

Advances in Formal Slavic Linguistics 2022

Edited by

Berit Gehrke

Denisa Lenertová

Roland Meyer

Daria Seres

Luka Szucsich

Joanna Zaleska

Open Slavic Linguistics

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

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. Marušić, 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*.

ISSN (print): 2627-8324

ISSN (electronic): 2627-8332

Advances in Formal Slavic Linguistics 2022

Edited by

Berit Gehrke

Denisa Lenertová

Roland Meyer

Daria Seres

Luka Szucsich

Joanna Zaleska

Berit Gehrke, Denisa Lenertová, Roland Meyer, Daria Seres, Luka Szucsich & Joanna Zaleska (eds.). 2025. *Advances in Formal Slavic Linguistics 2022* (Open Slavic Linguistics). Berlin: Language Science Press.

This title can be downloaded at:

`redefine\lsURL`

© 2025, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

ISBN: no digital ISBN

no print ISBNs!

ISSN (print): 2627-8324

ISSN (electronic): 2627-8332

no DOI

Cover and concept of design: Ulrike Harbort

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: X_EL^AT_EX

`redefine \publisherstreetaddress`

`redefine \publisherurl`

Storage and cataloguing done by `redefine \storageinstitution`

Contents

Preface	iii
1 Russian verbal stress retraction as induced unstressability Ora Matushansky0000-0002-1420-7091	3
2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian Olav Mueller-Reichau	53
3 Polar questions in Czech and Russian: An exploratory corpus investigation Maria Onoeva & Anna Staňková	85
4 Dual preservation in Slovenian: The verb supports the noun in semi-spontaneous production Matic Pavlič, Artur Stepanov & Penka Stateva	103
5 Word prosodic structure and vowel reduction in Moscow and Perm Russian Margje Post0000-0003-3959-5905	121
6 Morphosemantic mismatches with pronouns as a consequence of their internal structure Zorica Puškar-Gallien0000-0003-4130-0598	145
7 Animacy influences segmental phonology: The velar–sibilant alternation in BCMS Marko Simonović0000-0002-9651-6399	187

Contents

8 How departicipial are “L-participle” nominalisations in Western South Slavic	
Marko Simonović0000-0002-9651-6399, Petra Mišmaš0000-0001-8659-875X, Stefan Milosavljević0000-0003-2305-2519, Boban Arsenijević0000-0002-1124-6319, Katarina Gomboc Čeh, Franc Lanko Marušič0000-0002-0667-3236 & Rok Žaucer0000-0001-7771-6937	209
9 Focus-sensitive particles in Bulgarian: Towards an adverbial-only analysis	
Carla Spellerberg	239
10 The Western South Slavic verbal suffix <i>-nV/-ne</i> is a diminutive affix with a theme vowel	
Ema Šarkl0000-0002-1328-8395, Marko Simonović0000-0002-9651-6399, Stefan Milosavljević0000-0003-2305-2519 & Boban Arsenijević0000-0002-1124-6319	273
11 Wh-indefinites in Russian	
Ksenia Zanon	309
Index	342

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, January 20, 2025

Berit Gehrke

Denisa Lenertová

Roland Meyer

Daria Seres

Preface

Luka Szucsich
Joanna Zaleska

0

Chapter 1

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). 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 deletion of a vowel retains its accentual specification. 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 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 thematic suffix, the tense suffix and agreement morphology.¹ The thematic suffix is

¹The transcriptions below closely follow Russian orthography and do not indicate: (i) palatalization before front vowels (/Ci/ → [C'i], /Ce/ → [C'e]), (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 /í/ (the front yer) and /ü/ (the back yer). The letters *u* (IPA [fç], see Padgett & Žygis 2007), *w* (IPA [s]), *ꝝ* (IPA [z]), *ꝑ* (IPA [çç], *ꝑ* (IPA [ts]) are traditionally rendered as č, š, ž, šč, and c.

Ora Matushansky

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 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 suffixed accents). The consistent post-stem stress 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).

²See Antonyuk et al. (2022–2023) for a range of opinions.

1 Russian verbal stress retraction as induced unstressability

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: -lub- ‘love’	l ^j ubl ^j -ú	l ^j ub-í-t	l ^j ub-í-l-a	l ^j ub-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 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 is a result of 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 a conflicted 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 §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.

§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

Ora Matushansky

vocalic suffixes. My explanation of the latter fact will be introduced in section §4: I will suggest that in the first conjugation the 1SG pattern arises from induced unstressability of the present-tense suffix.

§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. §6 provides such an explanation by ascribing the unstressability of the present-tense suffix to accentual conflict: 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.

§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-*, as shown in Table 2.

Table 2: Verbal conjugations, present-tense paradigms: *nestí* ‘to carry’, *vinít’* ‘to blame’

first conjugation		second conjugation	
singular	plural	singular	plural
1. nes-e-u → nesú	nes-e-m → nes'óm	vin-i-Ø-u → viníú	vin-i-Ø-m → viním
2. nes-e-šj → nes'óšj	nes-e-te → nes'óte	vin-i-Ø-šj → viníšj	vin-i-Ø-te → viníte
3. nes-e-t → nes'ót	nes-e-nt → nesút	vin-i-Ø-t → vinít	vin-i-Ø-nt → viníát

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 §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

1 Russian verbal stress retraction as induced unstressability

-l-), while the verb *vinít^j* ‘to blame’ contains the thematic vowel *-i-*, see Table 3. 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	<i>nes-l-ú</i> → <i>n^jós</i>		<i>vin-i-l-ú</i> → <i>viníl</i>	
F	<i>nes-l-a</i> → <i>neslá</i>	<i>nes-l-i</i> → <i>neslí</i>	<i>vin-i-l-a</i> → <i>viníla</i>	<i>vin-i-l-i</i> → <i>viníli</i>
N	<i>nes-l-o</i> → <i>nesló</i>		<i>vin-i-l-o</i> → <i>vinílo</i>	

The past tense (historically, the active past participle form) is segmentally uncontroversial, 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 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.

³I 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 *-ti-* is resolved for various consonants (see Lightner 1965, 1972). The alternation between the surface back vowel with a palatalized preceding consonant ([^jo]) under stress and the phonological /e/ in unstressed syllables in Table 2, Table 3 and elsewhere is allophonic (Lightner 1969, Boyd 1997).

⁴The 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 [m^ja] (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.

Ora Matushansky

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-ɛ-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.

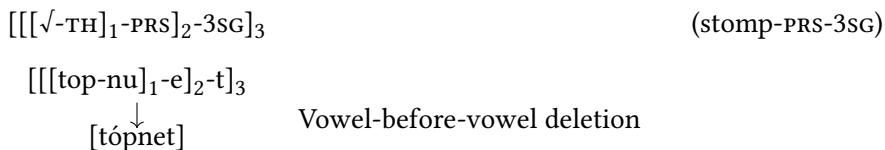


Figure 1: 3SG derivation

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

⁵Examples (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.

⁶Melvold (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.

1 Russian verbal stress retraction as induced unstressability

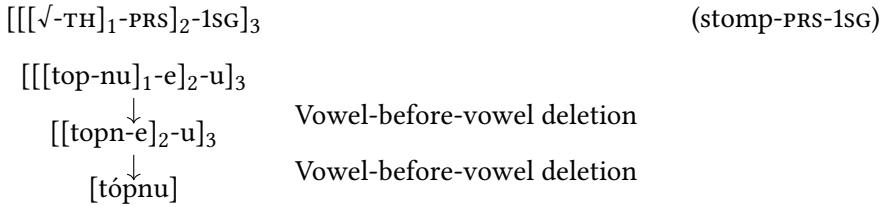


Figure 2: 1SG derivation

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 §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 (1989), Idsardi (1992), Halle & Idsardi (1995) and Halle (1997), where each syllable 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. $\begin{array}{c} (* \quad *) \\ zim \quad u \\ \downarrow \end{array}$
- b. $\begin{array}{ccc} * & & \\ (* \quad *) & & \\ zim \quad u & \rightarrow & zímu 'winter.SG.ACC' \end{array}$

Ora Matushansky

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 parenthesis 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. $\overset{*}{(\underset{*}{\text{lez}} \underset{*}{\text{l}} \underset{*}{\text{i}})}$

b. $\underset{*}{\text{klad}} \underset{*}{\text{l}} \underset{*}{\text{a}}$

(surface *klalá*)

c. $\underset{*}{\text{klad}} \underset{*}{\text{l}} \underset{*}{\text{i}}$

(surface *kláli*)

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, as 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)

1 Russian verbal stress retraction as induced unstressability

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. sekretár^j/sekretar^já/sekretar^jámi ‘secretary.SG.NOM/SG.GEN/PL.INS’

(10) a. zdoróv/zdorová/zdorovó/zdoroví ‘robust.F/M/N/PL’
b. t^jaž^jól/t^jaželá/t^jaželó/t^jaželí ‘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:

⁷The 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.). I will not attempt to address this discussion here.

Ora Matushansky

- (11) a. nominative singular

* * (*)
bu lav a

- b. genitive plural

* * ()
bu lav ū

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) as 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.

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-* as 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.

⁸Yet 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.

⁹The fifth class consists of just two verbal roots, *-mog-* (*močj* ‘to be able’) and the cranberry root *-im-/n'a-* (e.g., *prin'âtj* ‘to accept’) and their derivatives, which exhibit the 1SG pattern. I return to this matter in §7.1.1.

1 Russian verbal stress retraction as induced unstressability

Table 4: Accentual interaction in athematic ($\sqrt{-T-\phi}$) verbs

		accented PST-F.SG	unaccented PST-PL	accented PST-3SG	
a.	stem: <i>-lez-</i> 'climb'	A	<u>l</u> éz-l <u>a</u>	<u>l</u> éz-l-i	<u>l</u> éz-e-t
b.	post-stem: <i>-nes_-</i> 'carry'	PA+	nes_-l <u>á</u>	nes_-l-i	nes_- <u>j</u> ó-t
c.	variable (past): <i>-klad-</i> 'put'	UA	kla-l <u>á</u>	klá-l-i	klad- <u>j</u> ó-t
d.	retracting (past): <i>-griz_-</i> 'gnaw'	PA-	gríz_-l <u>a</u>	gríz_-l-i	griz_- <u>j</u> ó-t

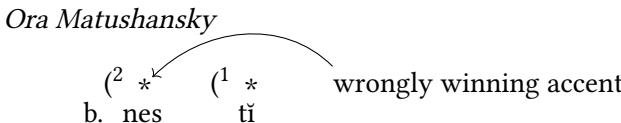
While Melvold (1989) implements this process by moving the relevant parenthesis one syllable to the left, Halle (1997) handles it by inserting a parenthesis before the preceding syllable.

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. lézt^j 'to climb', klást^j 'to put', grízt^j 'to gnaw'
 b. unesená 'carried away.F.SG',
 cf. perelézena 'climbed over', progrízena 'gnawed through', sprádena
 'spun'

The Basic Accentuation Principle (2) predicts that in a sequence of a post-accenting and a pre-accenting morphemes 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. nes tí



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.	griz l a ↓
	b.	(*) griz l a ↓ (*)
	c.	griz l a

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:

¹⁰This 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 will be absent from the tonal tier, and unstressable ones, from the metrical tier.

¹¹The 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.

1 Russian verbal stress retraction as induced unstressability

- (15) a. klad- e- m → klad^jóm
 put PRS 1PL

- b. klad- e- te → klad^jóte
 put PRS 2PL

- (16) a. nes_ - e- m → nes^jóm
 put PRS 1PL

- b. nes_ - e- te → nes^jóte
 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- ~~e~~- 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 *-jо-* (palatalizing [o]) under stress, and the 3PL suffix (which I take to be

Ora Matushansky

-*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 (-*ü*-) 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.

The fact that first-conjugation verbs (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 §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.

1 Russian verbal stress retraction as induced unstressability

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: <i>-top-</i> ‘stomp’	tóp-n-e-t	tóp-n-u	tóp-n-u-l-a	tóp-n-u-l-i
b. post-stem: <i>-max-</i> ‘wave’	max-n ^j -ó-t	max-n-ú	max-n-ú-l-a	max-n-ú-l-i
c. 1SG: <i>-obman-</i> ‘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: <i>-vid-</i> ‘see’	víd-i-t	víž-u	víd-e-l-a	víd-e-l-i
b. post-stem: <i>-vel-</i> ‘order’	vel-í-t	vel ^j -ú	vel-é-l-a	vel-é-l-i
c. 1SG: <i>-vert-</i> ‘spin’	vért-i-t	verč-ú	vert-é-l-a	vert-é-l-i

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 on 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

¹²One 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.,

Ora Matushansky

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 §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, Id-sardi 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 1sg pattern; for non-productive thematic classes the numbers given represent the number of unprefixed verbs in that class.¹³)

¹³ *Igalá/lgáli* ‘lied.F.SG/PL’). In the present the suffix is undetectable (*lgu/lž'ot* ‘lie.1SG/3SG’) due to hiatus resolution before the vocalic present-tense suffix.

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] arose from the underlying [dj]), *žáždatj* ‘to thirst’ and the substandard *stráždatj* ‘to suffer’.

- (i) a. *tájatj/táju/táet* ‘melt.INF/1SG/3SG’

- b. *stráždatj/stráždu/stráždet* ‘suffer.INF/1SG/3SG’ (Modern Russian
stradátj/stradáju/stradáet, literary variant with the *-a-/i-* thematic suffix

1 Russian verbal stress retraction as induced unstressability

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 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

stradát^j/stráždu/stráždet)

19 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's/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.

Ora Matushansky

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 §7.1.1) and with the second-conjugation *-e-* (five verbal roots, see §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 §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: when the thematic vowel is deleted, a sequence of two accents is created that should be resolved into one:¹⁵

- (19) a. $\sqrt{-\text{nu}} \text{ e } \text{t}$
 TH PRS 3SG
 ↓
 * ((*)
 b. $\sqrt{-\text{ny}} \text{ e } \text{t}$
 TH PRS 3SG
 ↓

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’, *počin’ú/počinit* ‘repair.1SG/3SG’ (1SG pattern) vs. the literary *učin’ú/učinít* ‘initiate.1SG/3SG’, *vozbužú/vozbudít* ‘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.

1 Russian verbal stress retraction as induced unstressability

- c. $\begin{array}{cccc} * & (*) \\ \sqrt{-} & n & e & t \end{array}$
 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. $\begin{array}{cccc} * & (*(*) \\ \sqrt{-} & n & e & t \end{array}$
 TH PRS 3SG
 ↓
 b. $\begin{array}{cccc} * & (*) \\ \sqrt{-} & n\emptyset & e & t \end{array}$
 TH PRS 3SG
 ↓
 c. $\begin{array}{cccc} * & (*) \\ (\sqrt{-} & n & e & t \end{array}$
 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 [i]:)¹⁶

- (21) a. vert-í ‘spin.IMP’, vert^j-á ‘spin.GER’ (cf. verčú/vértit ‘spin.1SG/3SG’)
 b. obman-í ‘cheat.IMP’ (cf. obmanú/obmánet ‘cheat.1SG/3SG’)
 c. l^jub-í ‘love.IMP’, l^jub^j-á ‘love.GER’ (cf. l^jubl^jú/l^júbit ‘love.1SG/3SG’)

¹⁶Stress in the active present participle generally patterns with non-1SG, but sometimes doesn’t (e.g., *ucús^j/účits'a* ‘study.1SG/3SG’ vs. *ucásčij^js'a* ‘studying.M.SG’, see also [Zaliznjak 1985: 29, 77](#)).

Ora Matushansky

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
 ↓
 * ((* *))
 √- nȳ e u
 TH PRS 3SG
 ↓
 * ((*))
 √- nȳ e 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 (§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 §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.

