ti   zev-uc-nu-∅-ti   (BCMS)
       $\sqrt{\text{YAWN}}$-nV-TV-INF $\sqrt{\text{YAWN}}$-DIM-nV-TV-INF
       'yawn a little'   'yawn a little'

Subsequent head movement derives the surface order.

$$vP$$
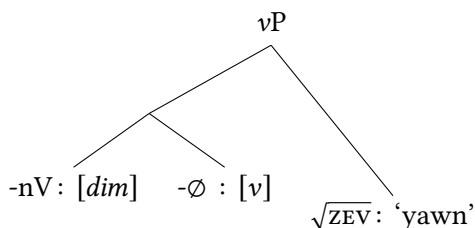
Figure 4: Syntactic representation of the verb *zevnuti* in (29)

-nV : [*dim*]   -∅ : [*v*] $\sqrt{\text{ZEV}}$: 'yawn'
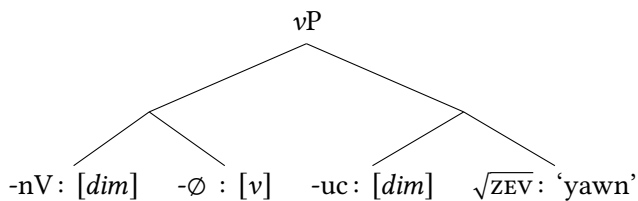
Figure 5: Syntactic representation of the verb *zevucnuti* in (29)

-nV : [*dim*]   -∅ : [*v*]   -uc : [*dim*] $\sqrt{\text{ZEV}}$: 'yawn'

5.4 Formal semantic description

In line with Pietroski (2005) and Arsenijević & Hinzen (2012), we take all syntactic heads to denote predicates and to mutually combine strictly in terms of predicate modification. We follow Arsenijević (2017, 2022) in taking the semantic content of the category feature to be a restriction of the referential domain in terms of the semantic ontological class and unit of counting. The head *v* restricts reference to eventualities, and optionally specifies the quantity structure of the referent of the eventually derived expression at the level of grammatical aspect in terms of neat units, in the sense of Landman (2011), assuming that the absence of this specification, i.e. the default interpretation, matches the messy quantity structure of the eventually derived description. Formally, hence, it is ambiguous between (30a) and (30b).

(30) a. $\lambda x.\text{EVENT}(x)$
      b. $\lambda x.\text{EVENT}(x) \wedge \text{NEAT}(x)$

In both cases, the category feature is a predicate over entities (x), such that the eventually generated expression refers in terms of units *x*, which are optionally *x* neat. For instance, a verb like *sleep* in its typical use (31a) involves a messy quantity structure as in (30a), where units are not strictly bounded and two units may share parts or be part of one another. By contrast, for a verb like the typical

use of *blink* (31b), the quantity structure of the predicate is neat, as in (30b), where units are strictly bounded and disjoint.

(31) a. John slept.
      b. Mary blinked.

We analyze the diminutive feature as a specification of a low degree on some measure function, as in (32a) (where $\textsc{m}(x)$ stands for the measure function applied to x). This measure function as well as the standard degree are both provided from the context. In the domain of concrete individuals, the measure function typically targets size, and in the domain of events their temporal duration. In the verbal structure, the diminutive feature may occur in two positions. One is to merge with the base from which the verb derives, typically a root or a complex structure, and apply diminution to it. This typically results in the choice of the measure of intensity of action or of the fit of the description (raising the interpretation of atypical nature of the eventuality with respect to the description used). This is structurally illustrated in (32b).

The other option is that it merges with the category head, typically receiving the measure of duration interpretation, i.e. the unit event has a shorter (temporal or other) interval than the standard for the event kind, as in (32c).[13] As the relation SMALLER entails boundedness, this imposes, by presupposition, restriction to neat predicates. As a result, the suffix *-nV* combines with neat *v*'s only, i.e. it accommodates neat quantity structure in the category head. When the diminutive feature adjoins to the category head, it is hence interpreted as specifying the bounded nature and small size of the unit eventuality. This is how for instance *trk-nu-ti* '$\sqrt{run}$-TV-INF' gets the interpretation of a small (i.e. atomic) instance of running.[14]

---

[13]Here we assume that the category head has the nature of a count classifier: it specifies the manner of reference, by specifying reference units (see Arsenijević 2022 for an elaboration and further references). We follow Milosavljević (2023b) in assuming that the verbal structure includes further projections dedicated to atomicity and grammatical number, where the units specified by the category head are further specified and structured to restrict the description and eventually reference too, quite parallel to the way this is traditionally modeled in the nominal domain.

[14]An anonymous reviewer raises the question of whether the neatness condition as part of the semantics of the suffix *-nV* is justified, given that this suffix can combine with non-verbal bases, i.e. may be added to stems that denote uncountable nouns or onomatopoeic words (in Polish). While in Western South Slavic too the suffix *-nV* combines with bases that are attested also as nominal (e.g. *korak-nu-ti* 'step'; with the noun *korak* 'step' and the Slovenian verb *nasmeh-ni-ti* 'smile' with the noun *nasmeh* 'smile'), or onomatopoeic (e.g. *tres-nu-ti* 'snap, crack' in BCMS

(32) a. $[dim] := \lambda x[\text{M}(x) < \text{STD}]$

    b. $\sqrt{\text{TRK}} := \lambda x[[\![\sqrt{\text{TRK}}]\!](x)]$
       by predicate modification:
       $[[dim]\sqrt{\text{TRK}}] := \lambda x[[\![\sqrt{\text{TRK}}]\!](x) \wedge \text{M}(x) < \text{STD}]$

    c. $[v] := \lambda x[\text{EVENT}(x) \wedge \text{NEAT}(x)]$
       by predicate modification:
       $[[dim][v]] := \lambda x[\text{EVENT}(x) \wedge \text{NEAT}(x) \wedge \text{M}(x) < \text{STD}]$

    d. by predicate modification with the root:
       $[[[dim][v]]\sqrt{\text{TRK}}] := \lambda x[\text{EVENT}(x) \wedge \text{NEAT}(x) \wedge \text{M}(x) < \text{STD} \wedge$
       $[\![\sqrt{\text{TRK}}]\!](x)]$

Considering that the suffix *-nV* realizes the diminutive adjoined to the category head and the suffix *-uc* the one composed with the root or other base, this analysis predicts that the suffix *-uc* will be ambiguous, while the suffix *-nV* will not be used with the meaning of low intensity without restriction to neat structure. Indeed, the latter is exactly what is discussed around example (29), while, as shown in (33), *-uc* may also have the pure low intensity interpretation, as all the verbs in (33) are ambiguous between the durative low intensity interpretation and that of an iteration of pointy intervals of the (low intensity or not) eventuality.

(33) svetl-uc-a-ti   bel-uc-a-ti   (BCMS)
     √LIGHT-DIM-TV-INF √WHITE-DIM-TV-INF
     'emit light a little'   'be white a little'
     svir-uc-a-ti   šet-uc-a-ti
     √PLAY-DIM-TV-INF   √WALK-DIM-TV-INF
     'play a little'   'walk a little'

5.5 Western South Slavic verbal suffixation

The proposed analysis postulates three syntactic positions in which verbal suffixes are generated in Western South Slavic (and possibly more generally Slavic).

and *tresk-ni-ti* in Slovenian, with the respective words *tres!* and *tresk!* also used as interjections expressing a sudden or sharp sound, like the sound of something breaking or snapping), such examples do not constitute a counterargument for our analysis. Namely, in our DM implementation, the suffix *-nV* merges with the category head or with a categorized root. This means that apparent onomatopoeic or nominal bases are verbalized before *-nV* enters the structure, so that these are not counterexamples to the verbal and/or neatness presupposition. More generally, there are two possibilities for "nominal" bases: either the root is nominalized by a nominal head, and then verbalized, or the same root appears in both nominal and verbal structures. In both cases, *-nV* would attach to the verbal category head (i.e. verbalized structure).

These are, bottom up: (i) a position merging with the base, be it a root or a category, in which ambiguous diminutive suffixes are generated (suffixes *-uc*, *-uš* in BCMS, *-ic*, *-k* in Slovenian), (ii) adjunct to the category head, also reserved for the diminutive suffix, but here realized as *-nV*, and (iii) the position of the imperfective (or biaspectual) verbal suffixes, traditionally associated with some aspectual projection. The last type of suffixes has been analyzed in Simonović et al. (2023) and Arsenijević et al. (2023) as consisting purely of theme vowels, and thus realizing the bare verbal category feature. This reduces the set of possible positions to only two: that below the verbal category head and the verbal category head itself.[15]

5.6 Comparison to previous analyses

Our analysis shares some properties with several others. Like Svenonius (2004) and Biskup (2023a,b, 2024), it relates the suffix with the verbal category. As Kwapiszewski (2020), our analysis attributes to the suffix specification of properties of quantity (the unit of counting), and as Arsenijević (2006), it associates it with diminutivity. Finally, in line with Armoškaitė & Sherkina-Lieber (2008), we associate the suffix with the unit of counting, and with Łazorczyk (2010), Taraldsen Medová & Wiland (2019), and Wiland (2019), we offer a bimorphemic analysis. Here is how our analysis accounts for the specific properties of the suffix presented above.

In terms of meaning, Semelfactives present the fully compositional interpretation of the suffix *-nV*: they denote one counting unit for the respective event predicate which is smaller than the standard for such an eventuality. Natural Perfectives are a special case, emerging when the event predicate specifies a salient atom. The salience of this interpretation imposes it as a pragmaticized meaning of the diminutive feature applying to the unit of counting specified by the event predicate. The Perfective Delimitative interpretation emerges when the event predicate specifies no salient counting unit. The diminutive feature presupposes such a unit, and by default takes bounded temporal intervals as the unit of counting. The salient natural class of bounded temporal intervals are points in time (no other length or type makes a natural class), resulting in semelfactivity. The Degree Achievement interpretation is not productive anymore, indicating that the suffix no longer contributes a meaning that derives it (see Rothstein 2008b for an explanation of the source of *-nV* degree achievements).

---

[15]Due to space limitations, we leave aside the status of suffixes that are used for integrating borrowed verbs, such as *-ir* (*kop-ir-a-ti* [copy-ir-tv-inf] 'copy') and *-is* (*determin-is-a-ti* [determine-is-tv-inf] 'determine') in BCMS.

The diminutive semantic component, which is at least latently always present with *-nV* (except in the unproductive class of degree achievements) is part of the meaning of the suffix. Telicity is part of the semantic specification of the meaning of the suffix, in the form of the presupposition of a unit of counting required by the meaning of smallness operating over the verbalizer which specifies properties of quantity. Perfectivity is generally strongly associated with telicity in Slavic (Borer 2005b, Arsenijević 2006, 2023, Łazorczyk 2010, Milosavljević 2022, 2023a,b), and the same mechanisms are likely at play with *-nV*. Modeling this suffix as the only one with additional syntactic/semantic content next to that borne by the theme vowel (Simonović et al. 2023, Arsenijević et al. 2023) enables capturing its being also the only one that imposes telicity and perfectivity.

By our analysis, *-nV* selects the TV ∅/*e*, i.e. the *-e* in the present stem is not part of the suffix but a TV. This fits the analysis where the diminutive feature realized as *-nV* is left-adjoined to the verbal category feature realized as the TV. Our view obviates the question about the complementary distribution of *-nV/-ne* with theme vowels, since the sequence *-nV/-ne* includes a TV.

The compatibility of the suffix *-nV* with secondary imperfectivizing suffixes in at least some Slavic languages (BCMS included), as well as the ability to stack with other imperfective suffixes, is not a problem for our approach since the suffix does not target the AspP, but a lower head (i.e. in the analysis by Arsenijević et al. 2023, *-nV* derives telic predicates, which then can be reverbalized).

Finally, unlike other analyses, ours also predicts that the suffix *-nV* combines with the root-level diminutive suffix *-uc* analogous to double diminution in nouns and adjectives.

6 Conclusion

The paper revisits the Slavic verbal suffix *-nV*, and highlights a range of new qualitative and quantitative observations and generalizations which have not yet been reported or supported by precise quantitative data in previous descriptive and theoretical accounts of this suffix. We observe a unique status of the suffix among verbal suffixes based on the properties of its use (e.g., it may combine with other verbal suffixes, which does not hold for other suffixes; it does not select the theme vowel the other suffixes do). To predict and explain the special properties of the suffix, we propose the decomposition of the suffix into two components, an actual suffix (*-nV*) and a theme vowel (∅/*e*), realizing diminution and the verbal category, respectively. We provide a formalization for the diminutive semantics, and a syntactic structure for the position of its base-generation.

Abbreviations

1	first person		NOM	nominative
3	third person		PASS	passive
ACC	accusative		PL	plural
ADJ	adjective		PREF	prefix
AUX	auxiliary		PRS	present tense
COMP	complementizer		PST	past
DIM	diminutive		PTCP	participle
F	feminine		REFL	reflexive
IMP	imperative		SG	singular
INF	infinitive		SI	secondary imperfective
INFL	inflectional ending		SUFF	suffix
M	masculine		TV	theme vowel

Acknowledgments

References

Armoškaitė, Solveiga & Marina Sherkina-Lieber. 2008. Event number suffixes in Russian and Lithuanian. In Andrei Antonenko, John F. Bailyn & Christina Y. Bethin (eds.), *Formal Approaches to Slavic Linguistics 16: The Stony Brook Meeting 2007*, 1–15. Ann Arbor: Michigan Slavic Publications.