1 Russian verbal stress retraction as induced unstressability

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. $\begin{array}{cccc} * & (*) & * \\ \sqrt{-} & \text{nu} & \text{e} & \text{u} \\ \text{TH} & \text{PRS} & \text{3SG} \\ \downarrow & & & \end{array}$
- b. $\begin{array}{cccc} * & (*) & * \\ \sqrt{-} & \text{ny} & \text{e} & \text{u} \\ \text{TH} & \text{PRS} & \text{3SG} \\ \downarrow & & & \end{array}$
- c. $\begin{array}{cccc} * & (*) & * \\ \sqrt{-} & \text{ny} & \text{e} & \text{u} \\ \text{TH} & \text{PRS} & \text{3SG} \\ \downarrow & & & \end{array}$

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 §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. $\begin{array}{cccc} * & (*) &) \\ \sqrt{-} & \text{nu} & \text{e} & \text{t} \\ \text{TH} & \text{PRS} & \text{3SG} \\ \downarrow & & & \end{array}$
- b. $\begin{array}{cccc} * & (*) &) \\ \sqrt{-} & \text{ny} & \text{e} & \text{t} \\ \text{TH} & \text{PRS} & \text{3SG} \\ \downarrow & & & \end{array}$
- c. $\begin{array}{cccc} (* & (\\ \sqrt{-} & \text{ny} & \text{e} & \text{t} \\ \text{TH} & \text{PRS} & \text{3SG} \\ \downarrow & & & \end{array}$

The final issue to be resolved is that of the 2PL ending *-te-*, which is wrongly predicted to be stressed:

Ora Matushansky

(25)	a.	$\sqrt{-\text{ny}}$	e	te	$*$
		TH	PRS	2PL	
		↓			$*$
	b.	$\sqrt{-\text{ny}}$	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 (§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 Es'kova 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)

1 Russian verbal stress retraction as induced unstressability

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 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: *kolebát^j/kolébl^ju* ‘rock.INF/1SG’, *kolixát^j/kolíšu* ‘sway.INF/1SG’, *alkát^j/álču* ‘crave.INF/1SG’, the archaic variant *stradát^j/stráždu* ‘suffer.INF/1SG’ and the two equally archaic pre-fixed derivatives of the cranberry root *-im-*, *vnimát^j/vném^jl^ju* ‘heed.INF/1SG’ and *prinimát^j/priéml^ju* ‘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 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 *vorkotáj* ‘to grumble’, *topotáj* ‘to stamp’ and *trepetáj* ‘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) can 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.

¹⁷The 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).

Ora Matushansky

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: <i>-maz-</i> ‘smear’	máž-e-t	máž-u	máz-a-l-a	máz-a-l-i
b.	post-stem: N/A; potential candidates among <i>j</i> -final stems can be assigned to the <i>-a-/Ø-</i> class				
c.	1SG: <i>-v^jaz-</i> ‘tie’	v ^j áž-e-t	v ^j áž-ú	v ^j az-á-l-a	v ^j az-á-l-i
d.	stem-final present: <i>-koleb-</i> ‘rock’	kolébl ^j -e-t	kolébl ^j -u	koleb-á-l-a	koleb-á-l-i

- (28) a. tópat^j ‘to stamp, tramp’: tópaju (1SG)/tópajet (3SG)
 b. topotát^j ‘to stamp’: topočú (1SG)/topočet (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.

- (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

¹⁸None of these nouns naturally forms a plural, which excludes this way of checking their accentuation. The fact that the post-accenting diminutive suffix *-ük-* derives a post-accenting noun (*xoxotók*) is also non-indicative (cf. Halle 1973: 340).

¹⁹See Melvold (1989) for the assumption that *-ow-* combines with the thematic suffix *-a-/i-* and

1 Russian verbal stress retraction as induced unstressability

- (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., *komádovat^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. *razzálobljú/razzálubit* ‘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. *bombljú/bombít* ‘bomb.1SG/3SG’ (inflection)
 (cf. *bómба/bómbu* ‘bomb.NOM/ACC’)

- (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 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: 147–147).

Ora Matushansky

- a. prizeml^jú/prizemlít ‘land.1SG/3SG’ (stem)
(cf. zeml^já/zémľú ‘ground.NOM/ACC’)
- b. poruč^jú/porúčit ‘entrust.1SG/3SG’ (1SG)
(cf. ruká/rúku ‘hand.NOM/ACC’)
- c. boron^jú/boroní ‘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; 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

- (34) a. accented:
uród/uródi ‘ugly person.SG.NOM/PL.NOM’ → uródovat^j ‘to disfigure’
- b. accented:
kritíka/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’

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 *-č-*, where stress is final (i.a), and the same result is obtained with the post-accenting diminutive suffix *-čk-* (i.b). Since, however, both suffixes contain yers, which are known to be unstressable, these cases are also non-definitive.

- (i) a. dvor/dvori ‘yard.SG.NOM/PL.NOM’
dvorc^j/dvorcí ‘palace.SG.NOM/PL.NOM’
- b. kazák/kazákí ‘Cossack.SG.NOM/PL.NOM’
kazačák/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 *-čk-* yields stress on the first one (e.g., the unaccented root *-vetír-* ‘wind’ yields a post-accenting simple diminutive *veterák/veterká* ‘wind.DIM.NOM/GEN’ and a stem-accented double diminutive *veteróček/veteróčka* ‘wind.DIM.NOM/GEN’).

1 Russian verbal stress retraction as induced unstressability

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 §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, in function of the fate of the thematic vowel.

Ora Matushansky

6 The 1sg pattern as an accentual conflict

I have suggested (§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 (§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 (§2.3). Likewise, the lack of accentual variability in the past tense of most thematic verbs (§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 (§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. $\begin{array}{ccccc} * & * & [& & (*) \\ \checkmark & -nu & l & a & \\ TH & PST F.SG & & & \\ & & \downarrow & & \\ * & [* & & (*) \\ \checkmark & -nu & l & a & \\ TH & PST F.SG & & & \end{array}$
- b. $\begin{array}{ccccc} * & [* & & (*) \\ \checkmark & -nu & l & a & \\ TH & PST F.SG & & & \end{array}$

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

1 Russian verbal stress retraction as induced unstressability

the right parenthesis at the right edge of the word deleted in (38b) because it is immediately preceded by another parenthesis (see §2.2):²¹

- (38) a.

TH PRS 3SG
- b.

TH PRS 3SG
- c.

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 (38b): the same syllable cannot be simultaneously accented and post-accenting. I propose that the conflicted position is deleted from the metrical tier (38c).

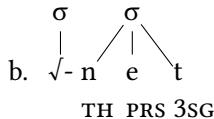
Once again, as no stress-bearing elements are contained between the two parentheses in (38c), the second parenthesis is deleted (39a). Because the remaining parenthesis ends up word-final, stress, like with post-accenting stems, will surface on the final syllable of the stem (39b):

- (39) a.

TH PRS 3SG

²¹The 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.

Ora Matushansky



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-*, after the deletion of the thematic vowel (40b) and the removal of the present-tense suffix from the metrical tier (40c), the present-tense suffix is deleted before another vowel (40d) and its accent is realized on the vowel of the 1SG suffix:

- (40) a. $\begin{array}{cccc} * & * & [& * \\ \sigma & \sigma & \sigma & \sigma \\ | & | & | & | \\ \sqrt{-} & nu & e & u \end{array}$
 TH PRS 1SG
 \downarrow
 $\begin{array}{cc} * & (*[* \\ \sigma & \sigma \\ | & | \\ \sqrt{-} & \cancel{nu} & e & u \end{array}$
 TH PRS 1SG
 \downarrow
 $\begin{array}{cc} * & (* \\ \sigma & \sigma \\ | & | \\ \sqrt{-} & \cancel{nu} & e & u \end{array}$
 TH PRS 1SG
 \downarrow
 $\begin{array}{cc} * & (* \\ \sigma & \sigma \\ | & | \\ \sqrt{-} & \cancel{nu} & e & u \end{array}$
 TH PRS 1SG
 \downarrow
 $\begin{array}{cc} * & (* \\ \sigma & \sigma \\ | & | \\ \sqrt{-} & n & \cancel{e} & u \end{array}$
 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:

1 Russian verbal stress retraction as induced unstressability

- (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. Moskv ič a
 Moscow NMLZ SG.GEN
- ↓
- * ((* p *)
- b. Moskv ič a
 Moscow NMLZ SG.GEN
- ↓ *
- * * (*)
- c. Moskv ič a
 Moscow NMLZ SG.GEN

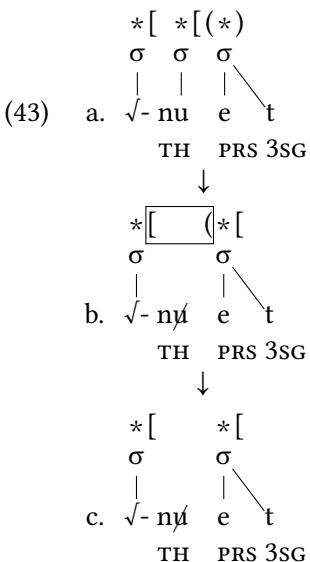
I will continue to use the enriched representation with square brackets because I believe that it not only encodes a valuable intuition about the source of the unstressability of the present-tense suffix, but also makes it possible to explain how this effect is nullified when the L-stem is post-accenting (see §5 on the correlation between unaccented *-ot-* verbs with the 1SG pattern and post-accenting *-ow-* verbs with the retracting pattern). More specifically, in section §6.1 I will show how post-stem stress is correctly predicted for post-accenting L-stems, and in section §6.2, how *-a/-i-* verbs with post-accenting L-stems give rise to the stem-final pattern alternating with the 1SG pattern.

Ora Matushansky

6.1 The role of a post-accenting stem

As discussed in §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 now shown, 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:



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

²²If 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).

1 Russian verbal stress retraction as induced unstressability

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 §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 §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 in Table 9 below.

Table 9: Accentual interaction with the 1st conjugation suffix *-a/-i-*

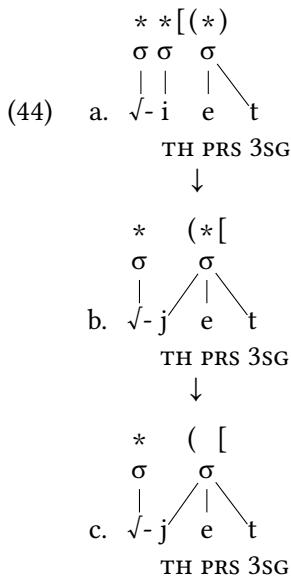
	accented PRS-3SG	accented PRS-1SG	accented PST-F.SG	unaccented PST-PL
a. stem: <i>-maz-</i> ‘smear’	máž-e-t	máz-u	máz-a-l-a	máz-a-l-i
b. post-stem: N/A; potential candidates among <i>j</i> -final stems can be assigned to the <i>-a/-Ø-</i> class				
c. 1SG: <i>-v^jaz-</i> ‘tie’	v ^j áž-e-t	v ^j až-ú	v ^j az-á-l-a	v ^j az-á-l-i
d. stem-final present: <i>-koleb-</i> ‘rock’	kolébl ^j -e-t	kolébl ^j -u	koleb-á-l-a	koleb-á-l-i

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,

Ora Matushansky

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 thematic vowel before another vowel gives rise to an accentual conflict (38b). The conflicted 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).



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 conflicted position, as in (45b–45c)):

- * [*(*)] 1 Russian verbal stress retraction as induced unstressability
- (45) a.
$$\begin{array}{c} \sigma \quad \sigma \quad \sigma \\ | \quad | \quad | \\ \sqrt{-} \quad i \quad e \quad t \end{array}$$

TH PRS 3SG
↓

$$\begin{array}{c} *[\quad](*)[\\ \sigma \quad \sigma \\ | \quad | \quad | \\ \sqrt{-} \quad j \quad e \quad t \end{array}$$

TH PRS 3SG
↓

$$\begin{array}{c} [* \quad (*)] \\ \sigma \quad \sigma \\ | \quad | \quad | \\ \sqrt{-} \quad j \quad e \quad t \end{array}$$

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.
$$\begin{array}{c} [* \quad (\quad [\\ \sigma \quad \sigma \\ | \quad | \quad | \\ \sqrt{-} \quad j \quad e \quad t \end{array}$$

TH PRS 3SG
↓
e.
$$\begin{array}{c} [* \quad (\quad [\\ \sigma \quad \sigma \\ | \quad | \quad | \\ \sqrt{-} \quad j \quad e \quad t \end{array}$$

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.

Ora Matushansky

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-i-t	govor-i-l-a	govor-i-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

1 Russian verbal stress retraction as induced unstressability

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-e]₂-i]₃-t]₄ → [[[gor-∅]₂-i]₃-t]₄ → [gorit] (vowel deletion)
 b. [[[gor-e]₂-∅]₃-t]₄ → [[[gor-i]₂-∅]₃-t]₄ → [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 10-c) 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. l^jub i ∅ t
 ✓ TH PRS 3SG
 ↓
 * (* [
 σ σ
 // //
 b. l^jub i ∅ t
 ✓ TH PRS 3SG
 ↓
 * ([
 σ σ
 // //
 c. l^jub i ∅ t
 ✓ TH PRS 3SG

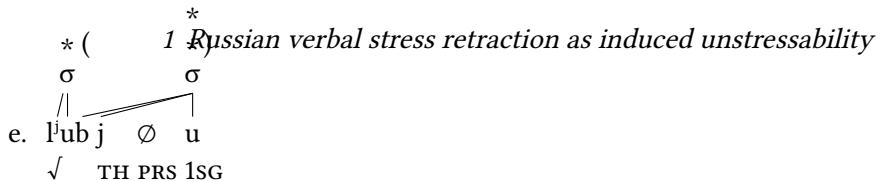
Ora Matushansky

In the 1SG, on the other hand, the conflicted 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?

- (48) a. $\begin{array}{cccc} * & (& * &) \\ \sigma & \sigma & \sigma \\ / & / & | \\ \text{l}^{\text{j}}\text{ub} & \text{i} & \emptyset & \text{u} \\ \checkmark & \text{TH PRS 1SG} \\ \downarrow \\ * & (& * &) \\ \sigma & \sigma & \sigma \\ / & / & | \\ \text{l}^{\text{j}}\text{ub} & \text{i} & \emptyset & \text{u} \\ \checkmark & \text{TH PRS 1SG} \end{array}$

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 conflicted 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 conflicted 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.

- (48) c. $\begin{array}{cccc} * & (& * &) \\ \sigma & \sigma & \sigma \\ / & / & | \\ \text{l}^{\text{j}}\text{ub} & \text{i} & \emptyset & \text{u} \\ \checkmark & \text{TH PRS 1SG} \\ \downarrow \\ * & (& * &) \\ \sigma & & \sigma \\ / & \diagup & | \\ \text{l}^{\text{j}}\text{ub} & \text{j} & \emptyset & \text{u} \\ \checkmark & \text{TH PRS 1SG} \\ \downarrow \end{array}$



If the L-stem is post-accenting, the derivation proceeds as in (43): the deletion of the second of the two accents not divided by metrical material (occurring before the present-tense suffix can influence the outcome) removes the accentual conflict.

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 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 stem is hypothesized to remove the problematic thematic accent and so to not 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, Revithiadou 1999,

²³While 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.

Ora Matushansky

(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 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 being 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 possible 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).

²⁴Intuitively, 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.

1 Russian verbal stress retraction as induced unstressability

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 below.

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 (§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^ja-* (*prin^ját^j* ‘to accept’, *podn^ját^j* ‘to raise’, *obn^ját^j* ‘to hug’, etc.).

Table 12: Two athematic 1SG pattern verbs

	PST-F.SG	PST-PL	PRS-1SG	PRS-3SG
a. <i>-pri.m/pri.n^ja-</i> ‘accept’	<i>pri.n^ja-l-á</i>	<i>prí.n^ja-l-i</i>	<i>pri.m-_-ú</i>	<i>prí.m-e-t</i>
b. <i>-mog_-</i> ‘be able’	<i>mog_-l-á</i>	<i>mog_-l-i</i>	<i>mog_-_-ú</i>	<i>móž_-e-t</i>

The verb *prin^ját^j* ‘to accept’ in Table 12-a exhibits accentual variability in the past, as expected from an unaccented stem, while the verb *moč^j* ‘to be able’ in Table 12-b 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 a conflicted position. 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^janút^j* ‘to remember’, *obmanút^j* ‘to

Ora Matushansky

cheat', *vzgljanút^j* 'to glance', and *minútⁱ* 'to elapse') and can therefore be regarded as semelfactive, while two are imperfective (*tonút^j* 'to drown' and *tjanút^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 a conflicted position. 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) *-gljad-* 'look'
- a. i. *gljánu/gljánet* 'will glance.1SG/3SG' (semelfactive suffix *-nu-*)
(stem)
 - ii. *proglijánet* 'will glance through.3SG, impers.' (ibid.)
 - b. *gljažú/gljadít* 'look.1SG/3SG' (suffix *-e-*)
(post-stem)
 - c. i. *vzgljanú/vzgljánet* 'will glance.1SG/3SG'
(1SG)
 - ii. *zagljanú/zagljánet* '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 §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.

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', *závíset^j* 'to depend', *videt^j* 'to see', *ne-návidet^j* 'to hate', and *obídét^j* 'to offend', the last three are 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

²⁵There is no clear difference in meaning between (49a-i) and (49c-i), except (49a-i) sounds slightly archaic.

1 Russian verbal stress retraction as induced unstressability

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. p̄órdnu/p̄órdnet ‘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 11-c: if the L-stem accent is removed, it becomes unaccented. However, only five *e*-verbs show the 1SG pattern (*deržáti* ‘to hold’, *terpéti* ‘to tolerate’, *smotréti* ‘to look’, *vertéti* ‘to turn’, and *díšáti* ‘to breathe’). The prevalence of the post-stem pattern Table 11-b 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-/Ø-* 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

²⁶The 1SG and the gerund forms of the verb *stonáti* ‘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

Ora Matushansky

- (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. *dišát^j*: ‘to breathe’:
 -*nu*-: *dixnú/dixn^jót* ‘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 examples (54) show, they may be expected not to: -*ow*- is post-accenting, and the -*nu*- derivation may involve a different, if related, root.

7.2 L-stem accentuation across verb classes

The evidence (§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 and that the accentual relation between semelfactive *nu*-verbs and their imperfective counterparts in -*i*- does not seem to be predictable: 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):

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.

1 Russian verbal stress retraction as induced unstressability

- (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), číl^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. šmignút^j (post-stem)/šmígat^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 is derived from. I leave the issue of apparently indeterminate accentuation of thematic L-stems for future research.

Abbreviations

1	first person	IPFV	imperfective
2	second person	M	masculine
3	third person	NOM	nominative
ACC	accusative	N	neuter
DIM	diminutive	NMLZ	nominalizer
F	feminine	PL	plural
GEN	genitive	PRS	present tense
GER	gerund	PST	past tense
INF	infinitive	SG	singular
INS	instrumental	TH	thematic suffix

Acknowledgments

Place your acknowledgements here and funding information here.

Ora Matushansky

References

- Alderete, John. 1999. *Morphologically governed accent in Optimality Theory*. Amherst, MA: 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](https://doi.org/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](https://doi.org/10.2307/308583).
- Boyd, Michael S. 1997. *Palatalization and coronalization in Russian and Czech: A non-linear approach*. Columbus, OH: 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](https://doi.org/10.1515/9783110919745.196).
- Butska, Luba. 2002. *Faithful stress in paradigms: Nominal inflection in Ukrainian and Russian*. New Brunswick, NJ: Rutgers University. (Doctoral dissertation).
- Choi, Sung-ho. 1996. A study on Russian inflectional accentology. *THE KRN COMMAND CAUSES ISSUES [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](https://doi.org/10.2307/412859).
- Crosswhite, Katherine. 1999. *Vowel reduction in Optimality Theory*. Los Angeles, CA: 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.
- Es'kova, Natalja A. 2008. *Нормы русского литературного языка XVIII–XIX веков: словарь. пояснительные статьи* [Norms of the Russian literary language of the XVIII–XIX centuries. Dictionary. Clarifying articles]. Moscow: Языки славянской культуры.

1 Russian verbal stress retraction as induced unstressability

- Es'kova, Natalja A. 2014. *Словарь трудностей русского языка. Ударение. Грамматические формы* [The dictionary of Russian language difficulties. Stress. Grammatical forms]. Moscow: Языки славянской культуры.
- 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](https://doi.org/10.1515/tlir.2010.015).
- Halle, Morris. 1963. О правилах русского спряжения [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](https://doi.org/10.2307/412457).
- Halle, Morris. 1975. On Russian accentuation. *The Slavic and East European Journal* 19. 104–111. DOI: [10.2307/306217](https://doi.org/10.2307/306217).
- Halle, Morris. 1997. On stress and accent in Indo-European. *Language* 73. 275–313. DOI: [10.2307/416020](https://doi.org/10.2307/416020).
- Halle, Morris. 2004. Russian phonology: The core. Manuscript, 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](https://www.jstor.org/stable/4178524).
- Idsardi, William J. 1992. *The computation of prosody*. Cambridge, MA: Massachusetts Institute of Technology. (Doctoral dissertation).
- Itkin, Ilja B. 2007. *Русская морфонология* [Russian morphonology]. Moscow: Gnozis.

Ora Matushansky

- 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](https://doi.org/10.1080/00437956.1948.11659338).
- Kayne, Richard S. 1967. Against a cyclic analysis of Russian segmental phonology. Manuscript, 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: 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, MA: 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: [10.1515/978311349121-032](https://doi.org/10.1515/978311349121-032).
- 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](https://doi.org/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](https://doi.org/10.3765/amp.v10i0.5446).
- Matushansky, Ora. 2023b. *Russian verbal stress clash as a tonal conflict*. Manuscript, 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. In Tatiana Bondarenko, Peter Grishin & Anton Kukhto (eds.), *Formal Approaches to Slavic Linguistics 30: The third MIT meeting*.
- Meillet, Antoine. 1934. *Le slave commun*. Paris: Champion.

1 Russian verbal stress retraction as induced unstressability

- Melvold, Janis L. 1989. *Structure and stress in the phonology of Russian*. Cambridge, MA: 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 Źygis. 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](https://doi.org/10.5334/gjgl.59).
- Red'kin, Valerij A. 1965. Об акцентных соотношениях имени и глагола в современном русском литературном языке [About accentual relationships of nomina and verbs in the Modern Russian language]. *Вопросы языкоznания [Questions of linguistics]* 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 iota-tion. *Language* 77(1). 26–60. <https://www.jstor.org/stable/3087027>.
- Sims, Andrea D. 2006. *Inflectional defectiveness in a paradigmatic theory*. Columbus, OH: The Ohio State University. (Doctoral dissertation).
- Slioussar, Natalia. 2012. Некоторые сведения о формообразовательных классах русских глаголов [Some data on the inflectional classes of Russian verbs]. Manuscript, 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 Janda. 1996. *Common and comparative Slavic: Phonology and inflection*. Columbus, OH: Slavica.
- Zaliznjak, Andrej A. 1977. *Грамматический словарь русского языка* [Grammatical dictionary of Russian language]. Moscow: Izdatel'stvo Russkij Jazyk.
- Zaliznjak, Andrej A. 1985. *От праславянской акцентуации к русской* [From Proto-Slavic accentuation to Russian one]. Moscow: Nauka.
- Zaliznjak, Andrej A. 2019. К истории одного нефонетического изменения [On the history of one non-phonetic change]. In Anatolij F. Zhuravlev (ed.), *Славянское и балканское языкознание: Русистика. Славистика. Компаративистика. Сборник к 64-летию С. Л. Николаева* [Slavic and Balkan Linguistics: A collection in honor of S. L. Nikolaev's 64th birthday]. Moscow: Nauka.

Ora Matushansky

tics: Rusistics, slavistics, comparative studies. Festschrift for the 64th birthday of S.L. Nikolaev], 164–204. Moscow: Institute of Slavic Studies.

Chapter 2

Perfectivity in Russian, Czech and Colloquial Upper Sorbian

Olav Mueller-Reichau^a

^aLeipzig 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 and imperfective forms with each other. 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 incompleteness is that it is simply false, as there are well-known occurrences of imperfectives under reference to completed events, as demonstrated in detail for Russian in Grønn (2004). Such cases thus have 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,

Olav Mueller-Reichau

as we know at the latest since the foundational work of Dickey (2000). Other authors addressing inner-Slavic differences in aspect 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 behaviors 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 want 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 Muzhakov; own translation] (Šćerba 1973: 121).¹ The present paper argues for a nuanced view of perfectivity in Slavic.

The paper is structured as follows. In §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 §3 I will then introduce some theoretical background assumptions that I take for granted in the line of argumentation to follow. §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, §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 that it has to be made more precise to fully capture the data. To this end, I will offer a proposal in §6, showing that this correctly models the distribution of aspectual forms. Having laid out my proposal so far in prose, I will pin it down formally in §7. §8 serves to place my analysis into the broader context of scalarity-based approaches to Slavic aspect. Finally,

¹Quoted after Scholze (2008: 230) and Werner (2018: 144).

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

§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. §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 from each other, 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 masculine personal gender, the obligatory presence of articles in noun phrases, a particularly high amount of lexical borrowings from German, a passive construction with the German loan verb *wordować* ‘become’; the use of adverbial preverbs in the function of verbal prefixes like in *hrómadźestajeć* ‘put together’ (together+put) or *nutřčinc* ‘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

Olav Mueller-Reichau

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 předal*.
 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 form⁴ 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 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:

²This 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.

³Not surprisingly, Standard Upper Sorbian aspectual morphology works in the same way, see [Scholze \(2008: 230\)](#).

⁴Throughout 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 irritation, but I think it would be even more irritating if I chose a different label for those forms that have perfective morphology from the point of view of other Slavic languages.

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

- (5) *Jurij jo rune jen text šeložil, hdyž sym ja nutř šišol.*
 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 the respective 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 przełožić.*
 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). And, secondly, there is no ban on perfective infinitives after phase verbs, consider (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 započal jowo začíšće napisáć.*
 he AUX start.PST.PFV his impressions write.INF.PFV
 'He began to write down his impressions.'

Olav Mueller-Reichau

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 the following 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 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 means, roughly saying, 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, 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 always⁵ denote a state-changing event, at the same time entailing the existence of an eventuality subsequent to the 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 *pjoti* ‘start going’, for instance. In any

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.

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

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 verbalized 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

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:

⁶A 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).

Olav Mueller-Reichau

- (11) *Hdyž hromadzé za blidom sedžeštaj, 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 wěšci fönwe něke.*
 well he REFL surely blow-dry.PRS.IPFV now
 ‘He surely is blow-drying his hair right now.’
- (14) *Na tón so wěšci 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.’

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

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 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.

Olav Mueller-Reichau

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 occurring 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 how many events precisely. 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 aeterminative (it is generic). And 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):

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

- (21) *Dyš sem ja do lodna šišla, jo sej wona rune jabuka brala.*
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 šišla, 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 and argue that interpreting the apples as a single unit not only leads to terminativity, but in addition 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.’

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

⁸Note that both main predicates are translated into English by means of the progressive ('was taking').

Olav Mueller-Reichau

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 that keeps 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:

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

- 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 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 like events.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.

⁹A 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.

Olav Mueller-Reichau

- (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, the event should return to where it was before.¹⁰

With respect to (29), two things need to be noted. First, the movement along a path associated with the learning of a poem is no motion in the physical sense, but rather a metaphorical motion along the words of the poem as they are considered by the learner. And, 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

¹⁰I 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 \(2001\)](#).

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

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 non-use of the perfective 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 usually goes under the terminology “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 behavior of the individuals allows for *inducing* the generalization. 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 generalization 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 chětř 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

Olav Mueller-Reichau

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 like (33) a dispositional generic. 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*

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

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 the time is consumed to proceed 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.

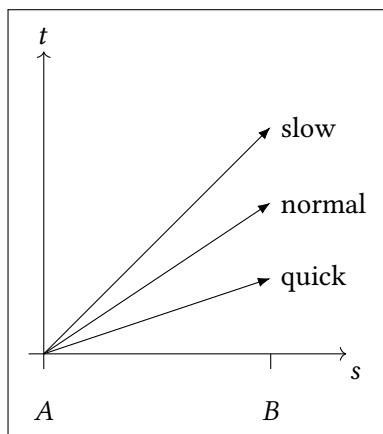


Figure 1: speeds of learning a poem

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):
- that he learns poems very slowly
 - that he learns poems slowly
 - that he learns poems at normal speed

Olav Mueller-Reichau

- 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 normalized distance \bar{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

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

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:**
 $\text{IPFV} \Rightarrow \lambda P \lambda t \exists e. P(e) \wedge t \circ \tau(e)$

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:**
 $\text{PFV}_{CUS} \Rightarrow \lambda P \lambda t \exists e. P(e) \wedge \text{DET}(P) \wedge t \circ \tau(e)$

In contrast to that, the perfective operator of Czech may formally be pinned down as in (37).

- (37) **Perfective operator in Czech:**
 $\text{PFV}_{CZ} \Rightarrow \lambda P \lambda t \exists e. P(e) \wedge \text{DET}(P) \wedge t \circ \tau(e) \wedge f_{end}(\tau(e)) \subseteq t$

As for the Russian perfective operator, I propose the following:

- (38) **Perfective operator in Russian:**
 $\text{PFV}_{RU} \Rightarrow \lambda P \lambda t \exists e. P(e) \wedge \text{DET}(P) \wedge t \circ \tau(e) \wedge f_{end}(\tau(e)) \subseteq t \wedge f_{end}(t) \subseteq f_{target}(e)$

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:**
 $\forall P. \text{UNI}(P) \leftrightarrow \forall e \forall e' \forall q. P(e) \wedge e' \leq e \wedge q = \text{TRACE}(e') \rightarrow \exists p. p = \text{TRACE}(e) \wedge q \leq p$

Olav Mueller-Reichau

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 one is yes. As for the second question, I will insert this small section to answer it carefully.

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, in effect, 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

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

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 as well as 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),

¹¹See Mueller-Reichau (2023) for discussion of attempts to reduce the two covert operators to one.

Olav Mueller-Reichau

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 tire change.PRS.IPFV
 ‘I am learning how to change a tire.’

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 substitution of imperfective *wukne* by 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 tire follows a standardized 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).

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

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* ‘tire 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 nawukłा 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 *nawukłा* '(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 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?

¹³The 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/>

Olav Mueller-Reichau

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) by its imperfective counterpart *wukla* (Scholze 2008: 245). Werner (2013) presents the following example:¹⁵

- (47) *Wón je list čital a je šet 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 *nawukla* 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 will be done, ‘she’ will 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 (what she will probably have to demonstrate at school the next day). In (46), in other words, the predicate *learn the poem* is understood in abstraction away 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). This path is directed, so my argument, and 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.

¹⁵Note that I treat the motion verb *šet* 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.

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

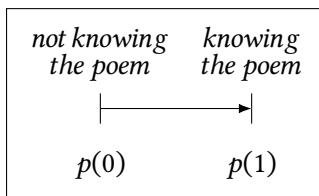


Figure 2: path of learning the poem in (46)

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, or 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,

Olav Mueller-Reichau

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 *chěr* ‘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 contacting 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 neighboring 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.

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

Abbreviations

3	third person	IPFV	imperfective
AUX	auxiliary	PFV	perfective
DU	dual	PST	past
F	feminine	PRS	present
INF	infinitive	REFL	reflexive

Acknowledgements

For helpful input and discussion, special thanks to Tilman Berger, Petr Biskup, Simon Blum, Berit Gehrke, Matthias Irmer, Fabian Kaulfürst, Alena Paulík, Matthias Schlegel, Lenka Scholze, Till Voigt, Eduard Werner, and Karolina Zuchewicz. Thanks also to the audiences of the conferences *70 let sorabistiki – konferenca w Lipsku* at Leipzig University (May 19–22, 2022) and *Formal description of Slavic languages 15* at Humboldt-University Berlin (October 5–7, 2022).

References

- Alvestad, Silje Susanne. 2013. *Beware of Fakes! Fake imperfectives in the Slavic imperative*. Oslo: University of Oslo.
- Arregui, Ana, María 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](https://doi.org/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.

Olav Mueller-Reichau

- 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](https://doi.org/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](https://doi.org/10.1524/stuf.2012.0016).
- Carlson, Greg. 1995. Truth conditions of generic sentences: Two contrasting views. In Gregory Carlson & Francis Pelletier (eds.), *The generic book*, 224–237. Chicago, London: 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](https://doi.org/10.1093/acrefore/9780199384655.013.326).
- Comrie, Bernhard. 1976. *Aspect*. Cambridge: Cambridge University Press.
- Dickey, Stephen. 2000. *Parameters of Slavic aspect*. Stanford: CSLI.
- Dickey, Stephen. 2006. Aspectual pairs, goal orientation and *po*-delimitatives in Russian. *Glossos* 7. 1–37.
- Dickey, Stephen. 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, CA: 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](https://doi.org/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](https://doi.org/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.