Arsenijević, Boban. 2006. *Inner aspect and telicity* (LOT Dissertation Series 142). Utrecht: LOT.

Arsenijević, Boban. 2017. Gender, like classifiers, specifies the type of partition: Evidence from Serbo-Croatian. In Jessica Kantarovich, Tran Truong & Orest Xherija (eds.), *Proceedings from the Annual Meeting of the Chicago Linguistic Society 52*, 21–37. Chicago: Chicago Linguistic Society.

Arsenijević, Boban. 2022. Adjectives as a lexical category: A story of striving for extension. In Phoevos Panagiotidis & Moreno Mitrović (eds.), *A°: The lexical status of adjectives*, 121–152. Amsterdam: John Benjamins. DOI: 10.1075/lfab.17.04ars.

Arsenijević, Boban. 2023. Specification of telicity in Serbo-Croatian, without null prefixes. In Olav Mueller-Reichau, Petr Biskup, Marcel Börner & Iuliia Shcherbina (eds.), *Advances in formal Slavic linguistics 2021*, 1–37. Berlin: Language Science Press. DOI: 10.5281/zenodo.10123627.

Arsenijević, Boban & Wolfram Hinzen. 2012. On the absence of X-within-X recursion in human grammar. *Linguistic Inquiry* 43(3). 423–440. DOI: 10.1162/LING_a_00095.

Arsenijević, Boban, Stefan Milosavljević & Marko Simonović. 2023. Secondary imperfectivisation is reverbalisation is addition of theme vowels. Presented at the 56th Annual Meeting of the Societas Linguistica Europaea (SLE 56), Workshop Inflection and derivation in the Slavic verb, National and Kapodistrian University of Athens, August 29–September 1, 2023.

Arsenijević, Boban, Katarina Gomboc Čeh, Franc Lanko Marušič, Stefan Milosavljević, Petra Mišmaš, Jelena Simić, Marko Simonović & Rok Žaucer. 2024. *Database of the Western South Slavic Verb HyperVerb 2.0 – WeSoSlaV*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1846.

Bacz, Barbara. 2012. Reflections on semelfactivity in Polish. *Studies in Polish Linguistics* 7(1). 107–128.

Barić, Eugenija, Mijo Lončarić, Dragica Malić, Slavko Pavešić, Mirko Peti, Vesna Zečević & Marija Znika. 1997. *Hrvatska gramatika.* Zagreb: Školska knjiga.

Bešlin, Maša. 2023. Passive vP is not phasal in Bosnian/Croatian/Serbian. *Proceedings of the Linguistic Society of America* 8(1). 5499. DOI: 10.3765/plsa.v8i1.5499.

Biskup, Petr. 2019. *Prepositions, case and verbal prefixes: The case of Slavic* (Linguistik Aktuell/Linguistics Today 255). Amsterdam: John Benjamins. DOI: 10.1075/la.255.

Biskup, Petr. 2023a. Aspect separated from aspectual markers in Russian and Czech. In Olav Mueller-Reichau, Petr Biskup, Marcel Börner & Iuliia Shcherbina (eds.), *Advances in formal Slavic linguistics 2021*, 61–98. Berlin: Language Science Press. DOI: 10.5281/zenodo.10123631.

Biskup, Petr. 2023b. Deriving morphological aspect in Russian: A matryoshka way. In Christina Clasmeier & Julia Golbekand (eds.), *Linguistische Beiträge zur Slavistik: XXIX. JungslavistInnen-Treffen vom 09. bis 10. September 2021 an der Ruhr-Universität Bochum*, 9–30. Berlin: Peter Lang.

Biskup, Petr. 2024. An Agree analysis of the morphological aspect in Slavic. In Silke Fischer, Doreen Georgi, Fabian Heck, Johannes Hein, Anke Himmelreich, Andrew Murphy & Philipp Weisser (eds.), *Strictly cycling*, 55–74. Leipzig: Institut für Linguistik, Universität Leipzig. https://www.philol.uni-leipzig.de/fileadmin/Fakult%C3%A4t_Philo/Linguistik/Forschung/LAB/LAB_97/LAB97_04_biskup.pdf.

Borer, Hagit. 2005a. Some notes on the syntax of quantity. In Paula Kempchinsky & Roumyana Slabakova (eds.), *Aspectual inquiries* (Studies in Natural Language and Linguistic Theory 62), 41–68. Dordrecht: Springer. DOI: 10.1007/1-4020-3033-9_3.

Borer, Hagit. 2005b. *Structuring sense, vol. 2: The normal course of events.* Oxford: Oxford University Press.

Breznik, Anton. 1934. *Slovenska slovnica za srednje šole.* Celje: Založila Družba sv. Mohorja.

Caha, Pavel. 2009. *The nanosyntax of case.* Tromsø: University of Tromsø. (Doctoral dissertation).

De Belder, Marijke. 2008. Size matters: Towards a syntactic decomposition of countability. In Natasha Abner & Jason Bishop (eds.), *WCCFL 27: Proceedings of the 27th West Coast Conference on Formal Linguistics*, 116–122. Somerville: Cascadilla Proceedings Project. http://www.lingref.com/cpp/wccfl/27/paper1823.pdf.

De Belder, Marijke. 2011. A morphosyntactic decomposition of countability in Germanic. *Journal of Comparative Germanic Linguistics* 14(3). 173–202. DOI: 10.1007/s10828-011-9045-0.

De Belder, Marijke, Noam Faust & Nicola Lampitelli. 2014. On a low and a high diminutive: Evidence from Italian and Hebrew. In Artemis Alexiadou, Hagit Borer & Florian Schäfer (eds.), *The syntax of roots and the roots of syntax*, 149–163. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199665266.003.0007.

Dickey, Stephen M. & Laura A. Janda. 2009. Хохомнул, схимрил: The relationship between semelfactives formed with *-nu-* and *s-* in Russian. *Russian Linguistics* 33(3). 229–248. DOI: 10.1007/s11185-009-9044-9.

Diesing, Molly. 1998. Light verbs and the syntax of aspect in Yiddish. *The Journal of Comparative Germanic Linguistics* 1. 119–156. DOI: 10.1023/A:1009751908064.

Farris-Trimble, Ashley W. 2008. *Cumulative faithfulness effects in phonology.* Indiana University. (Doctoral dissertation).

Gladney, Frank Y. 2013. On the syntax, morphology, and semantics of Russian verbal aspect. *Slavic and East European Journal* 57(4). 628–648. https://www.jstor.org/stable/24642488.

Isačenko, Aleksandr V. 1962. *Die russische Sprache der Gegenwart: Teil I: Formenlehre.* Halle (Saale): Niemeyer.

Ivšić, Stjepan, Josip Vrana & Radoslav Katičić. 1970. *Slavenska poredbena gramatika.* Zagreb: Školska knjiga.

Jabłońska, Patrycja. 2007. *Radical decomposition and argument structure.* Tromsø: University of Tromsø. (Doctoral dissertation).

Janda, Laura A. 2007. Aspectual clusters of Russian verbs. *Studies in Language* 31(3). 607–648. DOI: 10.1075/sl.31.3.04jan.

Kagan, Olga. 2008. On the semantics of aspect and number. In Andrei Antonenko, John F. Bailyn & Christina Y. Bethin (eds.), *Formal Approaches to Slavic Linguistics 16: The Stony Brook Meeting 2007*, 185–198. Ann Arbor: Michigan Slavic Publications.

Kagan, Olga. 2010. Russian aspect as number in the verbal domain. In Brenda Laca & Patricia Hofherr (eds.), *Layers of aspect*, 91–112. Stanford: CSLI Publications.

Klimek-Jankowska, Dorota, Anna Czypionka, Wojciech Witkowski & Joanna Błaszczak. 2018. The time course of processing perfective and imperfective aspect in Polish: Evidence from self-paced reading and eye-tracking experiments. *Acta Linguistica Academica* 65(2–3). 293–351. DOI: 10.1556/2062.2018.65.2-3.4.

Kovačević, Peđa, Stefan Milosavljević & Marko Simonović. 2024. Theme-vowel minimal pairs show argument structure alternations. *Journal of Linguistics*. Advance online publication, 1–30. DOI: 10.1017/S0022226723000415.

Kuznetsova, Julia & Anastasia Makarova. 2012. Distribution of two semelfactives in Russian: -nu- and -anu-. *Oslo Studies in Language* 4(1: The Russian Verb). Atle Grønn and Anna Pazelskaya (issue eds.), 155–176. DOI: 10.5617/osla.231.

Kwapiszewski, Arkadiusz. 2020. Aspect and verbalising morphology in Polish nominalisations. In Astrid van Alem, Mirella De Sisto, Elisabeth J. Kerr & Joanna Wall (eds.), *Proceedings of ConSOLE XXVIII*, 75–99. Leiden: Leiden University Centre for Linguistics.

Kwapiszewski, Arkadiusz. 2022. *Aspect and event structure: The morphosyntax of Polish verbs from a cross-linguistic perspective*. Oxford: University of Oxford. (Doctoral dissertation).

Landman, Fred. 2011. Count nouns, mass nouns, neat nouns, mess nouns. *Baltic International Yearbook of Cognition, Logic and Communication* 6(1). 1–67. DOI: 10.4148/biyclc.v6i0.1579.

Łazorczyk, Agnieszka A. 2010. *Decomposing Slavic aspect: The role of aspectual morphology in Polish and other Slavic languages*. Los Angeles: University of Southern California. (Doctoral dissertation).

Ljubešić, Nikola & Filip Klubička. 2014. bs,hr,srWaC - web corpora of Bosnian, Croatian and Serbian. In Felix Bildhauer & Roland Schäfer (eds.), *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, 29–35. Gothenburg: Association for Computational Linguistics. DOI: 10.3115/v1/W14-0405.

Logar-Berginc, Nataša, Simon Krek, Tomaž Erjavec, Miha Grčar, Peter Halozan & Simon Šuster. 2012. Referenčni korpusi slovenskega jezika (cc)Gigafida in (cc)KRES. In *Zbornik Osme konference jezikovne tehnologije, 8. do 12. oktober*

2012: Zbornik 15. mednarodne multikonference Informacijska družba, vol. C, 57–62. https://nl.ijs.si/isjt12/proceedings/isjt2012_11.pdf.

Makarova, Anastasia & Laura A. Janda. 2009. Do it once: A case study of the Russian *-nu-* semelfactives. *Scando-Slavica* 55(1). 78–99. DOI: 10.1080/00806760903175417.

Markman, Vita G. 2008. On Slavic semelfactives and secondary imperfectives: Implications for the split 'AspP'. In *Proceedings of the 31st Annual Penn Linguistics Colloquium* (University of Pennsylvania Working Papers in Linguistics 14(1)), 255–268. Philadelphia: Penn Graduate Linguistics Society. https://repository.upenn.edu/handle/20.500.14332/44670.

Marušič, Franc Lanko, Rok Žaucer, Petra Mišmaš, Boban Arsenijević, Marko Simonović, Stefan Milosavljević, Katarina Gomboc Čeh & Jelena Simić. 2022. *Database of the Western South Slavic verb HyperVerb 1.0.* Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1683.

Mathieu, Éric. 2012. Flavors of division. *Linguistic Inquiry* 43(4). 650–679. DOI: 10.1162/ling_a_00110.

Milosavljević, Stefan. 2019. O upotrebi iterativnih glagola u perfektu za označavanje izbrojivog mnoštva radnji i njihovoj konkurenciji sa svršenim glagolima (na primerima iz novinarskog stila srpskog jezika). *Naučni sastanak slavista u Vukove dane* 48(3). 149–191. DOI: 10.18485/msc.2019.48.3.ch8.

Milosavljević, Stefan. 2022. The delimitative prefix *po-*, durative adverbials, and Slavic aspectual composition. In Annie Holtz, Iva Kovač & Rasmus Puggaard-Rode (eds.), *Proceedings of ConSOLE XXX*, 17–42. Leiden: Leiden University Centre for Linguistics.

Milosavljević, Stefan. 2023a. Simple imperfective verbs, the sequence of similar events interpretation, and Slavic aspectual composition. In Olav Mueller-Reichau, Petr Biskup, Marcel Börner & Iuliia Shcherbina (eds.), *Advances in formal Slavic linguistics 2021*, 265–292. Berlin: Language Science Press. DOI: 10.5281/zenodo.10123645.

Milosavljević, Stefan. 2023b. *Specification of event duration and aspectual composition in Slavic.* Graz: University of Graz. (Doctoral dissertation).

Milosavljević, Stefan & Boban Arsenijević. 2022. What differentiates Serbo-Croatian verbal theme vowels: Content or markedness? *Glossa: a journal of general linguistics* 7(1). 1–36. DOI: 10.16995/glossa.8535.

Milosavljević, Stefan, Petra Mišmaš, Marko Simonović, Boban Arsenijević, Katarina Gomboc Čeh, Franc Lanko Marušić, Jelena Simić & Rok Žaucer. 2023. *Database of the Western South Slavic Verb HyperVerb – derivation.* Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1855.

Nesset, Tore. 2012. One or several categories? The Old Church Slavonic *nǫ*-verbs and linguistic profiling. *Russian Linguistics* 36(3). 285–303. DOI: 10.1007/s11185-012-9093-3.