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

- Heck, Stefan. 2018. Verbal aspect in the Czech and Russian imperative. In Erik Fuß, Marek Konopka, Beata Trawiński & Ulrich Waßner (eds.), *Grammar and corpora 2016*, 249–258. Heidelberg: Heidelberg University Publishing. DOI: [10.17885/heiu.361.c4705](https://doi.org/10.17885/heiu.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. *Příruční 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 habituels 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 Carlson, Alice ter Meulen, Gennaro Chierchia & Godehard Link. 1995. Genericity: An introduction. In Gregory Carlson & Francis Pelletier (eds.), *The generic book*, 1–124. Chicago, London: 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/>.
- Lewaskiewicz, Tadeusz. 2002. Oborsorbisch. 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*

Olav Mueller-Reichau

- 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 genetics. 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. 2023. The morphosemantics of Russian aspect. In Berit Gehrke & Radek Šimík (eds.), *Topics in the semantics of Slavic languages*, xx–xx. 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, NJ: John Wiley. DOI: <https://doi.org/10.1002/9781118788516.sem134>.
- Petruchina, Elena V. 2000. *Aspektual'nye kategorii glagola v russkom jazyke v so-potavlenii 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.

2 Perfectivity in Russian, Czech and Colloquial Upper Sorbian

- 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, Natalja Ju., 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](https://doi.org/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](https://doi.org/10.1007/s10988-005-2466-y).

Chapter 3

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.

¹The 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 to appear](#)). 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 §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)

3 Polar questions in Czech and Russian: An exploratory corpus investigation

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 §2, we describe the method of annotation. §3 reports on the absolute values of the annotated features and the results of the inter-annotator agreement. In §4, we discuss the results. §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

²The complete annotation is available here: <https://bit.ly/3xKM9XX>

Onoeva & Staňková

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 (\approx semantic) and outer (\approx pragmatic) – and that they differ in their syntactic and semantic/pragmatic features. Inner negation is interpreted and licenses NPIs (NCIs in Czech and Russian), whereas outer negation does not trigger the negative operator per se and licenses 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 tedka 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 possible) 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?*, $p = \text{it is raining}$.

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

3 Polar questions in Czech and Russian: An exploratory corpus investigation

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.
- d. A u tebja pomimo sobački est' eščě 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.

³Sudo (2013) suggests that epistemic/speaker bias also can include deontic or bouleptic states. Since these are too complex to judge based on written text, we only work with speaker's beliefs.

Onoeva & Staňková

We used affirmative prejacent and not question radicals to decide whether or not speakers had any prior belief. The same applies for evidential bias. With the aid of prejacent, 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 §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 mladší?
 how RAZVE Aleksandr Aleksandrovič youngest
 ‘Wait a second, is Aleksandr Aleksandrovič the youngest?’ (Ru)
 $\phi = \text{Aleksandr Aleksandrovič}$ 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’.

3 Polar questions in Czech and Russian: An exploratory corpus investigation

- (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 osmý nějak?
you leave around eight somehow
'Are you leaving around eight?' (Cz)
 ϕ = 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)
 ϕ = 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 Ryžík sitting in the speaker's spot.

Sp: Ryžík ty čto moë mesto zanjal?
Ryžík you what my spot taken
'Ryžík, have you taken my spot?' (Ru)
 ϕ = Ryžík 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.

Onoeva & Staňková

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.

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

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 excep-

3 Polar questions in Czech and Russian: An exploratory corpus investigation

tional cases where the polarity of speaker/evidential bias was not in accord with the question's polarity.

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)

In our sample, initial verb position (V INI) 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 INI in non-null EB 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 NON-INI).

⁴We thank an anonymous reviewer for pointing out the inconsistencies in our reporting of the results.

Table 4: V-position–evidence correlation ($p < .001$)

	V NON-INI	V INI
EB ±1	103 (87.4)	27 (42.6)
EB 0	211 (226.6)	126 (110.4)

As for indefinites, they were present in 59 Czech PQs. 54 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*. 38 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 prejacent 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

3 Polar questions in Czech and Russian: An exploratory corpus investigation

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.

⁵INTERROGATIVITY PRINCIPLE: Ask a question $?p$ only if the context set c does not entail a complete answer to $?p$.

Onoeva & Staňková

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 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 cautiously analyse the results of the semantic/pragmatic features for Russian, since the inter-annotator agreement was poor for both annotators. The potential explanation for this is the nature of the Russian corpus. Compared 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 least some agreement between the annotators.⁶ Such PQs usually

⁶We thank an anonymous reviewer who recommended to check it.

3 Polar questions in Czech and Russian: An exploratory corpus investigation

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 Dobrovolskij & 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 idět?* '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 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).

Onoeva & Staňková

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

The study was funded by the Czech Science Foundation (GAČR), project No. 21-31488J. We would like to acknowledge the valuable input (alphabetically) of Katerína Hrdinková, Roland Meyer, Mariia Razguliaeva, and Radek Šimík. We would like to express our gratitude to the organizers and audience of FDSL 15 for the discussion and insightful comments. We also thank our annotators who

3 Polar questions in Czech and Russian: An exploratory corpus investigation

helped us during the investigation, the reviewers and the editors of the volume for their valuable feedback.

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](https://doi.org/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](https://doi.org/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/10.1177/001316446002000102>.
- Daneš, František, Zdeněk Hlavsa & Miroslav Grepl. 1987. *Mluvnice češtiny*, vol. 3. Prague: Academia.
- Dobrovol'skij, Dmitrij & Irina Levontina. 2014. Diskursivnye slova v obščevo-prositol'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. Zažarov, 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. <https://www.dialog-21.ru/en/digest/2014/>.
- 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-poql-li-not-li.pdf>.

Onoeva & Staňková

- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.16995/glossa.6487).
- Goodhue, Daniel. 2018. *On asking and answering biased polar questions*. Montreal, QC: 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 korpusse russkogo jazyka. *Nacionalnyj 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. *Nacionalnyj korpus russkogo jazyka 2006–2008. Novyje rezul'taty 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](https://doi.org/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. Kálman & A. Kornai on the occasion of their 60th birthdays*, 247–280. Budapest: MTA Nyelvtudományi Intézet. DOI: [10.18135/kk120.2019](https://doi.org/10.18135/kk120.2019).

3 Polar questions in Czech and Russian: An exploratory corpus investigation

- 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://www.korpus.cz/>.
- 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 Mleinck. 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](https://doi.org/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](https://doi.org/10.1163/9789004183988_008).

Onoeva & Staňková

- 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](https://doi.org/10.1023/B:LING.0000033850.15705.94).
- Rudin, Deniz. 2022. Intonational commitments. *Journal of Semantics* 39(2). 339–383. DOI: [10.1093/jos/ffac002](https://doi.org/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](https://doi.org/10.1515/9783110912111.385).
- Šimík, Radek. to appear. 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](https://doi.org/10.1163/9789004183988_009).

Chapter 4

Dual preservation in Slovenian: The verb supports the noun in semi-spontaneous production

Matic Pavlič^a, Artur Stepanov^b & Penka Stateva^b

^aFaculty of Education – University of Ljubljana ^bCenter 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 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 (Tesanière 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 the 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

Pavlič, Stepanov & Stateva

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 dual in verbs is more preserved also preserve more dual morphology in (subject) nouns than the dialects in which 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.” The markedness of

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).

4 Dual preservation in Slovenian

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) are 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 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).

- (1) Ta dva stara tovornjaka
 these.DU two.DU old.DU lorries.DU
 'these two old lorries'

²“>” stands for “is followed by” on the diachronic trajectory.

Pavlič, Stepanov & Stateva

- (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 spray 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

4 Dual preservation in Slovenian

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 it is ...?").

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

³The number given by N_{loc} represents the number of locations for which a response to a SLA

Pavlič, Stepanov & Stateva

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).

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

question was available (out of 413 locations).

⁴It is not clear, though, why contacts with German on the North does not lead to the dual loss.

4 *Dual preservation in Slovenian*

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 mas-

Pavlič, Stepanov & Stateva

ciline (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 preservation of dual.

DIALECTS	Elicitation SLA	Verb <i>delati</i>	Noun <i>brat</i>	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

4 Dual preservation in Slovenian

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.

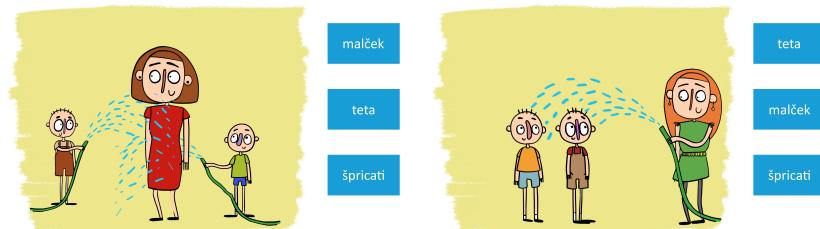


Figure 1: A target set of graphical stimuli.

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 ($\frac{1}{2}$ masculine and $\frac{1}{2}$ feminine), number ($\frac{1}{2}$ singular and $\frac{1}{2}$ plural), and sentence function ($\frac{1}{2}$ subject and $\frac{1}{2}$ 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).

Pavlič, Stepanov & Stateva

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

4 Dual preservation in Slovenian

(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.

	more preserved dual on verb			less preserved dual on verb	
Dialect	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

Pavlič, Stepanov & Stateva

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.

Finally, we modelled the results. We put the target response (the dual of a noun referring to two entities) as a reference value in the model 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

⁵The 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.

4 Dual preservation in Slovenian

Table 5: Analyzed target nouns by sentence function and dialect group.

		NOUN		
		subject	object	
VERB	+dual	p<0.025		
	-dual	p<0.0001	15/353 (04%) 50/260 (19%)	45/386 (12%) 42/248 (17%)
				p>0.05

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 less 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.

Table 6: Statistical model.

FACTORS	OR	CI	p
(intercept)	0.01	0.00–0.07	<0.001
Function [object]	6.31	0.69–57.33	0.102
Group [+dual on verb]	0.11	0.01–0.99	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	

The statistically significant odds ratio (OR) for a non-target response in the

Pavlič, Stepanov & Stateva

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 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.

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 ($\beta = -4.15$, 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 ($\beta = 1.84$, 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).

4 Dual preservation in Slovenian

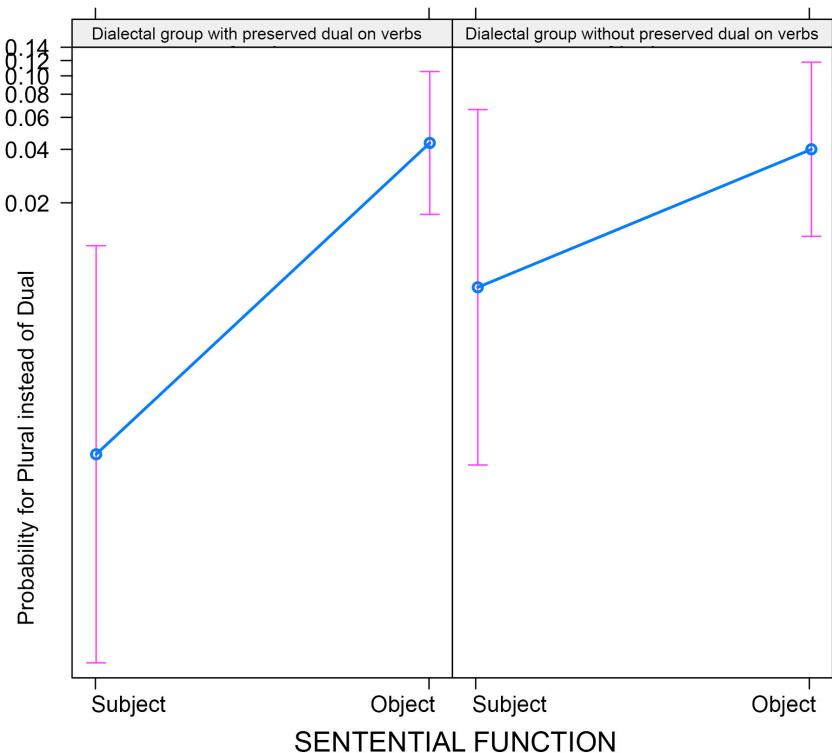


Figure 2: Comparison of the probability of false plural when syntactic position is crossed with dialect type.

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.

Pavlič, Stepanov & Stateva

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.

⁶This 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.

4 *Dual preservation in Slovenian*

Acknowledgments

The research was financially supported by the ARIS research project “Linguistic transfer in the pragmatic domain: Slovenian speakers in a multilingual environment” no. J6-2580 (PI: Penka Stateva).

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](https://doi.org/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](https://doi.org/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.

Pavlič, Stepanov & Stateva

- 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](https://doi.org/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.
- 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](https://doi.org/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](https://doi.org/10.1515/9783110209303.4.581).
- Slobodchikoff, Tatyana G. 2019. *The evolution of the Slavic dual: A biolinguistic perspective*. Lanham, MD: 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](https://doi.org/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](https://doi.org/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, Jérémie & Florian Schwarz. 2018. *PennController for internet based experiments (IBEX)*. <https://www.pcibex.net/>.

Chapter 5

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

Margje Post

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 §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. §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

¹Moderate 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.

²The 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 vowels, all vowels except /u/ are produced

5 Word prosodic structure and vowel reduction in Moscow and Perm Russian

into two different allophones, depending on their position in the word. In CSR, the word /molo'ko/ 'milk' is pronounced as [mələ'ko], 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 §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

as front vowels.

³Most 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 [ə] 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).

Margje Post

the phonological status of two-degree reduction (cf. Barnes 2006, Bethin 2006, Molčanow 2015, *inter alia*).

Much less is known about unstressed vowels and the word prosodic structure in other, non-central varieties of Russian (§1.2) and in the other East Slavic languages (§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₋₂CV₋₁'CV₀C with the vowels /a, o/ in a large number of Russian varieties. He discerned 12 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'saripova 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

5 Word prosodic structure and vowel reduction in Moscow and Perm Russian

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 some conservative central dialects of Russian and Belarusian ([Broch 1916, Bethin 2006; *inter alia*](#)).⁵

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.

⁵Note that [Borise \(2017\)](#) could not confirm a local high peak before a fall on the stressed syllable

Margje Post

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 Lukaszewicz 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 (Lukaszewicz & Mołczanow 2018). 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.

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).

5 Word prosodic structure and vowel reduction in Moscow and Perm Russian

The variation of vowel reduction patterns in the East Slavic languages and their dialects show 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

Margje Post

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 §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'CV₀C 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 á₀. 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. §4.2 and §4.3 below).⁷

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

⁶Russian text, including Russian surnames, is transliterated using Comrie & Corbett's (1993) transliteration system.

⁷As 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. §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.

5 Word prosodic structure and vowel reduction in Moscow and Perm Russian

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.

Margje Post

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

5 Word prosodic structure and vowel reduction in Moscow and Perm Russian

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 (\acute{a}_0) being significantly longer than the vowels in the first posttonic syllable (a_{-1}), which in turn are longer than the vowels in the second posttonic 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 2024). This shows that the earlier mentioned gender effect is due to the difference in Perm.

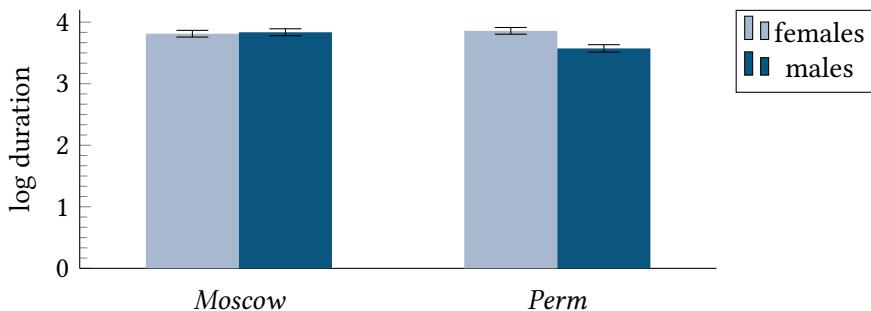


Figure 1: GENDER vs. LOCATION: Mean duration (log-transformed) of vowels (all three positions together), for female speakers (light-coloured) and Perm speakers (dark-coloured), with Moscow speech on the left and Perm speech on the right

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

⁸Bar 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.

Margje Post

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.

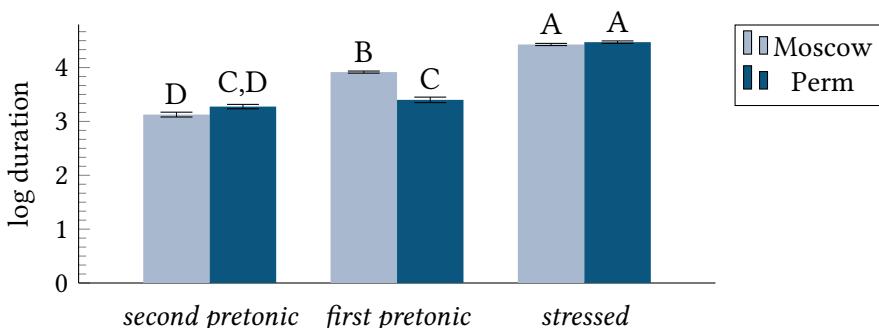


Figure 2: Mean duration (log-transformed) according to LOCATION and DISTANCE-TO-STRESS (light-coloured bars = Moscow; dark-coloured bars = Perm; both genders). Error bars represent standard errors, with Compact Letter Display of pairwise comparisons of the mean durations from the post-hoc pairwise Tukey HSD test. Bars not connected by the same letter are significantly different.

In Moscow, the actual mean durations (not log-transformed) of the three consecutive vowels ($a_{-2} : a_{-1} : a_0$) 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

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 ($F [2, 431] =$

⁹For details on mean durations and an elaborate discussion of the durational measurements from a sociolinguistic perspective, see Post (2024).

5 Word prosodic structure and vowel reduction in Moscow and Perm Russian

1.4552, $p > .05$), for the boys used the same relative durations as the girls in both cities (see Post 2024).

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 \dot{a}_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 F1 significantly according to each position, but the Tukey HSD test showed that the female speakers produce relatively higher F1 values than the men in \dot{a}_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 (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 \dot{a}_0 receiving the letters E, B, and C, respectively. Unlike for duration, a_{-1} has not lower, but higher mean F1 values than \dot{a}_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).

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

Margje Post

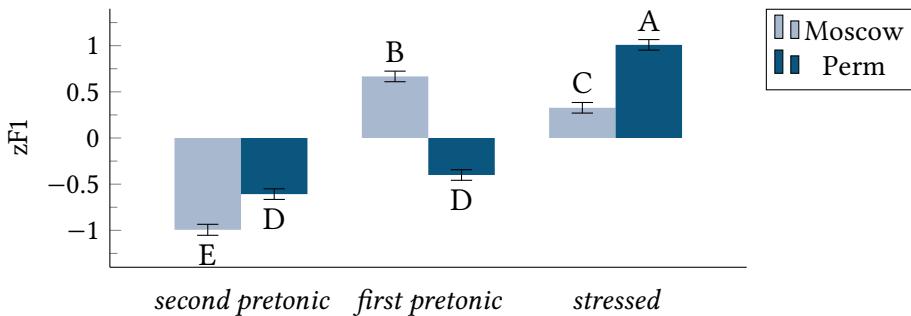


Figure 3: Mean zF1 values according to LOCATION and DISTANCE-TO-STRESS (both genders, light-coloured bars = Moscow; dark-coloured bars = Perm), with letter report from the pairwise comparison Tukey test

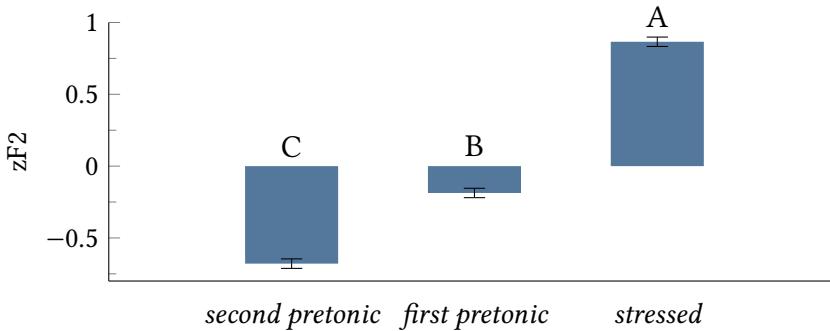


Figure 4: Mean zF2 values according to DISTANCE-TO-STRESS (both genders and both locations), including letter report from the pairwise comparison Tukey test

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 large space, especially in F2, than the stressed vowels (circle markers). The female (top) and male speakers (bottom) show the same patterns.

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

5 Word prosodic structure and vowel reduction in Moscow and Perm Russian

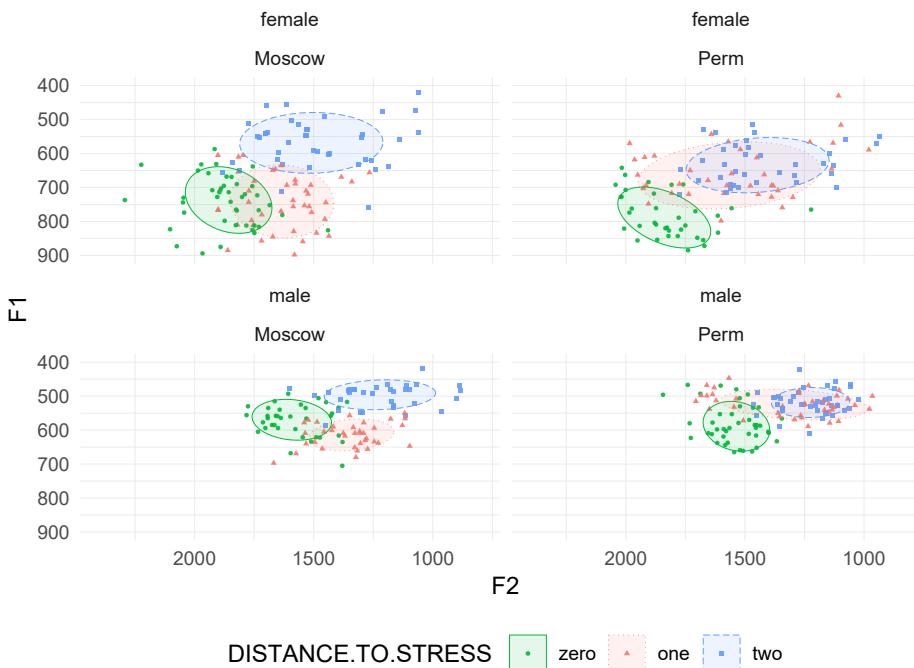


Figure 5: Vowel plots of F1 (y-axis) and F2 (x-axis) of the second pretonic (squares; dashed line), first pretonic (triangles; dotted line) and stressed (circles; solid line) vowels in the three target words produced by female speakers and male speakers in Moscow and Perm, with ellipses representing one standard deviation from the mean

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 2024). 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 2024).

Margje Post

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 2024) and proved not statistically significant. Unlike our data, Erofeeva's data were based on spontaneous speech from 4 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, as 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.

5 Word prosodic structure and vowel reduction in Moscow and Perm Russian

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 \dot{a}_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 \dot{a}_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; 3 of the 4 speakers in Barnes 2006). The even higher values in prestressed than in stressed position can be caused by the lowering effect of the final palatalized consonant following \dot{a}_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₋₂t(akat')* vs. (*to*)*po₋₁t(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.

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

¹⁰The 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).

Margje Post

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 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. In addition, 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) observations 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 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 pre-

¹¹The 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.

5 Word prosodic structure and vowel reduction in Moscow and Perm Russian

tonic 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 not only 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 \acute{a}_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

Margje Post

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

A range of people have helped me during this dive into phonetics, a field I knew too little about. First of all, I want to thank Benedikte Fjellanger Vardøy for her fieldwork in Perm and for our exciting first exploration of the data, Svetlana Djachenko for doing most segmentations, and Bistra Andreeva for vital statistical analyses and for advice and support. Alexander Krasovitsky, Sergej Knjazev,

5 Word prosodic structure and vowel reduction in Moscow and Perm Russian

Elena Erofeeva, Brechtje Post, Mitko Sabev and Carl Börstell answered questions on phonetics and statistics, and Elaine Schmidt and Dirk Jan Vet wrote the Praat scripts. I am also indebted to the members of the Phonetics group at Saarland University for their support during my research stay in 2022, and to Ivan Yuen and to the two anonymous reviewers, whose constructive and insightful comments led to a substantial improvement of this paper. I cannot possibly have incorporated all advice, so all remaining shortcomings are my sole responsibility. Last but not least, I thank the speakers and the secondary schools in Moscow and Perm they attended for their participation.

This study was made possible by various grants from the Meltzer Foundation and from the University of Bergen. The data collection was approved by the Norwegian Social Science Data Services (NSD).

References

- Al'muxamedova, Zel'fa M. & Ravza È. Kul'saripova. 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 Akadēmii 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](https://doi.org/10.1080/00085006.1995.11092083).
- Avanesov, Ruben I. 1984. *Russkoe literaturnoe proiznošenie*. 6th ed. Moscow: Prosveščenie.
- Avanesov, Ruben I. & Sofja 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](https://doi.org/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 sovremennoj 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.

Margje Post

- 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. Orfoèipeskie pravila akan'ja v sovremenном белорусском языке в сопоставлении с диалектами. In Ljudmila È. Kalnyn' (ed.), *Issledovaniya 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 veroyatnostnye struktury: Sociolinguisticheskij aspekt (na materiale foneticheskogo urovnya)*. 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-Universität 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 sovremennoj russkogo literaturnogo jazyka. In Aleksandr 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/S002226712000102](https://doi.org/10.1017/S002226712000102).
- 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.

5 Word prosodic structure and vowel reduction in Moscow and Perm Russian

- 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 sviaznoj reči: Èksperimental'noe issledovanie na materiale russkogo jazyka*. Saint Petersburg: Saint Petersburg University Press.
- Labov, William. 2001. *Principles of linguistic change*, vol. 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](https://doi.org/10.1121/1.1912396).
- Łukaszewicz, Beata & Janina Mołczanow. 2018. The role of vowel parameters in defining lexical and subsidiary stress in Ukrainian. *Poznań Studies in Contemporary Linguistics* 54(3). 355–375. DOI: [10.1515/pscl-2018-0014](https://doi.org/10.1515/pscl-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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.15845/bells.v7i0.1138).
- Post, Margje. 2024. 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*

Margje Post

- papers from Methods in Dialectology XVII*, 253–280. Berlin: Language Science Press.
- 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](https://doi.org/10.2478/9788366675728-012).
- Potebnja, Aleksandr A. 1866. *Dva issledovaniya o zvukax russkogo jazyka: I. O polnoglasii. II. O zvukovyx osobennostjakh russkix narečij*. Voronež: Tipografija V. Gol'dstejna.
- 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](https://doi.org/10.1017/jlg.2021.5).
- Verbickaja, Ljudmila A. 1977. Variantnost' sovremennoj proiznositel'noj 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. Regional'nye osobennosti realizacii russkoj reči (na fonetičeskem urovne). *Vestnik Leningradskogo universiteta* 8(2). 71–80.
- Vysotskij, Sergej S. 1973. O zvukovoj strukture slova v russkix govorax. In Sofja 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](https://doi.org/10.1121/10.0013410).
- Zlatoustova, Ljubov' V. 1981. *Fonetičeskie edinicy russkoj reči*. Moscow: Moscow State University Press.

Chapter 6

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 $\succ \phi$ -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

Zorica Puškar-Gallien

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 §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) \succ dependent (ACC, GEN) \succ oblique (DAT) \succ 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

6 Morphosemantic mismatches with pronouns

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. §2 introduces the pronominal paradigms and morphosemantic mismatches. A short overview of previous literature and certain issues raised from it will be presented in §3. The proposal on the internal structure of pronominal elements will occupy §4. Subsequently, §5 will inspect the consequences of the proposal for syntax and interpretation in more detail. §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)-*.

Zorica Puškar-Gallien

Table 1: Strong pronouns vs. clitics in BCMS

	1SG	2SG	1PL	2PL	3SGM/N	3SGF	3PL
NOM	<i>ja</i>	<i>ti</i>	<i>mi</i>	<i>vi</i>	<i>on-∅/-o</i>	<i>on-a</i>	<i>on-i/-e/-a</i>
GEN	<i>m-en-e</i>	<i>t-eb-e</i>	<i>na-s</i>	<i>va-s</i>	<i>nje-ga</i>	<i>nj-e</i>	<i>nj-ih</i>
DAT	<i>m-en-i</i>	<i>t-eb-i</i>	<i>na-ma</i>	<i>va-ma</i>	<i>nje-mu</i>	<i>nj-oj</i>	<i>nj-ima</i>
ACC	<i>m-en-e</i>	<i>t-eb-e</i>	<i>na-s</i>	<i>va-s</i>	<i>nje-ga</i>	<i>nj-u</i>	<i>nj-ih</i>
INS	<i>m-n-om</i>	<i>t-ob-om</i>	<i>na-ma</i>	<i>va-ma</i>	<i>nj-im</i>	<i>nj-om</i>	<i>nj-ima</i>
LOC	<i>m-en-i</i>	<i>t-eb-i</i>	<i>na-ma</i>	<i>va-ma</i>	<i>nje-mu</i>	<i>nj-oj</i>	<i>nj-ima</i>

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

(2) *Clitics vs. pronouns in a PP*

¹See Stegovec (2019) for a tripartite distinction between Slovenian strong, clitic and P-pronouns, present in earlier stages of BCMS.

6 Morphosemantic mismatches with pronouns

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.SG.M.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*

Zorica Puškar-Gallien

- 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 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.SG.M.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

6 Morphosemantic mismatches with 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 §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.))

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.

²The discussion here is restricted to third-person clitics.

³The 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’.

Zorica Puškar-Gallien

- (8) *Sloppy readings of strong pronouns as complements of P*
- 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.'
 - 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.'
 - 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*
- 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).'
 - 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).'

6 Morphosemantic mismatches with pronouns

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 identity 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*

Zorica Puškar-Gallien

- 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 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 her.' [+HUM] [-HUM] (Despić 2011: 246)
- (13) *Focused inanimate pronoun*
Malo ko obilazi muzeje oko gradske crkve_i. Nju_i *(samu),
few who visits museums around city chuch 3.F.SG.ACC 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)

⁴See Browne (1974), Zec & Inkelaš (1991), Franks & Progovac (1994), Godjevac (2000) on clitics requiring prosodic prominence, Godjevac (2000) on focus requiring prosodic prominence, and Despić (2011: 244) on further interactions between the two.

6 Morphosemantic mismatches with pronouns

The animacy properties, the ability to be bound and the sloppy readings outlined in §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 disguise.

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 question 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 others, 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-Branched Extraction of a nominal modifier is possible with a lexical 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 reject the NP/DP distinction as a culprit for the difference in the behaviour of nominal 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),

Zorica Puškar-Gallien

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, it 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

6 Morphosemantic mismatches with pronouns

(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). And 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 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

⁵Cardinaletti 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).

Zorica Puškar-Gallien

Y-model of syntax. They have been tied to person by Sichel & Toosarvandani (to appear, 2024), 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 [ϕ P [NP]]]). The ϕ P 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.

The base of the noun consists of a nominal root and a nominalizing head “ (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 “P as consisting solely of the categorizing head “ (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

⁶Puškar-Gallien (to appear) 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.

6 Morphosemantic mismatches with pronouns

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.

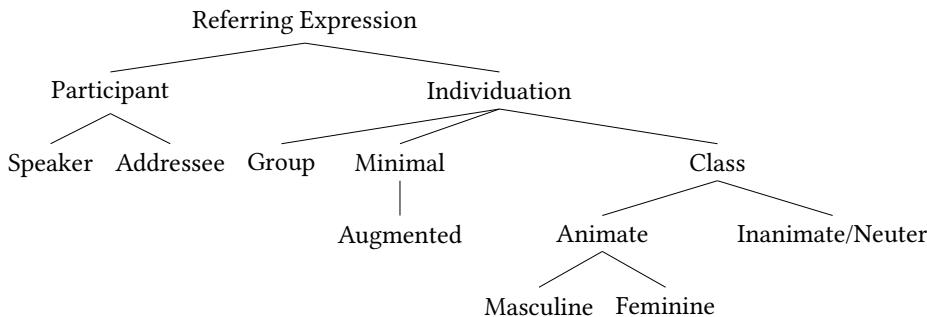


Figure 1: Structural hierarchy of ϕ -features (Harley & Ritter 2002: 486)

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 ‘‘P base. This models the idea that ‘‘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.

⁷Precursors for this idea include Caha (2021), who models RefP as an additional syntactic projection above the ‘‘P, albeit without providing much detail on its purpose or interpretation. Sichel & Toosarvandani (2024) 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.

Zorica Puškar-Gallien

Sichel & Toosarvandani to appear), 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 (to appear, 2024) 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 ‘’ (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 categories on different categorising heads (and sharing their extended projections). Making ‘’ 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.

⁸Puškar-Gallien (to appear) offers a revision of this model and provides further detail on how animacy and humanness can be encoded on the ‘’ base.