Nesset, Tore. 2013. The history of the Russian semelfactive: The development of a radial category. *Journal of Slavic Linguistics* 21(1). 123–169. https://www.jstor.org/stable/24600451.

Nordrum, Maria. 2019. *Together and apart: Perfective verbs with a prefix and the semelfactive suffix* -nu- *in Contemporary Standard Russian*. Tromsø: UiT The Arctic University of Norway. (Doctoral dissertation). https://hdl.handle.net/10037/17718.

Ott, Dennis. 2011. Diminutive-formation in German: Spelling out the classifier analysis. *Journal of Comparative Germanic Linguistics* 14(1). 1–46. DOI: 10.1007/s10828-010-9040-x.

Pietroski, Paul M. 2005. *Events and semantic architecture*. Oxford: Oxford University Press.

Progovac, Ljiljana. 2005. *A syntax of Serbian: Clausal architecture*. Bloomington: Slavica Publishers.

Rijkhoff, Jan N.M. 1991. Nominal aspect. *Journal of Semantics* 8(4). 291–309. DOI: 10.1093/jos/8.4.291.

Rothstein, Susan. 2008a. Telicity, atomicity and the Vendler classification of verbs. In Susan Rothstein (ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect* (Linguistik Aktuell/Linguistics Today 110), 43–77. Amsterdam: John Benjamins. DOI: 10.1075/la.110.04rot.

Rothstein, Susan. 2008b. Two puzzles for a theory of lexical aspect: Semelfactives and degree achievements. In Johannes Dölling, Tatjana Heyde-Zybatow & Martin Schäfer (eds.), *Event structures in linguistic form and interpretation*, 175–198. Berlin: Walter de Gruyter. DOI: 10.1515/9783110925449.175.

Sauerland, Uli. 2003. A new semantics for number. In Robert B. Young & Yuping Zhou (eds.), *SALT 13: Proceedings of the 13th Semantics and Linguistic Theory Conference*, 258–275. Ithaca: CLC Publications. DOI: 10.3765/salt.v13i0.2898.

Schoorlemmer, Maaike. 1997. The role of the internal argument in the Russian aspectual system. In Uwe Junghanns & Gerhild Zybatow (eds.), *Formale Slavistik*, 229–238. Frankfurt am Main: Vervuert Verlag.

Schoorlemmer, Maaike. 2004. Syntactic unaccusativity in Russian. In Artemis Alexiadou, Elena Anagnostopoulou & Martin Everaert (eds.), *The unaccusativity puzzle: Explorations of the syntax–lexicon interface*. Oxford: Oxford University Press.

Simonović, Marko. 2015. *Lexicon immigration service: Prolegomena to a theory of loanword integration* (LOT Dissertation Series 393). Utrecht: LOT.

Simonović, Marko, Stefan Milosavljević & Boban Arsenijević. 2023. Serbo-Croatian secondary imperfectivisers consist of theme vowels. *Journal of Slavic linguistics* 31(FASL 30 issue). 1–27. http://ojs.ung.si/index.php/JSL/article/view/178.

Smith, Carlota S. 1997. *The parameter of aspect*. 2nd edn. Dordrecht: Kluwer.

Sokolova, Svetlana. 2015. Rabotnul na slavu-gul'ni smelo!: -NU- as a universal aspectual marker in non-standard Russian. In Misushi Kitajo (ed.), *Aspektual'naja semantičeskaja zona: Tipologija sistem i scenarii diaxroničeskogo razvitija. Sbornik statej V Meždunarodnoj konferencii Komissii po aspektologii meždunarodnogo komiteta slavistov*, 271–277. Kyoto: Tanaka Print.

Stanojčić, Živojin & Ljubomir Popović. 2008. *Gramatika srpskog jezika (za gimnazije i srednje škole)*. Beograd: Zavod za Udžbenike.

Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. *Nordlyd* 36(1: Special issue on Nanosyntax). 1–6. DOI: 10.7557/12.213.

Štarkl, Ema, Stefan Milosavljević, Marko Simonović & Boban Arsenijević. 2024. *The Western South Slavic verbal suffix -nv/-ne*. Database. DOI: 10.5281/zenodo.13912034.

Stevanović, Mihailo. 1986. *Savremeni srpskohrvatski jezik i*. Beograd: Naučna knjiga.

Stojanović, Milica. 2016. *Poliprefigirani glagoli u srpskom jeziku*. Belgrade: University of Belgrade. (Doctoral dissertation). https://hdl.handle.net/21.15107/rcub_nardus_8566.

Svenonius, Peter. 2004. Slavic prefixes and morphology: An introduction to the Nordlyd volume. *Nordlyd* 32(2). 177–204. DOI: 10.7557/12.67.

Taraldsen Medová, Lucie & Bartosz Wiland. 2019. Semelfactives are bigger than degree achievements: The nanosyntax of Czech and Polish semelfactive and degree achievement verb stems. *Natural Language & Linguistic Theory* 37(4). 1463–1513. DOI: 10.1007/s11049-018-9434-z.

Toporišič, Jože. 2000. *Slovenska slovnica*. Maribor: Založba Obzorja.

Townsend, Charles E. 1968. *Russian word-formation*. New York: McGraw-Hill.

Vidovič Muha, Ada. 2011. *Slovensko skladenjsko besedotvorje*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.

Wiland, Bartosz. 2019. *The spell-out algorithm and lexicalization patterns: Slavic verbs and complementizers* (Open Slavic Linguistics 2). Berlin: Language Science Press. DOI: 10.5281/zenodo.2636394.

Wiltschko, Martina. 2006. Why should diminutives count? In Hans Broekhuis, Norbert Corver, Riny Huybregts, Ursula Kleinhenz & Jan Koster (eds.), *Organizing grammar: Linguistic studies in honor of Henk van Riemsdijk*, 669–678. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110892994.669.

Zdziebko, Sławomir. 2017. On the structure and interpretation of Polish passives. *Acta Linguistica Academica* 64(4). 563–617. DOI: 10.1556/2062.2017.64.4.4.

Chapter 22

Wh-indefinites in Russian

Ksenia Zanon

University of Cambridge

The goal of this paper is to chart the expanse of environments that license wh-indefinites in Russian. Primarily a descriptive endeavor, this study provides a more exhaustive empirical coverage of the phenomenon than what has heretofore been documented. Appearing in a proper subset of *nibud'*-licensing contexts, wh-indefinites require a clausebounded nonveridical operator and exhibit sensitivity to scalarity. The central analytical import concerns the dichotomy "clitic" vs. "nonclitic". Instead of a rigid binary taxonomy, I endorse the view that there is a continuum clitic⟷non-clitic, which accommodates elements of transitional flavor. Wh-indefinites are just such elements: not quite clitics proper, they are not full tonic forms either.

1 Introduction

Wh-indefinites have the morphological shape of a wh-word and the interpretation of an indefinite. A postverbal *kto* in a polar question (1a) functions as an indefinite pronoun in contrast to the sentence-initial one in (1b), interpreted as a wh-word.

(1) a. Prišel kto?   YN question
        came who.INDF
        'Did anybody come?'

     b. Kto prišel?
        who came
        'Who came?'

In addition to (1a), four other contexts in (2) reportedly enable licensing of wh-indefinites (examples (2a)–(2c) appear in Yanovich 2005, (2d) – in Hengeveld et al. 2018). However, it turns out that not all subjunctives tolerate wh-indefinites but only those that embed some negative component. Likewise, matrix negation is insufficient by itself: my informants deem (2d) degraded.[1]

(2)   a.   Esli kto   pridet, pozovi menja.   Antecedent of conditional
           if   who.INDF comes call   me
           'If anybody comes, call me.'

      b.   Možet, kto   prixodil.   Modal adverbs
           maybe who.INDF came
           'Maybe somebody came.'

      c.   Petja zaper dver', čtoby   kto   ne   vošel.   Subjunctive
           Petja locked door that.SBJV who.INDF NEG entered
           'Peter locked the door, lest somebody enter.'

      d.   (?*) Ne poxože, čto Vasja kogo   uvidel.   Matrix negation
                NEG similar that Vasja whom.INDF saw
                'It does not look like Vasja saw anybody.'

A wh-indefinite shares a requirement for a licenser with a (better studied) *nibud'*-indefinite. Neither is possible in past episodic declaratives like (3).

(3)   * Včera   kto-nibud' kto umer.
        yesterday who-*nibud'*   who.INDF died
        Intended: 'Yesterday someone died.'

Nibud'-indefinites are morphologically decomposable into a wh-element and an invariable suffix -*nibud'*: e.g., *kto-nidud'* 'who.NOM-*nibud'*, *čto-nidud'* 'what-*nibud'*, etc. Roughly, -*nibud'*-indefinites are eligible in nonveridical contexts (questions, conditionals, imperatives, in modal, future and iterative constructions, subjunctives of all flavors, under propositional attitude verbs like *doubt, hope*) as well as the scope and restriction of universal quantifiers (Fitzgibbons 2010, Padučeva 2016, Pereltsvaig 2008). In the next section I show that wh-indefinites appear in a proper subset of *nibud'*-environments and identify the conditions that impede or enable the licensing of wh-indefinites.

---

[1] Data are elicited from five informants on the scale 1–5. Judgments are presented in the following format: '*' = 1, '?*' = 2, '??' = 3, '?' = 4. In controversial cases, I provide all obtained values (e.g., '*/?*'). Positive data are mostly sourced from the national corpus (ruscorpora.ru) or found online. To keep the exposition unencumbered, I indicate the type of source instead of providing a long URL (specifics should be reconstructible via a reverse search). Angle brackets, i.e. '⟨ ⟩', are used to indicate elicited alternatives in the naturally occurring or reported examples.

Before diving in, two short asides are in order. First, a handful of constructions have been excluded from the present consideration on the grounds that the relevant wh-element does not fit the profile of a prototypical indefinite in an obvious way (or if its status is controversial). These include: (i) Modal-Existential configurations (MECs) like (4) and (ii) two subspecies of relatives in (5). On MECs, I refer the reader to Šimík (2017) for a concise literature overview on the topic.[2] Constructions like (5a) were first noted in Rudin (2007) for Bulgarian (see also Caponigro & Fălăuş 2022). Correlatives like (5b) (example provided by Reviewer 1) are discussed in Citko (2009) with antecedents in Izvorski (1996) (for a more general literature overview see Lin 2020).[3]

(4) Mne est' čto gde počitat'.
     to.me is what where to.read
     'I have something to read.'

(5) a. My otpravili, kto skol'ko   naskreb.
         we sent   who how.much scraped.together
         'We sent however much each scraped together.'

     b. Kto kogo uvidit (na večerinke), tot   s   tem   i
         who whom will.see at party   that.one with that.one and
         pozdorovaetsja.
         will.greet
         'Whoever sees whomever at the party will greet them.'

The second point concerns the shape of the indefinite itself. In Russian, it need not be a bare wh-word: complex expressions (i.e., *which X*, as in (6)) are admissible in all the licensing contexts catalogued in the ensuing sections.

(6) Byt' možet zavtra   kakoj   ukazik sverxu   spustjat i
     to.be may tomorrow which.INDF edict from.above will.issue and
     togda kotu pod xvost vse ego trudy (...)
     then to.cat under tail all his labors
     'It may well be that tomorrow they'll issue some edict from above and
     then all his labors are for naught.'   (S. Xabliev. *Povtornye ogni.* 2002)

---

[2]But see Šimík (2009) for arguments that wh-elements in MECs are (Hamblin) indefinites, after all.

[3]Belyaev & Haug (2020) defend the position that the wh-elements in these constructions owe their provenance to indefinites. Arsenijević (2009) treats the wh-elements in correlatives as "extreme non-specific expression(s)". It is worth pointing out that the latter two analyses view correlatives as subtypes of conditionals.

There are, however, gaps in terms of the membership in major ontological categories ('person', 'thing', 'time', 'place', 'manner', 'reason', etc., on which see Haspelmath 1997: 29–31, and references therein). In particular, reason and manner and categories (attempted in (7b) and (8b), respectively) prove to be unfit for bare indefinites but open to the *nibud'*-series (as attested by the (a)-examples).

(7) a. Esli že počemu-nibud' emu nel'zja budet priexat' ko
    if FOC why-*nibud'* to.him impossible will.be to.come at
    vremeni moego priezda,...
    time of.my arrival
    'If it would be impossible for some reason for him to come by the
    time of my arrival,...' (P. Tchaikovskii. Letters. 1884.)

  b. *Esli emu počemu nel'zja budet priexat'...
    if to.him why.INDF impossible will.be to.come
    Intended: 'If for any reason it would be impossible for him to come...'

(8) a. Možno ėto kak-nibud' ispravit'?
    possible this how-*nibud'* to.fix
    'Is it possible to fix it in some way?' (beauty forum. 2023)

  b. *Možno eto kak ispravit'?
    possible this how.INDF to.fix
    Intended: 'Is it possible to fix it in any way?'

Finally, I would be remiss not to point out the crosslinguistic ubiquity of wh-indefinites (Gärtner 2009), sometimes accommodated under the rubric of "indeterminate" pronouns in the literature (Kratzer & Shimoyama 2017). A wh-indefinite is easy to spot (it looks just like a wh-word), but the environments that render it happy differ across languages: some, like Dutch (Postma 1994) or Passamaquoddy (Bruening 2007), do not impose licensing requirements; others, like Chinese, do (see, e.g., Bruening 2007, Lin 2014; a more recent theoretical debate is found in Chierchia & Liao 2015, Giannakidou & Lin 2016, Liu & Yang 2021). Then there is the question of how wh-words and wh-indefinites are related: whether their syncretism is a matter of homophony or homonymy is addressed in Bhat (2004). All of this is to say that there is a massive volume of scholarly output on the topic, which I cannot hope to address in any detail here. Mine is a case study of the licensing conditions of the Russian wh-indefinite.

2 Licensing contexts

The default licensing requirement for wh-indefinites is the configuration that enables "epistemic neutrality", understood as Giannakidou's "nonveridical equilibrium" defined in (9). Such "prototypical inquisitiveness", i.e. genuine noncommitment of an epistemic agent to one of the polar values in the partitioned information state, arises in neutral Yes / No (YN) questions, conditionals, and under possibility modals (*might*).

(9)   An information state W is in nonveridical equilibrium iff W is partitioned into p and $\neg p$, and there is no bias towards p or $\neg p$.

(Giannakidou 2013: 121)

Equilibrium might be disrupted in a variety of ways: intonation, tags, adverbs, NPIs, etc. all tilt the balance, inducing the effect of speaker bias. For instance, although *John speaks English, doesn't he?* retains its nonveridical properties, it also supplies an inference that the proposition is true.

Generally speaking, while *-nibud'*-items are compatible with nearly all nonveridical contexts (whether biased or not), the conditions on wh-indefinites are more stringent. The "default" licenser must contribute to the representation consistent with epistemic neutrality (Section 2.1). But there are multiple ways to bypass this requirement: by introducing an extrapropositional (epistemic) speech act adverb (Section 2.2), by integrating an explicit scale whose value is set to be "less than" the alternatives (Section 2.3), or by embedding the indefinite in the context of "high" negation (Section 2.4). The ensuing exposition is best construed as an empirical exercise, designed to fit the novel data into some general theoretical schemes. In other words, I am not necessarily making any analytical commitments – rather, I am using the existing theoretical apparatus to systematize the facts. Insofar as the proposals acquire an explicit shape (most notably in Section 2.4), I attempt no exhaustive treatment of the phenomena involved.

2.1 Default contexts

Wh-indefinites are robustly attested in conditional antecedents (10a) and embedded (or root) YN questions (10b). The affinity between conditionals and questions has been observed by multiple authors (for an overview and further references see Bhatt & Pancheva 2017). In fact, wh-indefinites turn up with admirable regularity in precisely these two contexts. Following standard practice, we may assume that the responsible licensing party here is a Q/conditional operator, merged in CP.

(10) a. Esli kto   / ⟨kto-nibud'⟩ pridet   v futbolke, vygonju!
          if   who.INDF   who-*nibud'* will.come in t-shirt   kick.out
          'If somebody shows up in a t-shirt, I'll kick them out!' (Twitter. 2019)

   b. ...proverjali, [smogut li kogo   / ⟨kogo-nibud'⟩ obmanut']?
          checked   are.able Q whom.INDF   whom-*nibud'* to.cheat
          '(They) checked whether they would be able to hoodwink
          somebody.'   (M. Semenova. *Volkodav.* 2003)

From (11) and (12), we glean that wh-indefinites are unhappy in deontic contexts, independent of the quantificational strength of the modal – universal in (11a) or existential in (11b). They are, however, compatible with epistemic modality, provided that the modal is of a possibility (12a) rather than of a necessity (12b) variety.

(11) a. Ty dolžen   s"est' čto-nibud'   / *s"est' čto.
          you must.M.SG to.eat what-*nibud'*   to.eat what.INDF
          '(I am not letting you out hungry). You must eat something.'

   b. Možeš'   posmotret' čto-nibud'   / ?*posmotret' čto.
          may.2.SG to.watch   what-*nibud'*   to.watch   what.INDF
          '(Because you behaved today), you may watch something.'

(12) a. Razmery   mogut   komu   / ⟨komu-nibud'⟩ i
          dimensions may.3.PL to.whom.INDF   to.whom-*nibud'* FOC
          prigodit'sja.
          to.be.of.use
          '(I am sharing this information, because) the dimensions might be of
          use to somebody.'   (car forum. 2017)

   b. Lekcija dolžna   ?*/??kogo   / kogo-nibud' zainteresovat'.
          lecture must.F.SG   whom.INDF whom-*nibud'* to.be.of.interest
          'The lecture must be of interest to somebody (though there are no
          guarantees of robust attendance).'

An enduring generalization that epistemics consistently outscope other sentential operators (including negation and root modals) formed the basis for formulating the analyses under which epistemics occupy a clause-peripheral position, high enough to take the widest scope (for various implementations see Butler 2003, Drubig 2001, Cormack & Smith 2002, a.o.).

   If so, the patterns above conform to the following generalization. A wh-indefinite licenser must be merged in a position presumably related to the (split) C-domain, which houses interrogative, conditional and epistemic operators. But

there is a further semantic requirement necessary for convergence: given the contrast in (12), the operator must be compliant with epistemic neutrality. While the possibility modal ensures epistemic equilibrium, the necessity one coaxes a stronger statement – one that is biased towards *p*. A similar effect is detectable in future contexts like (13).

Predictive future in Giannakidou & Mari (2013) is likewise nonveridical (since the outcome of the future event is unknown), but positively biased ("probably"), since it "presupposes confidence [of the speaker] that the actual world to come is a *p* world" (119).

(13) Zavtra   kto-nibud' / ⟨*kto⟩   sdelaet   obaldennoe kino [...].
     tomorrow who-*nibud'*   who.INDF will.make exciting   movie
     'Tomorrow somebody will produce an exciting film.' (kinometro.ru. 2012)

Clearly, universal epistemics and the future induce a similar effect: they appear to be too strong for wh-indefinites. In Giannakidou (and Mari)'s work this strength (to wit, bias) arises at the "not-at-issue" (presuppositional) level, defined as the speaker's measure of the likelihood of the event/actual-world-to-come. Epistemic *must* and the future come with a default positive bias, which can be modified by speech act adverbs: *Maybe John will come* expresses less confidence in the occurrence of the future event than its adverb-less counterpart.

Tampering with the default bias in Russian yields the following results. A YN question with *razve*, an element strictly specialized for non-neutral questions in (14a), conveys negative bias (i.e., the speaker believes that nobody had doubts about Putin's intentions). Likewise, in a future configuration (14b), the introduction of *avos'* 'perhaps, maybe' weakens the statement enough to render the indefinite appropriate in this context.

(14) a. Razve kto   somnevalsja, čto Putin ne   ujdet na pensiju?
        really who.INDF doubted   that Putin NEG leave on pension
        'Did anybody really doubt that Putin wouldn't retire?'   (dk.ru. 2020)

     b. *(Avos') zavtra   kto   poučastvuet.
          maybe tomorrow who.INDF will.participate
        'Maybe someone will participate tomorrow.'   (car forum. 2009)

Assuming a skeletal structure in (15), we may conclude that the indefinites are licensed by an element above TP – a high modal or an operator in SpecCP, but not by a root modal. Furthermore, this licenser must introduce epistemic equilibrium. In situations when it does not – i.e. when the default bias is skewed towards a positive proposition – a weakening adverbial, overriding the default bias, may

salvage the configuration (as demonstrated by the contrast between (13) and (14b)).

(15) $[_{CP} Op_{Q/Conditional} [_{FP} Mod_{Epis/Fut} [... Mod_{root} ...]]]$

As it turns out, however, the adverbial need not induce weaker bias. In the next section I demonstrate that speech act adverbs are legitimate licensers for wh-indefinites, independent of the direction of their bias.

2.2 Speech Act adverbs

Yanovich (2005), enlisting the paradigm in (16), concludes that wh-indefinites are not licensed by certain adverbs like *dolžno byt'* 'must be'.

(16)   a.   Možet, {kto   / ⟨✓kto-nibud'⟩} prixodil.
            maybe who.INDF   who-*nibud'* came.M
            'Maybe someone came.'
       b.   Dolžno byt', {*kto   / ⟨✓kto-nibud'⟩} prixodil.
            must   be   who.INDF   who-*nibud'* came.M
            'It must be the case that someone came.'   (Yanovich 2005)

Indeed, (16b) is bad, but a small adjustment in the word order, as in (17), renders the sentence perfectly natural if a bit quaint. I attribute this contrast to PF constraints to be discussed in Section 4. For now, it suffices to concede that an adverbial *dolžno byt'*, which, in contrast to the neutral *možet*, introduces a higher degree of speaker confidence, is in principle compatible with wh-indefinites.

(17) Dolžno byt', {prixodil   / prišel}   kto   (raz   takoj porjadok).
      must   be   came.IPFV came.PFV who.INDF since   such order
      'Someone must've stopped by, given how clean the place is.'

Speech act modal adverbs (SpMAs) like *dolžno byt'* and *možet (byt')* are distinct from the agreeing modals encountered in (11) and (12): The former are adjuncts, the latter are integral to a proposition. SpMAs have an immutable form and appear in the environments with inflected verbs. Agreeing modals carry phi-features and take on the infinitive complements. Furthermore, SpMAs (*probably, perhaps, certainly*, etc.) are said to express subjective modality in contrast to objective modality, routinely encoded by modal adjectives (*[it is] probable, certain, possible*, etc.) (Ernst 2009, Krifka 2022, Wolf 2015, a.o.).[4] The basic intuition here

---

[4] Modal verbs are frequently ambiguous between the two (see, e.g., Papafragou 2006 and references therein).

is that SpMAs convey speakers' internal judgment of/commitment to the embedded proposition. This is opposed to some external (objective) assessment of the event's likelihood. There are also more tangible correlates of subjective modality: modal adverbs are deviant in non-assertive environments like (18a) and resist negation, as in (19a) (cf. the grammatical counterparts with agreeing modals in (b)). Krifka (2017, 2022) provisions a special syntactic position for SpMAs – one that is external to the core proposition: for him, objective epistemics are associated with TP (hence, proposition-internal, at-issue), while subjective ones relate to the Judgment Phrase, a position above TP (hence, proposition-external, relaying not-at-issue content).

(18) a. * Pojdet li segodnja, {možet / dolžno byt'}, dožd'?
          will.go Q today   maybe   must be   rain
          Intended: 'Will it (maybe, certainly) rain today?'

      b. {Možet   / dolžen}   li segodnja pojti dožd'?
          may.2.SG   must.M.SG Q today   to.go rain
          'Might/ must it rain today?'

(19) a. {(*Ne) možet / (*Ne) dolžno byt'}, Ivan doma.
           NEG maybe   NEG must   be   Ivan home
          Intended: 'Ivan cannot be home.'

      b. Oni {ne mogut   byt' / ne dolžny byt'} doma.
          they NEG may.2.PL to.be   NEG must.PL to.be home
          'It is not possible/probable that they are home.'

If (12b) externalizes objective modality and (17) subjective modality, then the requirement for a weaker speaker commitment only holds of the former. The basic insight here is that universal (objective) epistemics, future, and veridical past contexts are too strong for wh-indefinites. But when explicitly tempered at the illocutionary level, these three contexts become just fine for indefinites, as demonstrated by the trio in (20) for each environment, respectively. Note that "tempering" is equivalent to embedding any subjective modification. SpMAs in (20) range from weak (*vrjad li*) to neutral (*možet (byt')*) to strong (*očevidno, dolžno byt'*). As elements of epistemic/evidential/inferential flavor, they form a natural class.

(20) a. (?) {Vrjad li / edva   li} segodnja gde   dolžen   pojti dožd'.
          hardly Q hardly Q today   where.INDF must.M.SG to.go rain
          'It is unlikely that it must rain somewhere today.'

b. Gljadiš', komu   i   prigoditsja.
    see.2.SG to.whom.INDF FOC will.be.of.use
    'Perhaps, (it'll) be of use to someone.'   (multiple sources)

c. {Stalo byt' / očevidno}, obidel   ee kto   (raz   plačet).
    come to.be   obviously hurt.PST her who.INDF since cries
    'Evidently/obviously, someone hurt her, since she is crying.'

Finally, it should be noted that the licenser in the illocutionary domain must be of an epistemic variety, as no other speech act adverbs – discourse-oriented (*čestno, vkratce*) or evaluative (*k ščast'ju, uvy*) – are compatible with wh-indefinites (or *nibud'*-indefinites):

(21) * {K ščast'ju / uvy / čestno   / vkratce}, prišel kto.
         to fortune alas honestly briefly   came   who
         Intended: 'Fortunately/alas/honestly/in brief, someone came.'

2.3 Role of the scale

The inaugural (2) would have us believe that subjunctives and matrix negation are licit licensers for wh-indefinites. This is not quite accurate. Desiderative and root subjunctives in (22), and negated factive verbs in (23), prove to be unfit for purpose.[5] By contrast, in the previously reported examples, the subjunctive (2c) imparts a meaning somewhat akin to English *lest*-clauses (to be discussed separately in Section 2.4), while negation in (2d) accompanies a matrix verb of the epistemic flavor.