6 Morphosemantic mismatches with pronouns

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 '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 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

⁹GENDER 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.

Zorica Puškar-Gallien

the syntactic heads that project the corresponding phrases. These features can also include a small hierarchy of sub-features below them.¹⁰

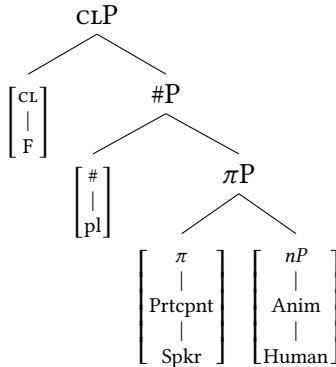
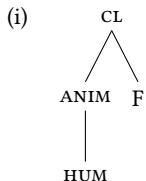


Figure 2: Basic pronominal functional spine in BCMS

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*

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.



6 Morphosemantic mismatches with pronouns

Hierarchy:

NOMINATIVE \succ ACCUSATIVE \succ GENITIVE \succ DATIVE \succ INSTRUMENTAL \succ COMMITATIVE. 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).

$$(17) \quad [\text{PP} \ P \ [K_{\text{OBL}} \ P \ K_{\text{OBL}} \ [K_{\text{DEP}} \ P \ K_{\text{DEP}} \ [K_{\text{UNM}} \ P \ K_{\text{UNM}} \ [\text{CLP} \ \text{CL} \ [\#P \ \# \ [\pi P \ \pi \ [·P \ “”]]]]]]]$$

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

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).

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

¹¹Modelling 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 subfeatures 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.

Zorica Puškar-Gallien

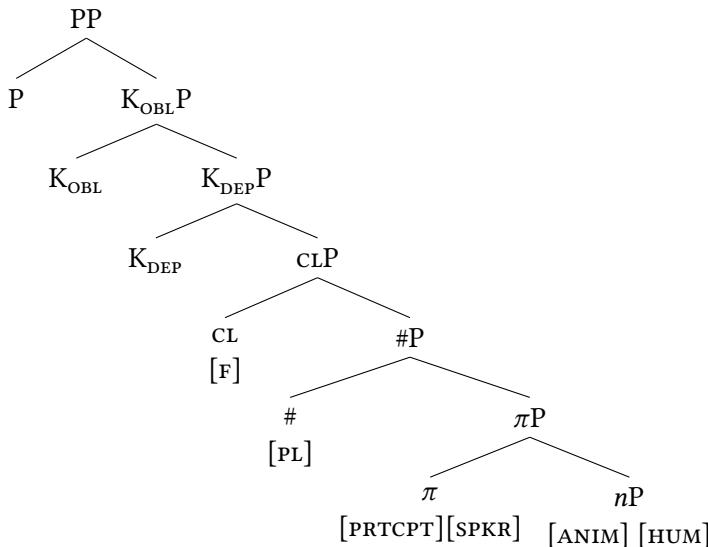


Figure 3: Proposed model of the structure of a BCMS pronoun

of KP). First person pronouns differ from second person ones in having the additional [SPKR] feature.¹²

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

The proposed structures for 3rd-person pronouns are presented in Figures 6–7. In the singular, due to the absence of number, their ‘P will be dominated by πP 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

¹²A reviewer wonders how local-person pronouns can control gender agreement without having overt grammatical gender features. Recall from §4.1 and footnote 10 that I assume that natural gender is present on the ‘P of local-person pronouns, following Puškar (2018), Puškar-Gallien (2019). From there it can enter agreement relations.

¹³Alternatively, 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.

6 Morphosemantic mismatches with pronouns

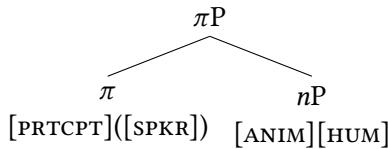


Figure 4: Singular local-person pronoun

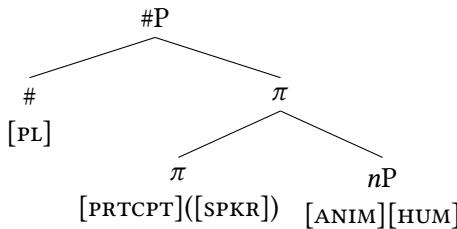


Figure 5: Plural local-person pronoun

#P. The combination of these two phrases will define the inflectional affixes of the pronouns. The ‘‘P lacks features if the pronoun denotes an inanimate entity. With an animate (or human) referent, these features will be present on the ‘‘P.

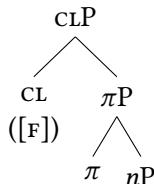


Figure 6: Singular 3rd-person pronoun

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 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

Zorica Puškar-Gallien

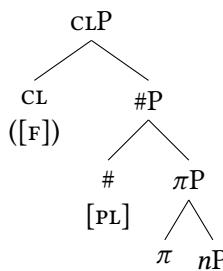


Figure 7: Plural 3rd-person pronoun

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 sub-nodes 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 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

¹⁴First 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, Djeebana, 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.

6 Morphosemantic mismatches with pronouns

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 §2.1, however this is outside the scope of the current paper.

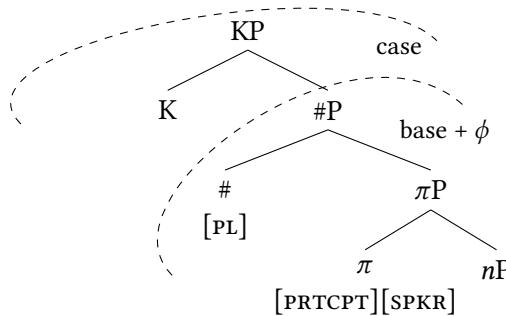


Figure 8: Local person

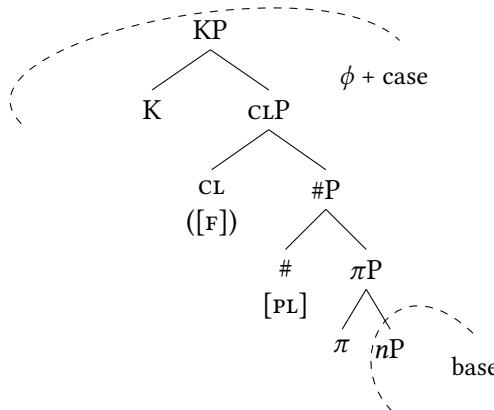


Figure 9: 3rd person

We will furthermore see that spelling out ‘‘P independently, i.e. effectively deleting it, is what enables a certain amount of flexibility to clitics that strong

Zorica Puškar-Gallien

pronouns lack. Under the assumption that the ‘‘P is a locality domain and as 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 ‘‘P deletion. I will build on this below in exploring the syntactic consequences of the given structures.

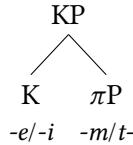


Figure 10: Local person clitic singular

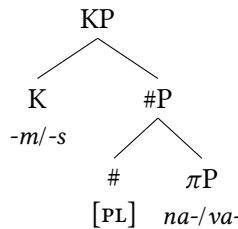


Figure 11: Local person clitic plural

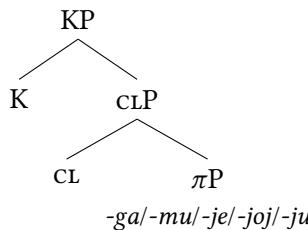


Figure 12: 3rd person clitic singular

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 ‘‘P, such that with clitics it is not realised.

6 Morphosemantic mismatches with pronouns

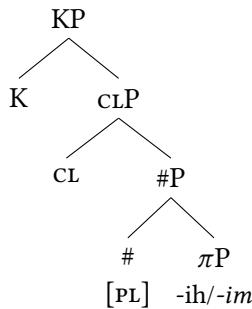


Figure 13: 3rd person clitic plural

4.5 Consequences for animacy and referentiality

The proposal above has direct consequences for the interpretational properties of pronouns presented in §2.2. Since clitics lack the ‘‘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 ‘‘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 interpretation

Zorica Puškar-Gallien

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 ‘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. §5.1 explores the general properties of movement of (pro)nominal elements, and §5.2–§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*

6 Morphosemantic mismatches with pronouns

- 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 vP 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),

Zorica Puškar-Gallien

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 §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 (Niloctic).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 *kēek*. This pronoun matches the fronted pronoun only partially – in number, but not in person.

- (20) a. wôök c̄ii bôl {kēek / *wôök} t̄iij
 1.PL PRF.OV Bol.GEN 3.PL 1.PL see.INF
 'Us, Bol has seen.'
- b. wêek c̄ii bôl {kēek / *wêek} t̄iij
 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 pronunciation 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

¹⁵See also Bošković (2001: et seq.) for a copy-based account of clitic placement in Serbo-Croatian.

6 Morphosemantic mismatches with pronouns

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 vP 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 ‘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 ‘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 ‘P (the locus of π under his account), the remaining copy need not match in person. In my account below, deleting the ‘P will exactly amount to spelling out a clitic, and I will assume that deleting the ‘P also deletes all of the contents of its sub-hierarchy.

5.2.2 Derivation

Following [van Urk \(2018\)](#), I will assume that pronominal ‘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 vP 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, ‘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 ‘P and the highest KP gets realised, but not the ‘P itself. In my system this amounts exactly to a realisation of a clitic, as illustrated in Figure 14.

The deletion of the ‘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

Zorica Puškar-Gallien

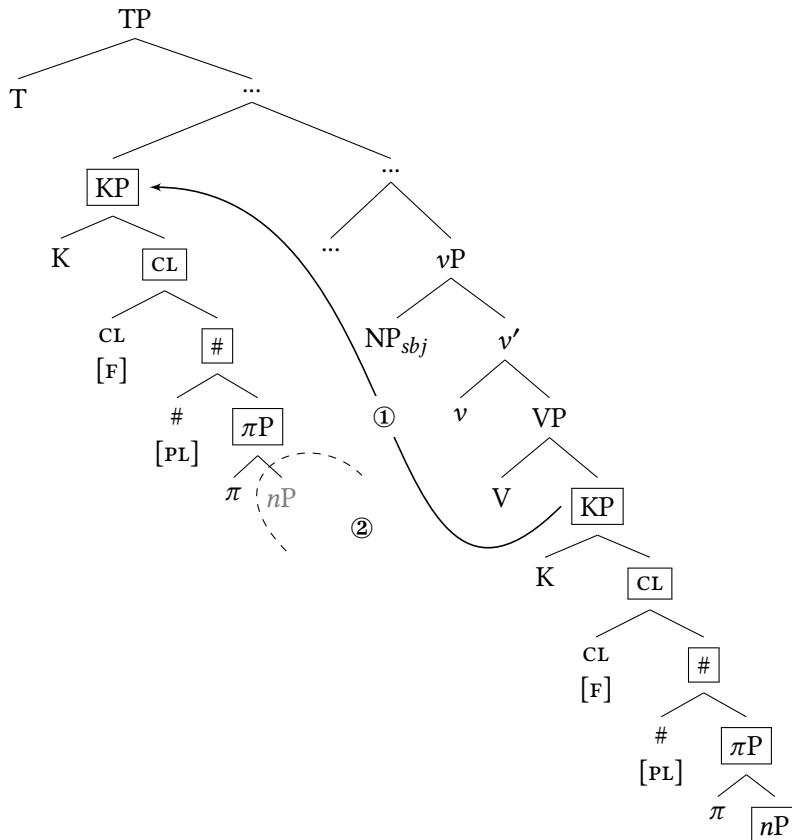


Figure 14: Pronoun movement, resulting in a clitic (here e.g. 3.F.PL)

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

$$[_{PP} <KP> P [KP \boxed{K} [CLP \boxed{CL} CL [\#P \# [\piP \pi [·P ·]]]]]]]$$


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 ‘‘P must be realised, and exactly in

6 Morphosemantic mismatches with pronouns

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 ‘‘P 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

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 ‘‘P or enforce its phonological realisation. Recall from examples (12) – (13) from §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

¹⁶The 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 ‘‘P 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 ‘‘P somewhat longer though, at least until the next phase head is merged. This would result in the possibility of realising the nP within the PP phase due to recoverability and economy, while the animacy features would be inaccessible.

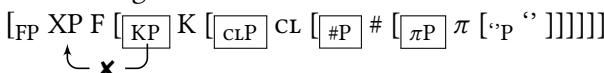
Zorica Puškar-Gallien

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 (to appear). 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 ‘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



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

¹⁷As 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

6 Morphosemantic mismatches with pronouns

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.

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 ‘‘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 ‘‘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 ‘‘P leaves us with a clitic, interpreted as either animate or inanimate, and either sloppy or strict. The deletion of the ‘‘P was implemented using van Urk’s (2018) theory of pronominal copying. A benefit of this analysis was that cases where the ‘‘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

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).

Zorica Puškar-Gallien

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

I would like to thank Artemis Alexiadou, Karolina Zuchewicz, the members of ZAS's Research Area 3 "Syntax and Lexicon" and the audience at FDSL 15 for their valuable comments and feedback. I also appreciate the comments of two anonymous reviewers and the editors, whose insights were of great help in writing the paper.

6 Morphosemantic mismatches with pronouns

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, MI: Michigan Slavic Publications.
- Abels, Klaus. 2012. *Phases: An essay on cyclicity in syntax*. Berlin: de Gruyter.
- Acquaviva, Paulo. 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. *Philologija 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 Milošević (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](https://doi.org/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. 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, NY: 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 & Ken Hale. 1996. The structural determination of case and agreement. *Linguistic Inquiry* 27(1). 1–68. <https://www.jstor.org/stable/4178925>.

Zorica Puškar-Gallien

- Borer, Hagit. 2005. *Structuring sense. Volume 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, MA: University of Massachusetts.
- Brody, Michael. 2000. Mirror theory: Syntactic representation in perfect syntax. *Linguistic Inquiry* 31(1). 29–56. DOI: [10.1162/002438900554280](https://doi.org/10.1162/002438900554280).
- Brody, Michael & Anna Szabolcsi. 2003. Overt scope in Hungarian. *Syntax* 6(1). 19–51. DOI: [10.1111/1467-9612.00055](https://doi.org/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, MI: 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](https://doi.org/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](https://doi.org/10.1162/002438902760168554).
- Despić, Miloje. 2011. *Syntax in the absence of Determiner Phrase*. Storrs, CT: 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](https://doi.org/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](https://doi.org/10.17161/SLS.1808.4408).
- Elbourne, Paul. 2005. *Situations and individuals*. Cambridge, MA: MIT Press.

6 Morphosemantic mismatches with pronouns

- Fassi Fehri, Abdelkader. 2000. Distributing features and affixes in Arabic subject verb agreement paradigms. In Jacqueline Lecarme, Jean Lowenstamm & Uri 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, MD: Lexington Books.
- Foley, Steven & Maziar Toosavandani. 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](https://doi.org/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](https://doi.org/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, IN: Indiana University.
- Godjevac, Svetlana. 2000. *Intonation, word order and focus projection in Serbo-Croatian*. Columbus, OH: The Ohio State University. (Doctoral dissertation).
- Harbour, Daniel. 2007. Against PersonP. *Syntax* 10. 223–243. DOI: [10.1111/j.1467-9612.2007.00107.x](https://doi.org/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*. Cambridge, MA: 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](https://doi.org/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](https://doi.org/10.1093/jos/ffs005).

Zorica Puškar-Gallien

- 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](https://doi.org/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, Boston: 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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.), *Језик, književnost, teorija*, 153–168. Niš: Faculty of Philosophy.
- Moskal, Beata. 2015a. *Domains on the border: Between morphology and phonology*. Storrs, CT: 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](https://doi.org/10.1162/LING_a_00185).

6 Morphosemantic mismatches with pronouns

- 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](https://doi.org/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](https://doi.org/10.1007/s11049-011-9150-4).
- Noyer, Rolf. 1992. *Features, positions, and affixes in autonomous morphological structure*. Cambridge, MA: 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, MA: 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.
- Progrovac, 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](https://doi.org/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](https://doi.org/10.5334/gjgl.587).
- Puškar-Gallien, Zorica. to appear. 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](https://doi.org/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](https://doi.org/10.4467/23005920SPL.21.006.13960).
- Runić, Jelena. 2014. *A new look at clitics, clitic doubling, and argument ellipsis: Evidence from Slavic*. Storrs, CT: University of Connecticut. (Doctoral dissertation).