(22) a. * Ja {xotel   / dobivalsja} togo, čtoby   kto   priexal.
           I   wanted   strove   that that.SBJV who.INDF came
           Intended: 'I wanted for (tried to get) somebody to come.'

     b. */?* Ja by   čto   sejčas posmotrel.
            I SBJV what.INDF now   watched
            Intended: 'I would watch something now.'

(23) * Ivan ne   {podtverdil / znal}, čto   prišel kto.
       Ivan NEG confirmed   knew   that came who.INDF
       Intended: 'Ivan didn't confirm / know that anybody came.'

---

[5]As an aside, *nibud'*-indefinites are perfect in (22) (though somewhat awkward in (23)).

In fact, matrix negation is not directly relevant – what matters is the type of the embedding predicate. As it turns out, wh-indefinites may be licensed under epistemic non-factives (*think*), emotives (*hope*), and dubitatives (*doubt*) – i.e., those verbs that in Romance are variable in selecting either subjunctive or indicative complements (Anand & Hacquard 2013, Farkas 1992); for arguments that they incorporate nonveridical components see Giannakidou & Mari (2016). Crucially, even in these contexts, wh-indefinites cannot simply appear "as is": they are most natural in the presence of a scalar adverb *xot'* "even, at least".

As for desideratives, a substantial body of work provisions a comparative semantics for the subjunctive-embedding attitudes (Anand & Hacquard 2013, Heim 1992, Villalta 2000, 2008, a.o.). Though the proposals vary in details, it will suffice for my purposes that *want*-type predicates introduce a scale, which orders the proposition expressed by the complement relative to the contextually supplied alternatives. Applying this to the contexts in (22), one may surmise that wh-indefinites are sensitive to preference ordering: they are incompatible with the contexts where the proposition is ranked as more desirable than the alternatives. Interestingly, desideratives, just like the attitude-embedding predicates, become wh-indefinite-friendly upon the introduction of *xot'*.

To sum up, though both non-factives and subjunctives are nonveridical (and hence, potential licensers), this alone is not sufficient for the felicity of wh-indefinites – as we will see, these contexts become appropriate for indefinites if they incorporate a bottom-of-the-scale condition. Moreover, this amelioration procedure is also available in imperative and iterative contexts (which, in the absence of scalar adverbials, are likewise incompatible with wh-indefinites).

I begin with the attitude verbs. Since (23) established that matrix negation is not a licenser for wh-indefinites, I suggested that the relevant factor is the type of the embedding predicate. The latter claim is ostensibly contradicted by the datasets in (24) and (25): while there is some speaker variation, none of my informants find wh-indefinites under *think*, *doubt* or *hope* (whether negated or not) fully acceptable.

(24)   The weather is awful today. People will probably choose to stay in.

    a.   {$^{?}$*Ne dumaju / $^{??}$ne poxože}, čto pridet   kto   na
            NEG think   NEG seems   that will.come who.INDF to
        sobranie.
        meeting
        Intended: '{I don't think that / It doesn't look like} anybody will show up to the meeting.'

b. ^{??} Somnjevajus', čto pridet   kto   na sobranie.
    doubt   that will.come who.INDF to meeting
    Intended: 'I doubt anybody will show up to the meeting.'

c. ^{?*/??} Ne nadejus', čto pridet   kto   na sobranie.
     NEG hope   that will.come who.INDF to meeting
     Intended: 'I doubt anybody will show up to the meeting.'

(25) The weather is delightful today. Surely, people will be inclined to get out.

a. ^{?*/??} {Dumaju / Poxože}, čto pridet   kto   na sobranie.
     think   seems   that will.come who.INDF to meeting
     Intended: '{I think that / It looks like} somebody will show up to the meeting.'

b. ^{?*/??} Ne somnjevajus', čto pridet   kto   na sobranie.
     NEG doubt   that will.come who.INDF to meeting
     Intended: 'I doubt anybody will show up to the meeting.'

c. ^{?*/??} Nadejus', čto pridet   kto   na sobranie.
     hope   that will.come who.INDF to meeting
     Intended: 'I don't doubt that somebody will show up to the meeting.'

Before I show how to improve (24) and (25), consider an apparent non-sequitur in (26), whose purpose will become clear in a moment. Though the imperatives provide a felicitous environment for wh-indefinites in a handful of Slavic languages and beyond (Haspelmath 1997), evidently they are not legitimate licensers for wh-indefinites in Russian. Kaufmann (2012) develops a modal semantics for imperatives, where *Eat your broccoli!* is roughly equivalent to *You must eat your broccoli.* For Condoravdi & Lauer (2012: 49), imperatives integrate the speaker's "preferential attitudes – including his wishes and desires", rendering the imperative operator broadly similar to *want.* If so, the ill-formedness of (26) with a bare *čto* follows from the same principles that inhibit the appearance of wh-indefinites in either deontic contexts like (11) or desiderative contexts like (22).

(26) Privezi   {*čto   / čto-*nibud'*} iz   Pariža!
     bring.IMP   what.INDF   what-*nibud'* from Paris
     'Bring [me] something from Paris!'

The reason for these detours is to do with a uniform procedure that converts all the listed bad contexts into good ones. To recap, the "bad" contexts for wh-indefinites include: (a) root/desiderative subjunctives in (22); (b) complements of affirmative and negated propositional attitude verbs, *think, doubt, hope* in (24)

and (25); (c) imperatives in (26). In all three environments, the degradedness disappears upon the introduction of a scalar adverb *xot'* 'at least, even', which evinces two properties. First, its associate is obligatorily focalized (Haspelmath 1997). Second, *xot'* is itself eligible only in non-assertive (i.e., nonveridical) situations. For example:

(27) a. *On xot' raz (ne) ezdil   v Pariž.
          he even once NEG travelled to Paris
          Intended: 'He (hasn't) traveled to Paris at least/even once.'

  b. On xot' raz ezdil   v Pariž?
      he even once travelled to Paris
      'Has he been to Paris even once?'

With this in place, observe a considerable transformation induced by *xot'* in all the iffy contexts (28): root and desiderative subjunctives in (28a)–(28b), the imperative in (28c), and the attitude predicate in (28d), all become quite natural when accompanied by *xot'*.

(28) a. Ty by xot' raz komu   peredaču snesla.
          you SBJV even once to.whom.INDF parcel   brought
          'You could've taken a care package to someone at least once.'
                                                (R. Pal'. *Cvety večnosti.* 1990)

  b. My dobivali's togo, čtoby   xot' stročku nam kto   napisal.
      we tried   that that.SBJV even line   to.us who.INDF wrote
      'We tried to get somebody to respond to us at least once.'

  c. Ty xot' slovo komu   napiši,   bezdel'nik!
      you even word to.whom.INDF write.IMP laggard
      'Write at least a word to somebody, you laggard!'

  d. On s   nadeždoj dumal, čto xot' raz ego kto   uslyšit.
      he with hope   thought that even once him who.INDF will.hear
      'He hoped that at least once someone will hear him.'

In fact, *xot'* need not be overt if the context is appropriate, as demonstrated by the *hope*-type predicate in (29).[6] In both cases of (29), the locative adverb *tam* is focalized, which ensures the identical interpretation of (29a) and (29b) even in the absence of an explicit *xot'*.

---

[6]The availability of the implicit *xot'* may be the source of speaker variation reported above, as well as the disagreement of my informants with the judgments recorded in Hengeveld et al. (2018). In fact, Reviewer 1 reports that in their judgment, wh-elements under *hope* are not possible whether with or without *xot'*.

(29) Context: 'It seems John is unlucky in his romantic pursuits. He never
      even had a date in our small town. But he's moving to New York soon, …'

   a. ✓/? Nadejus', čto [tam] kogo   vstretit.
         hope   that there who.INDF meets
         'I hope he meets somebody there (at least).'

   b. ✓/? Nadejus', čto xot' [tam] kogo   vstretit.
         hope   that even there who.INDF meets
         'I hope he meets somebody there at least.'

Russian *xot'* works just like the Greek variable scale *esto* 'even, at least' (Gian-
nakidou 2007). Giannakidou argues that *esto* carries a negative existential pre-
supposition and a bottom-of-the-scale condition. Unlike other types of *even*, *esto*
does not introduce the likelihood scale itself, but rather relies on the context to
supply one.

    The central take-away point here is sensitivity to scale: In potentially licens-
ing environments, wh-indefinites are possible only in the presence of a scalar
element which supplies a (contextual) bottom-of-the-scale condition. If so, itera-
tive contexts with frequency adverbs such as (30) likewise comply with this "less
than" requirement: wh-indefinites are only possible with negative frequency ad-
verbs in contrast to their *nibud'*-cousins, which are fine with both, *rarely* and
frequently.

(30) a. Znakomyx   u menja v Moskve mnogo […], no ja redko
        acquaintances at me   in Moscow lots   but I rarely
        {kogo   / ⟨✓ kogo-nibud'⟩} vižu.
         whom.INDF   whom-*nibud'* see
        'I've many acquaintances in Moscow, but I rarely see anybody.'
                                         (M.Bulgakov. Letters.)

   b. Ona často   {*kogo   / ✓ kogo-nibud'} rugaet.
        she frequently whom.INDF   whom-*nibud'* chides
        'She chides somebody frequently.'

Finally, I would be remiss not to point out one recurrent theme. In all the licens-
ing contexts discussed so far, wh-indefinites exhibit sensitivity to their syntactic
environments – in that the relevant licenser must be contained in the same clause
as the licensee.

2.4 "High" negation

Yanovich's subjunctive from (2c) belongs in the same semantic cluster as the
examples in (31). I will refer to them as LEST-clauses. LEST-clauses are special,

because they freely admit negative concord items (*nikto*) as well as *nibud'*-indefinites (this alternation is treated in Padučeva 2016).

(31) a. ... pribrala, čtoby   {kto   / ⟨✓kto-nibud'⟩ / ⟨✓nikto⟩} ne
           picked.up that.SBJV who.INDF   who-*nibud'*   *ni*.who NEG
           podnjal.
           took
           '(I deliberately) picked [it] up, lest somebody take it.'
                                            (M. Bulgakov. *Master i Margarita*. 1928–40)

      b. Szadi, čtoby   {kto   / ⟨✓kto-nibud'⟩ / ⟨✓nikto⟩} ne sbežal
           behind that.SBJV who.INDF   who-*nibud'*   *ni*.who NEG ran.away
           dorogoju, exali na konjax dva monaxa.
           en.route   rode on horses two monks
           'Two monks were riding astride behind [them] lest someone make a
           run for it en route.'   (Ju. German. *Rossija molodaja*. 1952)

The point of oddity is that *nibud'*-indefinites are not licensed by clausemate negation. Verbal negation in Slavic famously requires negative concord, as in (32).

(32) *{Kto-nibud' / ✓Nikto} ne sbežal   dorogoju.
        who-*nibud'*   *ni*.who NEG ran.away en.route

Concerning the meaning differences induced by *kto-nibud'* vs. *nikto* in, e.g., (31a), Padučeva offers the paraphrases in (33) and accepts the two as logically equivalent. She argues that although *nibud'*-indefinites appear in the scope of "global" negation (as opposed to local negation in cases of *nikto*), they are licensed by a nonveridical clausal operator.

(33) a. With *nikto*: I picked it up so that (it is the case that) nobody takes it.
      b. With *nibud'*: I picked it up so that it is not the case that somebody
          takes it.

In this specification for a negative outcome, LEST-clauses are akin (though not fully identical) to "apprehensive subjunctives" like (34). A handful of verbs, denoting surveillance/supervision/warning (*prismatrivat'* 'keep an eye', *karaulit'* 'guard', *bereč'sja* 'beware, be safe', *smotret'* 'watch (out)') or psych states of an unpleasant nature (*bojat'sja* 'be afraid', *trevožit'sja* 'be anxious', *volnovat'sja* 'be uneasy'), select a subjunctive clause headed by *kak (by)* (Nilsson 2012). In fact, the matrix verb may be altogether absent, in which case a bare *kak by*-clause (absolutely coherent as a stand-alone sentence) is understood as an implicit warning or expression of fear.

(34) (Smotri   / Bojus',) kak   by   {kto   / ⟨✓kto-nibud'⟩ / ⟨✓nikto⟩}
       watch.out.ɪᴍᴘ / fear   how sʙᴊᴠ who.ɪɴᴅғ   who-*nibud'*   *ni*.who
       telefon ne   stibril   v   takoj tolpe!
       phone ɴᴇɢ snatched in such crowd
       'Watch out lest someone snatch your phone in this crowd./ I fear
       someone might snatch your phone in this crowd.'

In addition to subjunctive morphology, ʟᴇsᴛ-clauses (31) and apprehensive sub-junctives (34) also pattern alike in syntax – by requiring verbal negation and admitting NCIs as well as *nibud'*- and wh-indefinites.[7] Such similarity, in turn, suggests that the two constructions may be eligible for a uniform analysis.

Complements of fear verbs are said to contain "expletive negation" (EN), al-leged to be devoid of polarity reversing semantics despite the compulsory real-ization of negation on the verb. The theoretical status of EN remains murky: there is no consensus on what *ne* in (34) actually does. Is it a semantically contentful element that moves to a high position within its clause to negate the evalua-tive mood (as in Abels 2005), or a mood marker licensed by nonveridicality (as in Yoon 2011), or a weak epistemic (as in Makri 2016), or simply a semantically empty exponent of morphosyntactic negation (as in Brown & Franks 1995)? My proposal is closer in spirit to Abels (2005) (and consistent with Padučeva's 2016 insight on "global" vs. "local" negation). Suppose that there are multiple merge sites for negation available in Russian, as in (35). The lower one (NegP2) negates events and delimits the exclusive domain of negative concord. The higher one (NegP1), introduced in the illocutionary field above TP, does not license NCIs, but it is compatible with bare wh-indefinites. If so, (31) and (34) are ambiguous between the two structures – and hence, enable a seemingly free alternation of the indefinites and NCIs.

(35) [... [$_{\text{NegP1}}$ ɴᴇɢ ... [$_{\text{TP}}$ ... [$_{\text{NegP2}}$ ɴᴇɢ [AspectP/vP ...]]]]]

There are also constructions that are not ambiguous between the two negations, shown in (36) and (37). The former, featuring an *until*-clause, is standardly clas-sified as another species of EN. The latter features expletive negation in a very literal sense – the negator here is a taboo word (*dick*, glossed as X.ɴᴇɢ). Neither construction tolerates NCIs.

---

[7]Contrary to the standard claim that NCIs do not embed under fear-predicates (e.g., Abels 2005, Brown & Franks 1995), many such examples are attested online. My informants likewise indicate that (34) with the NCI is perfectly on a par with the negative concord version of (31). See also Nilsson (2012) for further empirical adjudication.

(36) Uvjazneš po samye stupitsy i   zagoraeš, poka {kto   /
     stuck   to very hubs   and tan   until who.ɪɴᴅꜰ
     ⟨ ✓kto-nibud'⟩ / ⟨ *nikto⟩} ne   vytaščit.
        who-*nibud'*   ni.who ɴᴇɢ will.pull.out
     'Your hubs get stuck and you hang out until somebody pulls you out.'
                                        (O. Efremov. *Rybak primor'ja.* 2003)

(37) Xuj   {kto   / kto-nibud' / ⟨ *nikto⟩} prišel.
     X.ɴᴇɢ who.ɪɴᴅꜰ who-*nibud'*   ni.who came
     'It is not the case that anybody came.'   (Erschler 2023)

Per Abels, the matrix proposition and the *until*-clause in (36) cannot be true at the same time: One is either stuck, in which case the extricating event has not happened, or one is extricated, in which case they are no longer stuck. His proposal is that negation raises at LF to scope over the *poka*-clause, which precludes NCIs (as the licenser ceases to be sufficiently local). My amendment is that high negation merges directly in that position. Similarly, for X-negation in (37), Erschler (2023) argues that the negator sits in the Spec of the TP-external PolP that does not license NCIs.

The data are summarized in Table 1. I ascribed the alternation NCI ~ wh-indefinite in the first two entries to syntactic ambiguity stemming from the position of merge: The lower negation requires negative concord, the higher one supplies an appropriate context for wh-indefinites. Because the last two contexts do not tolerate NCIs, the negators in both instances must be introduced higher – above TP.

Table 1: Distribution of wh-indefinites and NCIs

	NCIs	wh-indefinites
ʟᴇꜱᴛ-clauses	✓	✓
apprehensive subjunctives	✓	✓
until-clauses	✗	✓
X-ɴᴇɢ	✗	✓

It should be noted that I do not envision a fixed position for "high negation" – indeed, its behavior in various contexts is consistent with multiple merge sites in the illocutionary domain. Since considerations of space prevent me from dealing with this topic in any coherent detail, I confine myself to a bare bones sketch of

the proposal, leaving the details of implementation or, indeed, a comprehensive justification for future endeavors. Assume the structure in (38), adopted from Krifka (2022), where ActP is the locus of assertions (•) or questions (?), ComP is the domain of the speaker's social commitments to the proposition, and the already familiar J(udgment)P is the province of subjective epistemic attitudes. For explicitness, I also assume that ComP can be headed by a null bouletic element (alternatively, one may posit an independent projection, representing bouletic attitudes of the speaker as in, e.g., Sode & Truckenbrodt 2018).

(38) $[_{ActP} [_{Act^0} •][_{ComP} [_{Com_{BOUL}}] [_{JP} ...[_{TP} ...]]]]$

Given the above, I suggest that X.NEG and "global" negation in LEST-clauses apply at the level of ActP, which furthermore must contain an assertorial operator to render it consistent with Erschler's observation that X.NEG is impossible in questions (39).

(39)   * Xren   on xodil na rabotu?
        X.NEG he went to work
        Intended: 'Did he not go to work?'

On the other hand, in apprehensive subjunctives under *kak by* and *until*-clauses, negation appears lower – at the level of ComP or TP. That the NEG of a LEST-clause is distinct from the NEG of an apprehensive subjunctive/*until*-clause is confirmed by (40): X.NEG can replace *ne* in a LEST-clause (40a) but not the lower *ne* of the two EN contexts in (40b) and (40c).

(40)   a.   ..., čtoby   xren kto   sbežal   dorogoju, ...
             that.SBJV X.NEG who.INDF ran.away en.route
             '... so that it is not the case that somebody escapes en route, ...'

        b. * Bojus', kak by xren   kto   telefon stibril.
             fear   how SBJV X.NEG who.INDF phone snatched
             Intended: 'I don't want for anybody to steal the phone.'

        c. * ... poka xren   kto   vytaščit.
             until X.NEG who.INDF will.pull.out
             Intended: '... until someone pulls (us) out.'

The exposition is undeniably terse here, but the essential insight should be reasonably clear: Wh-indefinites are licensed by a negative operator, residing in the illocutionary domain. This distinguishes wh-indefinites from NCIs, whose felicity is predicated on the presence of a proposition-internal operator.

3 Intermediate summary

Wh-indefinites are possible in polar interrogatives (neutral or biased), conditionals (indicative or hypothetical) and under existential epistemic modals. While the future, episodic past and modal environments (with universal epistemics) are "too strong", they can be made compatible with wh-indefinites by manipulating subjective modality (i.e., by merging an epistemic speech act adverbial). Desiderative and root subjunctives, attitude predicates, iterative contexts and imperatives likewise create "potentially licensing" contexts – only in these situations, the felicity of wh-indefinites is parasitic on the presence of a scalar adverb (encoding a bottom-of-the-scale condition). Finally, wh-indefinites are happy under high (illocutionary) negation.

The lessons here are two. First, there are no contexts that license wh-indefinites to the exclusion of *nibud'*-indefinites. In fact, the requirements of the latter are substantially less stringent: *nibud'*-indefinites are perfectly acceptable with no additional conditions in desiderative, future, iterative, etc. contexts. In the interest of full disclosure, consider also (41), which shows that in contrast to wh-indefinites, *nibud'*-indefinites are fine in both the scope and the restriction of a universal.[8] Conversely, wh-indefinites are routinely banned in universally quantified contexts, independent of the quantifier's syntactic role (subject or object), its surface position or, indeed, its type (*vse* 'all', *oba* 'both', *každyj* 'each' are all deviant with wh-indefinites). Furthermore, my informants are reluctant to accept wh-indefinites even when a quantifier is embedded in an otherwise wh-indefinite-friendly environment, such as a polar interrogative in (42).

(41)   a.   Každyj {⟨*čto⟩   / čto-nibud'} slyšal o korole Arture.
            each   what.INDF what-*nibud'* heard of king   Arthur
            'Everybody heard something about king Arthur.'

                                                 (bookstore blurb, modified)

       b.   Každyj, kto   {*komu   / komu-nibud'}   zaviduet,
            each   who.REL to.whom.INDF to.whom-*nibud'* envies
            obladaet nizkoj samoocenkoj.
            possesses low   self-esteem
            Intended: 'Everyone who envies somebody has low self-esteem.'

(42)   Razve vse studenty {*/?*čto   / čto-nibud'} pročitali?
       really all students   what.INDF what-*nibud'* read
       'Didn't all students read something?'

---

[8]I refer the reader to Padučeva (2007), Pereltsvaig (2008) for discussion of Russian *nibud'*-indefinites in quantified contexts.

The second point concerns a recurrent locality issue. The felicitous contexts require a clausemate licenser of the relevant kind – adverbs aside, all other environments feature an operator associated with the C-domain. For instance, in (43), with the operator in the superordinate clause, wh-indefinites are unacceptable.

(43) a. Razve on govoril, čto {*kto   / kto-nibud'} sdal   ėkzamen?
          really he said   that who.INDF who-*nibud'* passed exam
          'Didn't he say that someone passed the exam?'

     b. Esli najti   v Rossii čeloveka, kotoryj {⟨$^{?}$*čto⟩   / čto-nibud'}
          if   to.find in Russia person   who   what.INDF what-*nibud'*
          sdelal v   pol'zu UNSO, to … .
          did   for benefit UNSO
          'If one were to find a person who did something to benefit UNSO,
          then [he might be prosecuted].'   (gazeta.ru. 2014)

This locality constraint is intuitively logical. While both local and distant licensers require full morphological specification, the medial one enables the spell-out of a bare indefinite, provided the environment is sufficiently negative. In other words, we may conceive of the polarity-sensitive pronouns as a hierarchy of sorts, i.e. *kto-nibud'* » *kto(-nibud')* » *nikto*, where *nibud'* is compatible with (almost any) nonveridical operator (medial or distant), *ni*-items are required under a local antiveridical operator, and wh-indefinites are somewhat in the middle – possible in a subset of nonveridical environments in close proximity to their licenser. This "intermediate" (and morphologically sterile) status also correlates with certain PF-related effects to be discussed in the next section.

4 Syntax-PF interactions

That wh-indefinites are crosslinguistically de-focalized is not a revelation (e.g., Haida 2008, Hengeveld et al. 2022). Hengeveld et al. (2022), in fact, state the requirement as a biconditional: wh-elements ("quexistentials" in Hengeveld et al.'s terminology) are obligatorily focalized in their interrogative interpretation; in their existential incarnation, on the other hand, they are never focalized.[9] What I will attempt to show here is that Russian wh-indefinites are not simply unable to bear contrastive focus: indeed, they are considerably fussier in selecting surface positions than other indefinites. The basic observation is that in addition

---

[9]Reviewer 2 points out that the first clause of this biconditional is falsified by Czech (see Šimík 2010) and perhaps Slovenian (Mišmaš 2017).

to resisting contrastive focus, wh-indefinites prefer to be adjacent to the element that realizes the main sentential stress. This property, along with a preference for clustering in a specific order as well as resistance to coordination, render them akin to clitics.

Sentential stress here is understood as in Yokoyama (1987), who argues that there are two basic types of intonation in Russian: Type 1 (neutral) and Type 2 ("expressive"), shown in (44a) and (44b), respectively. Type 1 entails an iterating sequence of intonational phrases with LH contour, accompanied by a downstep. The "new information" (or, in more familiar terms, an element bearing information focus) comes at the end with a falling (HL) contour, which basically corresponds to a neutral declarative sentence with falling intonation. This is demonstrated in (45a).[10] Obviously, "new information" need not be restricted to a single lexical item – an entire constituent may function in this manner. In Type 2 (45b), the fronted constituent (*doždiček*) realizes sentential stress, defined as the "stress which marks the knowledge item that would occur in utterance-final position, were the same sentence to be uttered with intonation Type 1 instead" (Yokoyama 1987: 191). Its properties are twofold: (i) It is the last intonational center of the utterance, and (ii) No rising tones can follow it. Abstracting away from the pitch details, '↘' will be used to indicate Type 1 intonation, '∗' (and small caps) to mark Type 2 intonation, and '|' to identify phonological phrase boundaries (call it πP).

(44)   a.   Nad Krakovom nakrapyval doždiček.
            over Krakow   drizzled   rain
            'The rain was drizzling over Krakow.'

       b.   Nad Krakovom Doždiček nakrapyval.
            over Krakow   rain   drizzled

(45)   a.   Nad Krakovom | nakrapyval | doždiček.
                 LH |ᶦ LH | HL (↘)

       b.   Nad Krakovom | Doždiček nakrapyval.
                 LH | HL (∗)

Yokoyama (1987) also shows that indefinite pronouns are ineligible to realize the final HL under neutral Type 1 intonation. Instead, the intonational core shifts to a "fully specified" constituent. In the case of (46), it is the verb. While the indefinite is ineligible to serve as the default intonational pivot here, it can be pronounced with the contrastive (Type 2) contour.

---

[10]This representation is borrowed from King (1993).

(46) Pojdemte kuda-nibud'.   (Type 1 with indefinites)
     lets.go   where-*nibud'* ($\searrow$)
     'Let's go somewhere.'

With these preliminaries in place, consider Yanovich's (2005) data in (47) again (repeated from (16)). Earlier it was established that a modal adverb like *dolžno byt'* is a legitimate licenser for wh-indefinites after all, provided its licensee complies with certain word order restrictions.

(47) a. Možet, {kto}   prixodil {kto}.
        maybe who.INDF came.M   who.INDF
        'Maybe someone came.'

     b. Dolžno byt', {*kto}   prixodil {kto}.
        must   be   who.INDF came.M   who.INDF
        'It must be the case that someone came.'

Focusing here on the intransitive verbs, consider the subject position permutations with *dolžno byt'*. A neutral sequence in (48a) requires a postverbal subject, which takes on the default Type 1 accentuation. On the other hand, (48b) is marked: now, the scrambled (contrastively focused) subject carries Type 2 sentential stress.