Zorica Puškar-Gallien

- Sauerland, Uli. 2013. Presuppositions and the alternative tier. *Proceedings of SALT 23*. 156–173. DOI: [10.3765/salt.v23i0.2673](https://doi.org/10.3765/salt.v23i0.2673).
- Sichel, Ivy & Maziar Toosarvandani. 2024. The featural life of nominals. *Linguistic Inquiry*. DOI: [10.1162/ling_a_00517](https://doi.org/10.1162/ling_a_00517).
- Sichel, Ivy & Maziar Toosarvandani. to appear. Nominal intervention and the extended PERSON domain. In Heidi B. Harley, Romero Diaz, Damian Yukio, Rebecca M. Whitney, Tianyi Ni, Jianrong Yu, Hall Hartley & et al. Lucy Jane (eds.), *39th West Coast Conference on Formal Linguistics*. DOI: [10.25422/azu.data.c.5325401.v2](https://doi.org/10.25422/azu.data.c.5325401.v2).
- Siewierska, Anna. 2013. Gender distinctions in independent personal pronouns. In Matthew S. Dryer & Martin Haspelmath (eds.), *WALS online (v2020.3) [Dataset]*. Zenodo. DOI: [10.5281/zenodo.7385533](https://doi.org/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 and Linguistic Theory* 37. 1029–1101. DOI: [10.1007/s11049-018-9425-0](https://doi.org/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 pridjevskog 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](https://doi.org/10.2478/qal-2018-0001). https://www.researchgate.net/publication/324482938_Mixed_gender_agreement_in_the_case_of_Russian_hybrid_nouns/fulltext/5acf7f6b4585154f3f46652c/Mixed-gender-agreement-in-the-case-of-Russian-hybrid-nouns.pdf.
- 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.

6 Morphosemantic mismatches with pronouns

- Stojanović, Danijela. 1997. Object shift in Serbo-Croatian. In James R. Black & Virginia Motapanyane (eds.), *Clitics, pronouns and movement*, 301–320. Amsterdam/Philadelphia: John Benjamins.
- Sudo, Yasutada. 2012. *On the semantics of phi features on pronouns*. Cambridge, MA: 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](https://doi.org/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](https://doi.org/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](https://doi.org/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. to appear. Complementizer agreement is clitic doubling: Evidence from intervention effects in Frisian and Limburgian. *Natural Language and Linguistic Theory*. 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](https://doi.org/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](https://doi.org/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, CA: CSLI Publications.

Chapter 7

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 the other hand, the DAT/LOC.SG morpheme is one of the examples where it is

¹To my knowledge, there are no major differences between the four varieties when it comes to assibilation. As clarified in §3, the empirical basis for this study originates from a Croatian and a Serbian corpus.

Marko Simonović

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 assibilations.

Table 1: Four assibilations 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/ → [bajʃi] ‘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 assibilations. 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 assibilations behavior. The rationale behind this choice not to employ an overabundance analysis lies in the observation that variable assibilations 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 assibilations in DAT/LOC.SG. The factors, as outlined in standard descriptions, will serve as predictors in a statistical model to anticipate assibilations encountered in the corpus (see Lečić 2016 for an overview of statistical modeling

7 Animacy influences segmental phonology

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. §2 provides an overview of the predictors of assibilation and their treatment in the descriptive literature on BCMS. In §3, I describe the two corpus studies and the rationale behind them. §4 presents the results of the corpus studies, with the main finding being that animacy is a strong predictor of assibilation. §5 discusses the results and their theoretical implications. Finally, §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

Marko Simonović

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) that 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 §1, the NOM/VOC.PL morpheme *-i* almost always 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.

7 Animacy influences segmental phonology

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 §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., [suraka] ‘magpie.GEN.PL’).² On the other hand, in nouns

²In all traditional descriptions, this GEN.PL ending contains a long vowel and causes a lengthening of the preceding vowel, e.g. [suraka] ‘magpie.NOM.SG’ vs. [sura:ka:] ‘magpie.GEN.PL’. I am

Marko Simonović

with a stem-final consonant cluster three endings are attested: *-i* (e.g. in [krijki] ‘disguise.GEN.PL’), *-aa*, whose first vowel breaks up the consonant cluster, (e.g. in [banaka] ‘bank.GEN.PL’), and, somewhat marginally, *-a* (e.g., in [?][krijka] ‘disguise.GEN.PL’ and [?][baŋka] ‘bank.GEN.PL’). If endings can play a role in favouring/blocking allomorphy in other members of the paradigm, nouns which have GEN.PL in *-i* should be less prone to assibilations in DAT/LOC.SG than nouns which have other endings in GEN.PL. In other words, if *-i*[GEN.PL] is a strong factor that blocks assibilations, most nouns should behave either as [krijka] ‘disguise’ (i.e., have a GEN.PL in *-i* and no assibilations in DAT/LOC.SG) or as [baŋka] ‘bank.GEN.PL’ (i.e., not have a GEN.PL in *-i* and exhibit assibilations in DAT/LOC.SG).

Table 3: Assibilations in [baŋka] and lack of assibilations in [krijka]

	SG	PL	SG	PL
NOM	baŋk-a	baŋk-e	krijk-a	krijk-i
GEN	baŋk-e	banak-a	krijk-e	krijk-i
DAT/LOC	bants-i	baŋk-ama	krijk-i	krijk-ama
ACC	baŋk-u	baŋk-e	krijk-u	krijk-e
VOC	baŋk-o	bank-e	krijk-o	krijk-e
INS	baŋk-om	baŋk-ama	krijk-om	krijk-ama

2.3 Semantic factors

All traditional descriptions contain lists of classes in which assibilations are 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 assibilations, overviews as the one sketched above justify the implementation of ANIMACY as one of the factors that blocks assibilations.

ignoring both vowel length and suprasegmental information here, because there is considerable variation in this respect, including large numbers of speakers who don’t have distinctive vowel length.

7 Animacy influences segmental phonology

Indirect evidence that animacy plays a role in blocking assibililation 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 assibililation becomes possible with the inanimate referent. Examples are *Podravka* (the inhabitant of the Podravina region or a factory in Koprivnica, Croatia), *Beogradanka* (the female inhabitant of Belgrade or a building in Belgrade), and *Japanka* (a Japanese woman) vs. *japanka* (flip flop). Most normative sources argue against assibililation 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 that assibililation 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 assibililation in DAT/LOC.SG, which allows for a considerable amount of variation, the relevant construct to be employed here is the ASSIBILATION RATIO. The assibililation ratio of a word is the proportion of the DAT/LOC.SG forms with assibililation, calculated as the number of DAT/LOC.SG forms with assibililation divided by the total number of DAT/LOC.SG forms. For instance, if three DAT/LOC.SG tokens of the word [lozin̊ka] ‘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, C₁-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, C₁-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 C₁-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 nouns and only eight *g*-final nouns. 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"]

³The 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.

⁴The 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.

7 Animacy influences segmental phonology

- b. [word = ".*ih"] [word = ".*ih"] [word = "fres(ki|aka|ka)"]

The nouns for which one of the searches 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, C₁-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, C₁-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 searches yielded an empty result

⁵Data 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.

⁶As 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.

Marko Simonović

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 C₁-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 generalized linear mixed model. The binary variables MONOSYLLABIC STEM, C₁-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](#)

7 Animacy influences segmental phonology

Table 4: Mean ASSIBILATION RATIO for each value of the binary variables

Variable	Mean AR for 1 (N)	Mean AR for 0 (N)	Difference
ANIMACY	0.04 (77)	0.51 (127)	-0.47
C ₁ -SONORITY	0.39 (160)	0.13 (44)	0.26
MONOSYLLABIC STEM	0.40 (80)	0.29 (124)	0.11

al. 2015). The complete script is available in Simonović (2024). The summarized results can be found in Table 5.⁷

```
(3) modell <- glmer(assib ~ mono + anim + clson + igenpl + (1 | noun), family = binomial(link = "logit"), data = analysis1)
```

Table 5: Generalized 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
MONOSYLLABIC STEM	0.1950	0.6303	0.309	0.75703	1.2153
ANIMACY	-6.7532	0.8073	-8.366	< 2 × 10 ⁻¹⁶ ***	0.0012
C ₁ -SONORITY	-4.9472	0.7643	-6.473	9.62 × 10 ⁻¹¹ ***	0.0071
-i[GEN.PL] RATIO	-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 corresponds to 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,

⁷The significance codes used in this report follow the standard R output format: ‘***’ for *p*-values ≤ 0.001, ‘**’ for *p*-values ≤ 0.01, ‘*’ for *p*-values ≤ 0.05, ‘.’ for *p*-values ≤ 0.1, and no extra symbol for *p*-values > 0.1. These codes are retained from the R output for consistency in reporting results.

Marko Simonović

we consider effects where $p < .05$ 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.

Summarizing the findings, Table 5 indicates that among the four predictors, only two exhibit a highly statistically significant relationship with ASSIBILATION: ANIMACY and C₁-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.

7 Animacy influences segmental phonology

Table 7: Mean ASSIBILATION RATIO for the three values of FINAL VELAR

Mean AR for k (n)	Mean AR for g (n)	Mean AR for x (n)
0.86 (248)	0.54 (82)	0.19 (19)

Both binary predictor variables discussed above, as well as the categorical predictor variable FINAL VELAR 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: Generalized 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
FINAL VELAR:G	-5.2410	0.8887	-5.897	$1.23 \times 10^{-9}***$	0.00529
FINAL VELAR:X	-10.0896	1.6726	-6.032	$1.62 \times 10^{-9}***$	4.15×10^{-5}
ANIMACY	-9.9452	1.2072	-8.238	$< 2 \times 10^{-16}***$	4.80×10^{-5}
MONOSYLLABIC STEM	-6.3654	0.7886	-8.072	$6.90 \times 10^{-16}***$	0.00172

First, it is important to note that the categorial variable FINAL VELAR was dummy coded, resulting in two of its values appearing in the list. The baseline value, FINAL VELAR:K, is used as the reference category for comparison.

The model results indicate that all incorporated predictors exhibit a negative association with ASSIBILATION. Specifically, for the variable FINAL VELAR, which was omitted in Study 1, both g and x demonstrate negative associations with the outcome. Similarly, ANIMACY maintains a negative association, consistent with the findings of Study 1. Notably, unlike in Study 1, MONOSYLLABIC STEM 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 §2 is in order.

Marko Simonović

Regarding the phonological factors, C₁-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 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 realize 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 two other 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', [svastika] 'sister-in-

⁸The 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 [kauga] 'conflict' as the only C_g-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

7 Animacy influences segmental phonology

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 don't 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 suggests 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 [t̪ex] 'Czech man.NOM.SG', [t̪esi] 'Czech man.NOM.PL', [t̪esima] 'Czech man.DAT/LOC/INS.PL').

⁹As 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.

Marko Simonović

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 formalizing 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

The author thanks the audience of FDSL 15 and the anonymous reviewers, as well as Peđa Kovačević and the participants of the course Collecting and Analyzing Corpus and Experimental Data in Hypothesis-Driven Linguistic Research held at the University of Novi Sad. Special thanks go to Igor Marchetti and Maja Miličević Petrović for all the help with the statistical analysis. This research was funded in whole by the Austrian Science Fund (FWF, Grant-DOI 10.55776/I6258).

7 *Animacy influences segmental phonology*

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](https://doi.org/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](https://doi.org/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, Sweden: Association for Computational Linguistics. DOI: [10.3115/v1/W14-0405](https://doi.org/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, MA.
- 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.

Marko Simonović

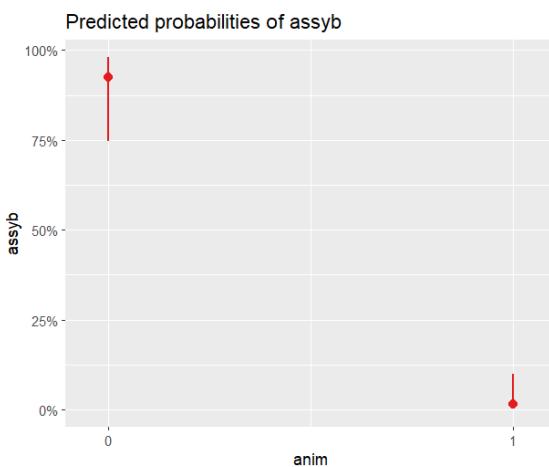
- Petrović, Dragoljub & Snežana Gudurić. 2010. *Fonologija srpskog jezika* [Phonology of Serbian]. Beograd: Institut za srpski jezik SANU / Beogradska knjiga / Matica Srpska.
- Simonović, Marko. 2024. *Sibilarisation in the Dative/Locative Singular in BCMS (Data from hrWaC and srWac) [Data set]*. DOI: [10.5281/zenodo.1382533](https://doi.org/10.5281/zenodo.1382533).
- Težak, Stjepko. 1986. Sibilarizacija u suvremenom 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.

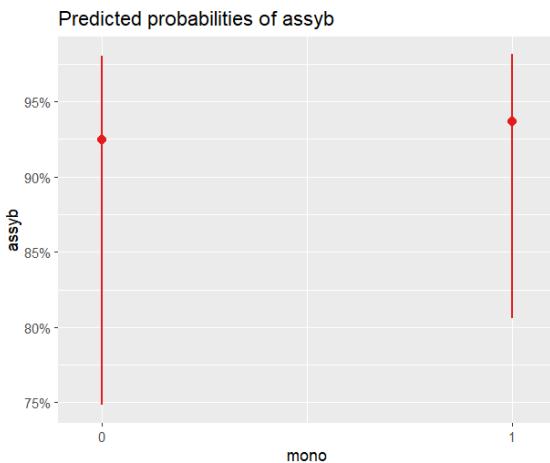
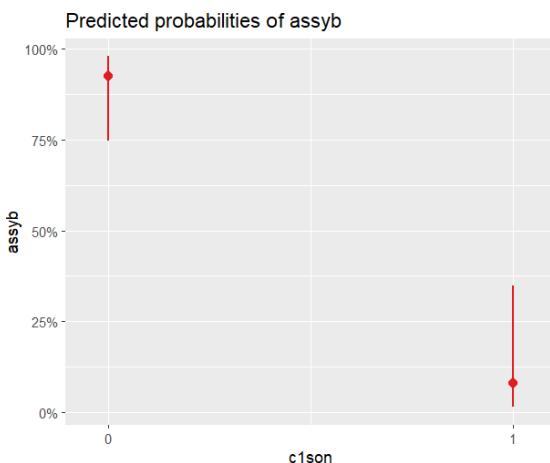
Appendix A Appendix: Plots for predicted probabilities

Below I report the plots for predicted probabilities for each predictor. These were obtained in R using the package the package sjPlot ([Lüdecke 2024](#)) in R. The formula implemented as `plot_model(model1, type = "pred", terms = "name_of_the_predictor")`.