(48) a. Dolžno byt', | umerla | koroleva.
        must   be | died | queen   ($\searrow$)
        'The queen must've died.'

     b. Dolžno byt', | KOROLEVA umerla.
                          *

On the other hand, the *nibud'*-indefinite in (49) can be placed either before or after the verb – but in either case *umer* serves as the default prosodic center of its prosodic phrase, i.e. both (49a) and (49b) display the Type 1 pattern, where the intonational pivot shifts along with the verb. The Type 2 scheme, found in (48b), is difficult to get for *nibud'*-indefinites. For whatever reason, in these contexts the *nibud'*-item resists contrastive focalization.

(49) a. Dolžno byt', | umer kto-nibud'.
        must   be | died who-*nibud'* ($\searrow$)
        'Somebody must've died.'

     b. Dolžno byt', | kto-nibud' umer.
        must   be   who-*nibud'* died ($\searrow$)

Finally, consider (50). The sentence is parsed into two πPs. Immediately excluded are instances like (50c) with the Type 2 (contrastive focus) intonation. The two incarnations of Type 1 prosody in (50a) and (50b) correspond to (49a) and (49b), respectively. I ascribe the deviance of (50b) to the convergence of two factors: the indefinite sits in the πP-initial position to the left of the element realizing default declarative prosody.

(50) a. Dolžno byt', | prixodil kto.
                               HL ($\searrow$)

     b. [?]*Dolžno byt', | kto prixodil.
                                      HL ($\searrow$)

     c. *Dolžno byt', | кто prixodil.
                      *

Taken independently, these two contingencies are no impediment for wh-indefinites. For instance, *možet* in (51) does not require a prosodic boundary after itself, which, in turn, ensures that the indefinite is not stranded in the initial position. In this situation, the indefinite can be left- or right-adjacent to the default prosodic host. Note that a heavier constituent – like *možet byt'* in (52) – is tougher to integrate into the utterance: with a pause after the adverbial, the indefinite feels awkward in the preverbal slot.

(51) a. Možet, umer kto.
         maybe died who.INDF ($\searrow$)

     b. Možet, kto   umer.
         maybe who.INDF died ($\searrow$)

(52) [?]*/^{??}Možet byt', | kto   prišel.
           may be   who.INDF came ($\searrow$)
     Intended: 'Maybe someone came.'

Conversely, (53) shows that a wh-indefinite may appear in the utterance-initial position but only if its host carries a non-default intonational contour, as is the case in the polar interrogative context schematized in (53b).

(53) a. Kto   PRIŠEL?
         who.INDF came
         'Did somebody come?'

     b. Kto PRIŠEL?
             *

The analytical payoff here is this. Wh-indefinites can appear neither in the positions of information focus (like other indefinites) nor in the positions of contrastive focus (unlike other indefinites). Additionally, they must obey certain added restrictions, which curb their presence in the πP-initial positions. These two properties suggest that the elements in question are of a special nature. In the remainder of this section I identify a few additional quirks of wh-indefinites that attest to their clitic-like qualities.

First, multiple wh-indefinites are possible in principle. In such situations, the indefinites prefer the sequence NOM >> DAT/ACC >> adjuncts. Violations are perceived to be non-lethal – certainly not on the level of ordering infractions in languages with pronominal clitics, yet my informants are consistent in their dislike for the alternative orders. Examples are found in (54).

(54) a. Videl li {kto   kogo   / [?]kogo   kto}   včera?
    saw Q who.INDF whom.INDF   whom.INDF who.INDF yesterday
    'Did someone see anybody yesterday?'

    b. Kak by {kto   gde   / ^{?/??}gde   kto}   ne
    how SBJV who.INDF where.INDF   where.INDF who.INDF NEG
    zastrjal!
    get.stuck
    '(I am afraid) someone might get stuck somewhere.'

Second, multiple wh-indefinites tend to form a cluster, as demonstrated by an embedded YN question in (55) and a conditional in (56). Under the most natural reading, in the deviant examples, the verbs (i.e., *razboltal* and *rasskažet*) form the prosodic core in the relevant intonational domains. The oddity of (55a) and (56a) follows from the non-adjacency of one of the indefinites to its (verbal, in this case) host. There are, however, strategies that improve split clusters. For instance, if *Ivan* from (56a) receives contrastive focus in the manner of (57), the sentence becomes rather natural. In other words, while the default configuration is one in which the indefinites form a bundle, split clusters are possible if the indefinites in question are adjacent to the appropriate host.

(55) a. *[?] (Ja ne znaju,) razboltal li komu   Ivan čto,   no vse
      I NEG know blabbed Q to.whom.INDF Ivan what.INDF but all
    uže   znajut naš sekret.
    already know our secret
    Intended: 'I don't know if Ivan blabbed something to someone, but everybody already knows our secret.'

b. Ja ne znaju, razboltal li komu   čto   Ivan,...
    I NEG know blabbed Q to.whom.INDF what.INDF Ivan

(56) a. ?? Esli komu   Ivan čto   rasskažet, ja budu v jarosti.
         if to.whom.INDF Ivan what.INDF will.tell   I will.be in fury
         Intended: 'If Ivan tells anybody anything, I will be furious.'

   b. Esli Ivan komu   čto   rasskažet, ...
       if Ivan to.whom.INDF what.INDF will.tell

(57) ✓ Esli komu IVAN čto rasskažet,...
              *

The third property requires a small digression. It is a well-established fact that multiple wh-phrases can be coordinated in Russian in the manner of (58a) (e.g., Gribanova 2009). The other cases in (58) are, perhaps, less famous (data are due to Paperno 2012). Paperno shows that the conjuncts in such configurations must be of the same type (i.e., indefinite+indefinite, universal+universal, etc.), cf. a mismatched indefinite+universal in (59).

(58) a. Kto i   kogo videl?
         who and who saw
         'Who saw whom?'

   b. Nikto i   nikogo ne pobedil.
      ni-who and *ni*-whom NEG defeated
      'Nobody defeated anybody.'

   c. Ponjal   li kto-nibud' i čto-nibud'?
      understood Q who-*nibud'* and what-*nibud'*
      'Did anybody understand anything?'

(59) * Ponjal   li kto-nibud' i vse?
        understood Q who-*nibud'* and everything
        Intended: 'Did anybody understand everything?'

The phenomenon of hybrid coordination is poorly understood (and I have nothing to add about the syntax of these structures). But whatever the mechanism, it is clearly unavailable to wh-indefinites: the pattern in (60) corroborates that wh-indefinites are not fantastic when coordinated, whereas in the absence of the coordinator (61), they are well-formed. One alternative (with a precedent in the literature) is to attribute their resistance to coordination to PF reasons.

(60)   a.   ?*/?? Možet, on komu   i   čto   privezet iz   Pariža.
            maybe he to.whom.INDF and what.INDF will.bring from Paris
            Intended: 'Maybe somebody will bring something from Paris.'

       b.   ?*/?? Esli on komu   i   čto   privezet,...
            if   he to.whom.INDF and what.INDF will.bring
            Intended: 'If he brings anything for anybody,...'

(61)   a.   Možet, on komu čto privezet iz Pariža.

       b.   Esli on komu čto privezet,...

Said precedent is found in Stepanov & Moussaoui (2020), who argue that in
Lebanese Arabic, *ʃu* 'what' evinces clitic-like properties – one of which, they
suggest, is resistance to coordination, as shown in (62) (this also holds of the
French *que* 'what').

(62)   a.   * ʃu   w   min bta-ʕref   b-hal-balad?
            what and who 2.SG-know in-this-country
            'What and who(m) do you know in this country?'

       b.   Amta w   kif ʕam t-rooħ-o   ʕa-l-masbaħ?
            when and how PROG 2-going-PL to-the-swimming.pool
            'When and how are you going to the swimming pool?'

                                                            (Lebanese Arabic)

Needless to say, coordinating "real" clitics (e.g., Bosnian/Croatian/Serbian (BCS)
pronominal clitics in (63)) is out of the question. Though the violations in (60)
are not nearly as bad as in BCS (63), my informants are consistent in assigning
the value ≤ 3 (out of 5) to the coordinated indefinites.

(63)   * Poklonila sam mu i   ga.
         gifted   am him and it
         Intended: 'I gifted it to him.'   (BCS)

All of this is to say that although coordination should not be taken as the sole
diagnostic of clitichood, when combined with earlier observations on the de-
pendent prosodic status of wh-indefinites, it may prove to be of explanatory
value.[11] Considering also that the mechanism is available to all types of quan-
tified elements in Russian, the baseline assumption would situate wh-indefinites

---

[11]See Citko (2013) and references therein for a discussion of syntactic factors involved in deriving
   non-standard coordination dependencies.

within the same array of elements that are amenable to coordination in principle. Their outlier behavior is hence best accommodated by appealing to their peculiar prosodic status.

In short, the contention here is that wh-indefinites are reminiscent of (albeit not fully tantamount to) clitics. They require adjacency to a prosodic host, but show flexibility in alignment (left or right). They form clusters which can nevertheless be broken under the right conditions. Within clusters, they tend to appear in a particular order, the violations of which are merely dispreferred (rather than fully unacceptable). Finally, unlike other quantified elements, they are far from ideal when coordinated. All of these properties, in turn, suggest that the binary division into clitic vs. non-clitic is too rigid. There must be room to accommodate items like wh-indefinites, which are not quite clitics proper but neither are they tonic forms. In other words, clitic⟷non-clitic represents a scale, with elements occupying various intermediate positions within this continuum.[12]

5 Conclusion

Wh-indefinites are "not quite" elements: not quite clitics, they require a weakly negative context, created by a clausebounded operator. They can always be replaced with *nibud'*-indefinites, but not *vice versa*. This "in-between" status correlates with bare morphology: while very local (antiveridical) and superordinate (nonveridical) operators call for full morphological specification (*ni-* or *nibud'*, respectively), the medial ones admit such morphologically deficient elements under certain circumstances. Though I have attempted to catalog what these circumstances are, it would be obviously desirable to uncover a unifying semantic mechanism that ensures the felicity of wh-indefinites in all the contexts from Section 2. Dwelling on the topic of further desideratum, it would be productive to establish specific phonetic correlates that underlie the weak prosodic status of non-polar elements within the clitic–non-clitic continuum.

With these caveats aside, the basic findings are as follows. First, wh-indefinites are possible in a proper subset of *nibud'*-indefinites. Encountered most frequently in polar interrogatives, in conditional antecedents, and under weak epistemic verbs, they can also be introduced in desiderative/root subjunctives, imperatives, iterative and future contexts, as well as under strong epistemics and attitude predicates. However, all the latter ("non-standard") contexts require further modification to render the indefinites happy – either a (subjective) epistemic or a

---

[12]Reviewer 1 points out that this property renders them rather akin to weak pronouns in the sense of Cardinaletti & Starke (1999)

scalar adverb. Second, Russian wh-indefinites occupy a peculiar PF niche: not only do they resist contrastive focalization (a well-established fact), they evince additional properties consistent with the typical behavior of clitics. Said properties include their preference for clustering (and a specific order within the clusters), their selectivity of hosts, and their inability to coordinate.

Abbreviations

2	second person		PFV	perfective
3	third person		PL	plural
FOC	focus		PST	past
IMP	imperative		PROG	progressive
INDF	indefinite		Q	question marker
IPFV	imperfective		REL	relative
M	masculine		SBJV	subjunctive
NEG	negation		SG	singular

Acknowledgments

References

Abels, Klaus. 2005. "Expletive negation" in Russian: A conspiracy theory. *Journal of Slavic Linguistics* 13(1). 5–74. https://www.jstor.org/stable/24599547.

Anand, Pranav & Valentine Hacquard. 2013. Epistemics and attitudes. *Semantics and Pragmatics* 6. 1–59. DOI: 10.3765/sp.6.8.

Arsenijević, Boban. 2009. {Relative {conditional {correlative clauses}}}. In Anikó Lipták (ed.), *Correlatives cross-linguistically*, 131–156. Amsterdam: John Benjamins. DOI: 10.1075/lfab.1.06ars.

Belyaev, Oleg & Dag Haug. 2020. The genesis and typology of correlatives. *Language* 96(4). 874–907. DOI: 10.1353/lan.2020.0065.

Bhat, Darbhe Narayana Shankara. 2004. *Pronouns*. Oxford: Oxford University Press.

Bhatt, Rajesh & Roumyana Pancheva. 2017. Conditionals. In Martin Everaert & Henk C. van Riemsdijk (eds.), *The Wiley Blackwell companion to syntax (second edition)*, 1–48. Hoboken: John Wiley & Sons. DOI: 10.1002/9781118358733.wbsyncom119.

Brown, Sue & Steven Franks. 1995. Asymmetries in the scope of Russian negation. *Journal of Slavic Linguistics* 3(2). 239–287. https://www.jstor.org/stable/24598950.

Bruening, Benjamin. 2007. Wh-in-situ does not correlate with wh-indefinites or question particles. *Linguistic Inquiry* 38(1). 139–166. DOI: 10.1162/ling.2007.38.1.139.

Butler, Jonny. 2003. A minimalist treatment of modality. *Lingua* 113(10). 967–996. DOI: 10.1016/S0024-3841(02)00146-8.

Caponigro, Ivano & Anamaria Fălăuş. 2022. The semantics of Rudin constructions in Romanian. In Giorgio Sbardolini Marco Degano Tom Roberts & Marieke Schouwstra (eds.), *Proceedings of the 23rd Amsterdam Colloquium*, 55–61. Amsterdam: ILLC.

Cardinaletti, Anna & Michal Starke. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In Henk van Riemsdijk (ed.), *5 clitics in the languages of Europe (Volume 5): Clitics in the languages of Europe (part 1)*, 145–234. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110804010.145.

Chierchia, Gennaro & Hsiu-Chen Daphne Liao. 2015. Where do Chinese wh-items fit? In Luis Alonso-Ovalle & Paula Menéndez-Benito (eds.), *Epistemic indefinites: Exploring modality beyond the verbal domain*, 31–59. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199665297.003.0002.

Citko, Barbara. 2009. What don't wh-questions, free relatives, and correlatives have in common? In Anikó Lipták (ed.), *Correlatives cross-linguistically*, 49–79. Amsterdam: John Benjamins. DOI: 10.1075/lfab.1.04cit.

Citko, Barbara. 2013. The puzzles of wh-questions with coordinated wh-pronouns. In Theresa Biberauer & Ian Roberts (eds.), *Challenges to linearization*, 295–329. Berlin: Mouton de Gruyter. DOI: 10.1515/9781614512431.295.

Condoravdi, Cleo & Sven Lauer. 2012. Imperatives: Meaning and illocutionary force. In Christopher Piñón (ed.), *Empirical Issues in Syntax and Semantics 9*, 37–58. Paris: CNRS. http://www.cssp.cnrs.fr/eiss9/.

Cormack, Annabel & Neil Smith. 2002. Modals and negation in English. In Sjef Barbiers, Frits Beukema & Wim van der Wurff (eds.), *Modality and its interaction with the verbal system*, 133–163. Amsterdam: John Benjamins. DOI: 10.1075/la.47.08cor.

Drubig, Hans Bernhard. 2001. *On the syntactic form of epistemic modality*. Ms. University of Tübingen. https://www.lingexp.uni-tuebingen.de/sfb441/b2/papers/DrubigModality.pdf.

Ernst, Thomas. 2009. Speaker-oriented adverbs. *Natural Language & Linguistic Theory* 27(3). 497–544. DOI: 10.1007/s11049-009-9069-1.

Erschler, David. 2023. Colloquial emphatic negation in Russian and morphology of negative concord. *Journal of Slavic Linguistics* 31(FASL 30 issue). 1–18. http://ojs.ung.si/index.php/JSL/article/view/171.

Farkas, Donka. 1992. On the semantics of subjunctive complements. In Paul Hirschbühler & E.F.K. Koerner (eds.), *Romance languages and modern linguistic theory: Selected papers from the XX Linguistic Symposium on Romance languages, University of Ottawa*, 69–104. Amsterdam: John Benjamins. DOI: 10.1075/cilt.91.07far.

Fitzgibbons, Natalia. 2010. *Licensers and meanings: Structural properties of dependent indefinites.* Storrs: University of Connecticut. (Doctoral dissertation).

Gärtner, Hans-Martin. 2009. More on the indefinite-interrogative affinity: The view from embedded non-finite interrogatives. *Linguistic Typology* 13(1). 1–37. DOI: 10.1515/LITY.2009.001.

Giannakidou, Anastasia. 2007. The landscape of EVEN. *Natural Language & Linguistic Theory* 25(1). 39–81. DOI: 10.1007/s11049-006-9006-5.

Giannakidou, Anastasia. 2013. Inquisitive assertions and nonveridicality. In Maria Aloni, Michael Franke & Floris Roelofsen (eds.), *The dynamic, inquisitive, and visionary life of φ: A Festschrift for Jeroen Groenendijk, Martin Stokhof, and Frank Veltmanof*, 115–126. Amsterdam: ILLC.

Giannakidou, Anastasia & Jing Lin. 2016. *The Mandarin NPI shenme is not exhaustive: A reply to Chierchia and Liao (2015).* University of Chicago. Ms. https://home.uchicago.edu/~giannaki/pubs/Final.LinGianna.Mar.11.2016.pdf.

Giannakidou, Anastasia & Alda Mari. 2013. A two dimensional analysis of the future: Modal adverbs and speaker's bias. In Maria Aloni, Michael Franke & Floris Roelofsen (eds.), *Proceedings of the 19th Amsterdam Colloquium*, 115–122. Amsterdam: ILLC.

Giannakidou, Anastasia & Alda Mari. 2016. Emotive predicates and the subjunctive: A flexible mood ot account based on (non)veridicality. In Nadine Bade, Polina Berezovskaya & Anthea Schöller (eds.), *Proceedings of the 20th Sinn und Bedeutung*, 288–305. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/issue/view/7.

Gribanova, Vera. 2009. Structural adjacency and the typology of interrogative interpretations. *Linguistic Inquiry* 40(1). 133–154. DOI: 10.1162/ling.2009.40.1.133.

Haida, Andreas. 2008. The indefiniteness and focusing of question words. In Tova Friedman & Satoshi Ito (eds.), *Semantics and Linguistic Theory 18*, 376–393. DOI: 10.3765/salt.v18i0.2510.

Haspelmath, Martin. 1997. *Indefinite pronouns.* Oxford: Oxford University Press.

Heim, Irene. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9(3). 183–221. DOI: 10.1093/jos/9.3.183.

Hengeveld, Kees, Sabine Iatridou & Floris Roelofsen. 2018. *Quexistentials I.* Handout, UvA & MIT. http://lingphil.scripts.mit.edu/papers/iatridou/Quexistentials_I.pdf.

Hengeveld, Kees, Sabine Iatridou & Floris Roelofsen. 2022. Quexistentials and focus. *Linguistic Inquiry.* 1–54. DOI: 10.1162/ling_a_00441.

Izvorski, Roumyana. 1996. The syntax and semantics of correlative proforms. In Kiyomi Kusumoto (ed.), *Proceedings of the 26th North East Linguistics Society*, 133–147. Amherst: GLSA. https://scholarworks.umass.edu/nels/vol26/iss1/11.

Kaufmann, Magdalena. 2012. *Interpreting imperatives.* Dordrecht: Springer.

King, Tracy Holloway. 1993. *Configuring topic and focus in Russian.* Stanford: CSLI.

Kratzer, Angelika & Junko Shimoyama. 2017. Indeterminate pronouns: The view from Japanese. In Chungmin Lee, Ferenc Kiefer & Manfred Krifka (eds.), *Contrastiveness in information structure, alternatives and scalar implicatures* (Studies in Natural Language and Linguistic Theory 91), 123–143. Leiden: Springer. DOI: 10.1007/978-3-319-10106-4_7.

Krifka, Manfred. 2017. *Assertions and judgements, epistemics, and evidentials.* Paper given at *Speech Acts: Meanings, Uses, Syntactic and Prosodic Realizations.* Handout. https://www.leibniz-zas.de/fileadmin/Archiv2019/veranstaltung_zas/workshops/samusar/Krifka.pdf.

Krifka, Manfred. 2022. Layers of assertive clauses: Propositions, judgements, commitments, acts. In Jutta M. Hartmann & Angelika Wöllstein (eds.), *Propositional arguments in cross-linguistic research: Theoretical and empirical issues.* Tübingen: Narr.

Lin, Jo-Wang. 2014. Wh-expressions in Mandarin Chinese. In C.-T. James Huang, Y.-H. Audrey Li & Andrew Simpson (eds.), *The handbook of Chinese linguistics*, 180–207. Wiley Online Library. DOI: 10.1002/9781118584552.ch8.

Lin, Jo-Wang. 2020. Correlatives. *The Wiley Blackwell Companion to Semantics.* 1–25. DOI: 10.1002/9781118788516.sem062.

Liu, Mingming & Yu'an Yang. 2021. Modal wh-indefinites in Mandarin. In Patrick Georg Grosz, Luisa Martí, Hazel Pearson, Yasutada Sudo & Sarah Zobel (eds.), *Proceedings of the 25th Sinn und Bedeutung*, 581–599. DOI: 10.18148/sub/2021.v25i0.955.

Makri, Maria Margarita. 2016. What 'not' might mean. Expletive Negation in attitude contexts. *Rivista di Grammatica Generativa* 38. 187–200. https://api.semanticscholar.org/CorpusID:73710352.

Mišmaš, Petra. 2017. Slovenian questions with short wh-movement and the low periphery. *Slovenski jezik/Slovene Linguistic Studies* 11. 111–126. DOI: 10.17161/1808.25034.

Nilsson, Nadezhda Zorikhina. 2012. Peculiarities of expressing the apprehensive in Russian. *Oslo Studies in Language* 4(1). 53–70. DOI: 10.5617/osla.164.

Padučeva, Elena V. 2007. *O semantike sintaksisa: Materialy k transformacionnoj grammatike russkogo jazyka.* Second edition. Moskva: URSS.

Padučeva, Elena V. 2016. Mestoimenija tipa *čto-nibud'* v otricatel'nom predloženii. *Voprosy jazykoznanija* (3). 22–36. DOI: 10.31857/S0373658X0001000-9.

Papafragou, Anna. 2006. Epistemic modality and truth conditions. *Lingua* 116(10). 1688–1702. DOI: 10.1016/j.lingua.2005.05.009.

Paperno, Denis. 2012. *Semantics and syntax of non-standard coordination.* Los Angeles: UCLA. (Doctoral dissertation).

Pereltsvaig, Asya. 2008. Russian *nibud'*-series as markers of co-variation. In Natasha Abner & Jason Bishop (eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics*, 370–378. Somerville: Cascadilla Proceedings Project. https://www.lingref.com/cpp/wccfl/27/paper1852.pdf.

Postma, Gertjan. 1994. The indefinite reading of wh. *Linguistics in the Netherlands* 11(1). 187–198. DOI: 10.1075/avt.11.19pos.

Rudin, Catherine. 2007. Multiple wh-relatives in Slavic. In Richard Compton, Magdalena Goledzinowska & Ulyana Savchenko (eds.), *Proceedings of the 15th Formal Approaches to Slavic Languages*, 282–307. Ann Arbor: Michigan Slavic Publications.

Šimík, Radek. 2009. Hamblin pronouns in modal existential wh-constructions. In Jodi Reich & Darya Kavitskaya (eds.), *Proceedings of the 17th Formal Approaches to Slavic Languages*, 187–202. Ann Arbor: Michigan Slavic Publications.

Šimík, Radek. 2010. Interpretation of multiple interrogatives: An information structure sensitive account. In Wayles Browne, Adam Cooper, Alison Fisher, Esra Kesici, Nikola Predolac & Draga Zec (eds.), *Proceedings of the 18th Formal Approaches to Slavic Languages*, 486–501. Ann Arbor: Michigan Slavic Publications.

Šimík, Radek. 2017. Existential wh-constructions. In Mark Aronoff (ed.), *Oxford bibliographies in linguistics, vol. 10.* DOI: 10.1093/OBO/9780199772810-0162.

Sode, Frank & Hubert Truckenbrodt. 2018. Verb position, verbal mood, and root phenomena in German. In Mailin Antomo & Sonja Müller (eds.), *Non-canonical verb positioning in main clauses*, 91–135. Hamburg: Helmut Buske.

Stepanov, Arthur & Ali Al Moussaoui. 2020. When a wh-word refuses to stay in situ. *Linguistic Inquiry* 51(2). 410–423. DOI: 10.1162/ling_a_00345.

Villalta, Elisabeth. 2000. Spanish subjunctive clauses require ordered alternatives. In Brendan Jackson & Tanya Matthews (eds.), *Semantics and Linguistic Theory 10*, 239–256. DOI: 10.3765/salt.v10i0.3123.

Villalta, Elisabeth. 2008. Mood and gradability: An investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy* 31(4). 467–522. DOI: 10.1007/s10988-008-9046-x.

Wolf, Lavi. 2015. *Degrees of assertion.* Beersheba: Ben Gurion University of the Negev. (Doctoral dissertation).

Yanovich, Igor. 2005. Choice-functional series of indefinite pronouns and Hamblin semantics. In Efthymia Georgala & Jonathan Howell (eds.), *Semantics and Linguistic Theory 15*, 309–326. DOI: 10.3765/salt.v15i0.2921.

Yokoyama, Olga T. 1987. *Discourse and word order.* Amsterdam: John Benjamins.

Yoon, Suwon. 2011. *"Not" in the Mood: The syntax, semantics, and pragmatics of Evaluative Negation.* Chicago: The University of Chicago. (Doctoral dissertation).

Name index

Advances in formal Slavic linguistics 2022

Advances in formal Slavic Linguistics 2022 brings together a collection of 22 articles originating as talks presented at the 15th Formal Description of Slavic Languages conference (FDSL 15) held in Berlin on 5–7 October, 2022. The contributions cover a broad spectrum of topics, including clitics, nominalizations, l-participles, the dual, verbal prefixes, assibilation, verbal and adjectival morphology, lexical stress, vowel reduction, focus particles, aspect, multiple wh-fronting, definiteness, polar questions, negation words, and argument structure in such languages as BCMS, Bulgarian, Czech, Macedonian, Polish, Russian, Slovenian, Ukrainian, and Upper Sorbian. The wide range of topics explored in this volume underscores the diversity and complexity of Slavic languages. The contributions not only advance our understanding of languages belonging to the Slavic group but also offer fresh perspectives for linguistics more broadly.