A.1 Study 1

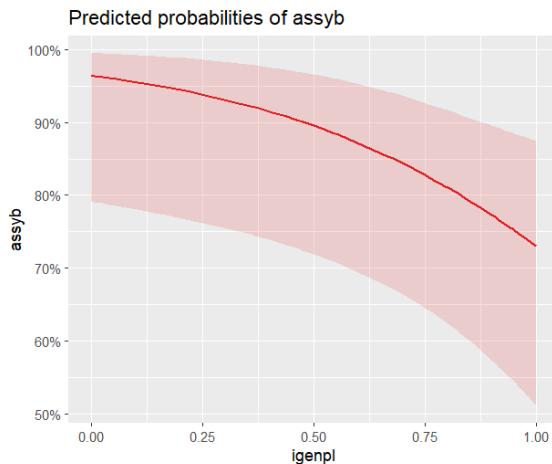
6.0.1 ANIMACY



*7 Animacy influences segmental phonology***6.0.2 MONOSYLLABIC STEM****6.0.3 C₁-SONORITY**

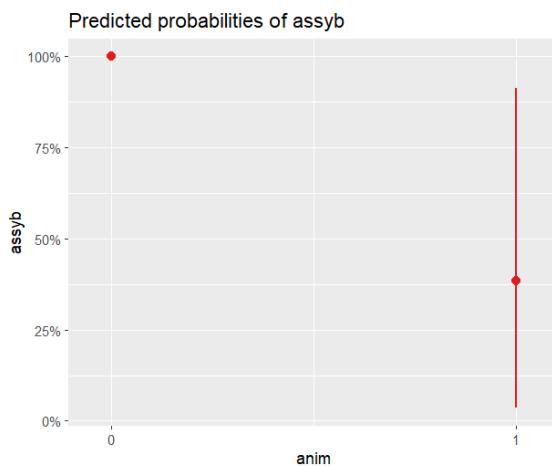
Marko Simonović

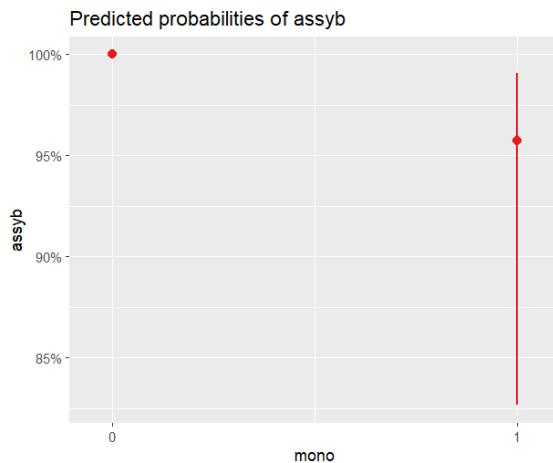
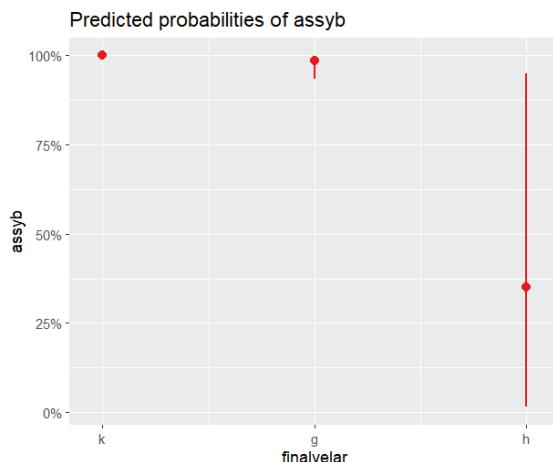
6.0.4 I[GEN.PL] RATIO



A.2 Study 2

6.0.5 ANIMACY



7 *Animacy influences segmental phonology***6.0.6 MONOSYLLABIC STEM****6.0.7 FINAL VELAR****Appendix B Appendix: 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:

Marko Simonović

```
[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
    tidyverse_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 8

How departicipial are “L-participle” nominalisations in Western South Slavic

Marko Simonović^a, Petra Mišmaš^b, Stefan Milosavljević^a, Boban Arsenijević^a, Katarina Gomboc Čeh^b, Franc Lanko Marušić^b & Rok Žaucer^b

^aUniversity of Graz ^bUniversity of Nova Gorica

We focus on nominalisations seemingly derived from L-participles, illustrated by *lec*-nominalisations in Slovenian, in order to establish the nature and position of the L-morpheme as well as the structure of these nominalisations in general. Our research is situated in the current debates on whether the item L in L-participles and L-nominalisations is the same morpheme or two different morphemes, and if the former, whether L-nominalisations are derived from L-participles. We argue that the L-morpheme is a root in both, but also show that it is not the case that *lec*-nominalisations contain L-participles. The *lec*-nominalisations are argued to contain a smaller structure than the corresponding L-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 L-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'

Simonović et al.

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 deparcipial nominalisations, which also include nominalisations illustrated in (3), standardly analysed as derived from the passive participles (e.g., [Toporišić 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

¹In 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’ ([Toporišić 2000](#): 163–164). We take this to be lexically conditioned allomorphy.

²Note 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. in preparation](#)).

8 “L-participle” nominalisations

- (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*-nominalisations (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 L-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 L-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’

³The 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.

Simonović et al.

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 L-morpheme and the passive-participle N/T-morpheme, correspond to the same conceptually empty root.⁴ Focusing on the L, we argue that L-participles contain a richer structure than the corresponding portion of L-nominalisations, while no such difference is found with the passive-participle N/T-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 L-initial deverbal suffixes into an L-morpheme and another suffix that is added on top of it (e.g., -ec). In §3, we discuss the nature of the L-morpheme in the nominalisations under consideration. §4 presents our account of the structural position of the L-morpheme and the theme vowel restrictions in the respective nominalisations. §5 concludes the paper.

2 Internal structure: Severing L 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 L-morpheme preceded by a theme vowel are by far the most common in Slovenian (see Marvin 2016, 2015 quoting Stramlič Breznik 1999).⁵ Example (7), taken from Marvin (2002: 98, (22)), illustrates three such nominalisations. According to Marvin, the three affixes added to the L-morpheme (-ec, -k and -Ø) are variants of the same affix deriving nouns of three different genders. All three nominalisations are generally related to external arguments, 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

⁴The 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.

⁵Stramlič 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.

⁶As 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.

8 “L-participle” nominalisations

- (7) a. brus-i-l-ec-∅
 sharpen-TV-L-er-NOM.SG
 ‘a sharpener.M’
- b. brus-i-l-k-a
 sharpen-TV-L-er-NOM.SG
 ‘a sharpener.F’
- c. brus-i-l-∅-o
 sharpen-TV-L-er-NOM.SG
 ‘a sharpening device’

Moreover, it has long been observed that various nominalisations and adjectivisations share the same L-final base. Some further affixes (both nominal and adjectival) that can combine with the L-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 PLACE are two different items. While *-išč* consistently results in a place interpretation, *-ic* is only associated with this interpretation when in the context of *-l* 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-TV-L-ADJ wait-TV-L-ADJ-PLACE-NOM.SG wait-TV-L-PLACE-NOM.SG
 ‘waiting.A’ | ‘waiting room’ | ‘waiting spot’

These items cannot be modified by an adverb such as *hitro* ‘fast.ADV’.

- (i) a. *hitro / hitrega brus-i-l-c-a
 fast.ADV fast.GEN.SG.M sharpen-TV-L-er-GEN.SG
 ‘a fast sharpener.M’
- b. hitre brus-i-l-k-e
 fast.GEN.SG.F sharpen-TV-L-er-GEN.SG
 ‘a fast sharpener.F’
- c. hitrega brus-i-l-∅-a
 fast.GEN.SG.N sharpen-TV-L-er-GEN.SG
 ‘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.

Simonović et al.

The existence of such families of related derivations is a strong argument for the decomposition of *-lec* into *L* and *ec*, but also an argument for recognising *L* 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 *L*-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'
 - b. hod- | hod-ec-∅
 $\sqrt{\text{WALK}}$ walk-*er*-NOM.SG
| 'walker'
 - c. krog | krog-ec-∅
circle cirle-*er*-NOM.SG
'circle' | 'small circle'
-
- (10) a. jekl-en | jekl-en-k-a
steel-ADJ steel-ADJ-*er*-NOM.SG
'steel' | 'gas cylinder'
 - b. hod- | hod-k-a
 $\sqrt{\text{WALK}}$ walk-*er*-NOM.SG
| 'walker'

⁷The 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.

⁸The 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.

8 “L-participle” nominalisations

c. adidas	adidas-k-a
Adidas	adidas- <i>er</i> -NOM.SG
	‘adidas-shoe’

These facts unequivocally demonstrate that the item *-lec* is complex, comprising two distinct items, namely L 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 L-morpheme in these nominalisations and its relation to the L-morpheme that surfaces in the L-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 L?

The question of the status of the L-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 L in nominalisations

For Slovenian, some authors take *-lec* to be a single morpheme, e.g., Toporišič (2000) treats *-lec*, as in example (1), and *-ec*, as in (9), as two separate items. The issue of L in *lec*-nominalisatons is considered in Stramljič Breznik (1999), who mentions as a possible answer Bajec’s (1950) proposal that L 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 L-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 L-participle encodes past. Given that L-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 L-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

⁹In fact, even as a part of the perfect form, traditionally analysed as past tense in BCMS, its

Simonović et al.

(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 form *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 T₂/Participle head¹⁰. 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 vP) 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,

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.
eaten.M.SG AUX.1.SG
'I ate/I've eaten. (i.e., I'm not hungry.)'

(BCMS)

¹⁰ Marvin (2002) distinguishes between two T(ense) heads, T₁ and T₂, 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.

8 “L-participle” nominalisations

unaccusatives do not have an external argument position). This tentative prediction is confirmed by the WeSoSlaV database (Arsenijević et al. in preparation), but see also Marjanović et al. (2013). Out of 728 *lec*-nominalisations in the Slovenian sub-base 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 not derive *lec*-nominalisations (**porumenelec* ‘(intended) someone who becomes yellow’).¹²

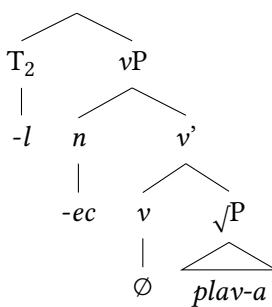


Figure 1: Structure proposed by Marvin (2002: 99, (25)).

The analysis just outlined, under which these nominalisations are derived from L-participles, finds its support in the fact that virtually all *lec/lac*-nominalisations can be derived from an L-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 L 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

¹²In fact, even these three examples can be successfully accounted for under an alternative account. 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)’ (form *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 L-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.

Simonović et al.

l-participle nominalisations". In these nominalisations the root together with the theme vowel is the complement of a Part(iciple)P (headed by the *l*-morpheme), which gets nominalised by *-ec/k/Ø*. Crucially, Marvin argues that in these root nominalisations (unlike the deverbal *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 (SpecvP).

- (12) a. rez- | rez-a-ti | rez-e-mo | rez-a-l-Ø-o | rez-i-l-Ø-o
 $\sqrt{\text{CUT}}$ | 'cut.INF' | 'cut.PRS.1PL' | 'cutter' | 'blade'
 b. barv- | barv-a-ti | barv-a-mo | barv-a-l-Ø-o | barv-i-l-Ø-o
 $\sqrt{\text{COLOR}}$ | 'color.INF' | 'color.PRS.1PL' | 'coloring device' | 'pigment'
 c. god- | gos-Ø-ti /god-Ø-ti/ | god-e-mo | god-a-l-Ø-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).

8 “L-participle” nominalisations

- (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 L, the fact that these examples are not derived from a verb clearly implies that they are also not derived from L-participles.

3.2.2 L is a root, and the importance of the theme vowel set

An account of the L that participates in deverbal derivations is given in Simonović & Mišmaš (2022), where the focus is on adjectival L-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 L-participles that we can find in complex tenses, and adjectival L-participles, which only derive from unaccusative verbs (see fn. 12), arguing that L is a root in both. This root can merge with either a root or a phrase. In adjectival participles, L is merged with a root, whereas in past participles it is merged with a verbalised structure. In addition, taking L to be a root then allows Simonović & Mišmaš (2022) to offer a unified account even for L 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 into allowed theme vowel classes, there is a discrepancy between the adjectival L-participles in Simonović & Mišmaš (2022) and the

¹³This 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).

Simonović et al.

derivations discussed in this paper. The set of morphemes that can precede the L-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 non-finite forms, which are relevant here, and 863 have *i*; the situation is similar in BCMS). On the other hand, in the nominalisations and adjectivisations discussed in this paper, what can precede the L-morpheme is exactly this set of theme vowels. This holds not only of examples such as those in (13), where the base verb is not attested independently. It also holds of agentive *lec*-nominalisations. (14) illustrates *lec*-nominalisations from four transitive verbs where theme vowels Ø and *e* are replaced by *i*.

- (14) a. ves-Ø-ti /vez-Ø-ti/ – vez-e-l /vez-Ø-l/ – vez-i-l-ec
embroider-TV-INF embroider-TV-L embroider-TV-*er.NOM.SG*
‘to embroider’ ‘embroidered’ ‘embroiderer’
- b. ples-Ø-ti /plet-Ø-ti/ – plet-e-l /plet-Ø-l/ – plet-i-l-ec
knit-TV-INF knit-TV-L knit-TV-*er.NOM.SG*
‘to knit’ ‘knitted’ ‘knitter’
- c. gnes-Ø-ti /gnet-Ø-ti/ – gnet-e-l /gnet-Ø-l/ – gnet-i-l-ec
knead-TV-INF knead-TV-L knead-TV-*er.NOM.SG*
‘to knead’ ‘kneaded’ ‘kneader’
- d. vrt-e-ti – vrt-e-l – vrt-i-l-ec
spin-TV-INF spun-TV-L spin-TV-L-*er-NOM.SG*
‘to spin’ ‘spun’ ‘spinner’

This change in the theme vowel is noted also in [Toporišić \(2000: 163–164\)](#), 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 Ø theme vowel), the vowel is realised as *i*. Based on the observation that verbs with a Ø theme vowel are nominalised with *-ilec*, [Stramlijč 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 L-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 L-participles the complement of L has been argued to be a root ([Simonović & Mišmaš 2022](#)).¹⁴

¹⁴Note that the same restriction to the theme vowels *a* and *i* is attested in much less productive nominalisations derived from short infinitives.

8 “L-participle” nominalisations

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 **L** under discussion, even the ones in (12), implies that all these nominalisations are in fact deverbal. This means that both L-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 L-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-ava-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-ava-ti – ??prikim-ova-ti – prikim-ova-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 L

The final type of evidence featuring in DM approaches to *lec*-nominalisation is their prosodic behaviour. All nominalisations and adjektivisations containing the L-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

Simonović et al.

can be seen in the L-participles, but these differences get neutralised in all other cases.¹⁵

- (16) a. mo'r-i-l vs. | mo'r-i-l-ec –
 murder-TV-L.M.SG murderer-TV-L-er.NOM.SG
 ‘murdered’ | ‘murderer.M’
 mo'r-i-l-k-a – mo'r-i-l-en
 murderer-TV-L-er-NOM.SG murder-TV-L-ADJ.M.SG
 ‘murderer.F’ ‘related to murder’
- b. 'mer-i-l vs. | me'r-i-l-ec –
 measure-TV-L.M.SG measure-TV-L-er.NOM.SG
 ‘measured’ | ‘measurer.M’
 me'r-i-l-k-a – me'r-i-l-en
 measure-TV-L-er.NOM.SG measure-TV-L-ADJ.M.SG
 ‘measurer.F’ ‘related to measuring’
- c. 'rez-a-l vs. | re'z-a-l-ec – re'z-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’
 re'z-i-l-∅-o
 cut-TV-L-er-NOM.SG
 ‘blade’
- d. i'gr-a-l vs. | i'gr-a-l-ec | i'gr-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’ |
 i'gr-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. 'merjen vs. | 'merjen-ec – 'merjen-ka –
 measure.PASS.PTCP measured-er.M measured-er.F
 ‘measured’ | ‘measured person.M’ ‘measured person.F’

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.

8 “L-participle” nominalisations

- 'merjen-je
 measured-*ing*
 'measuring'
- b. umor'jen vs. | umor'jen-ec umor'jen-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:

$$(18) \quad \sqrt{\text{MER(I)}} + \sqrt{L} + \sqrt{C} - \sqrt{\text{MER(I)}} + \sqrt{L} + \sqrt{N}$$

'measurer' 'measuring.ADJ'

Simonović et al.

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-l-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 *L* is not itself a root, but minimally a *vP*, 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 *L* 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 *L*-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 *me'r-i-l-ec*, given in (16b), is *me'r-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 *vP* *mer-i-* gets spelled out, then the *nP -l-c-* and then the case/number ending *-a* follows.¹⁷ Now, in the first cycle the faithful prosodic pattern wins: '*mer-i-*'. This output serves as the input to the second cycle, where there is a clash between the lexical prosody ('*merilc*') and the pattern imposed by the radical core ('*me'rilc*'). 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 *me'rilca*.

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 *L* 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

¹⁶A question that [Simonović \(2020\)](#) leaves open is the status of theme vowels (e.g., *mer-i* 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.

¹⁷See fn. 7 for the omission of *e*.

8 “L-participle” nominalisations

structure, while at the same time not containing full L-participles. The exact way in which these two verbal structures differ is the main focus of the following section.

4 An analysis

We pursue a unified analysis of the L morpheme in (i) L-participles, (ii) deverbal nominalisations with L, 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 L 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 L, the passive N/T or the T 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 only limit ourselves to N/T and L and leave short infinitives for future work. Second, prosodic patterns and the set of allowed theme vowels distinguish between L-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 in 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{EC} can be categorised (and in fact is categorised by an *n* head) and has the ability to select. It is 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 vP (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 vP results in a crash and thus requires

Simonović et al.

some extra operation or additional structure. We argue that insertion of the L-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 (N/T-) base, since passives are adjectival (and therefore an acceptable complement), cf. (9a).

In proposing root insertion of \sqrt{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{STROY} and a root extension \sqrt{DE} , which attaches to \sqrt{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{L} in nominalisations acts as an extension. However, unlike \sqrt{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{L} to trigger spell out, the roots \sqrt{L} and \sqrt{EC} form a radical core.

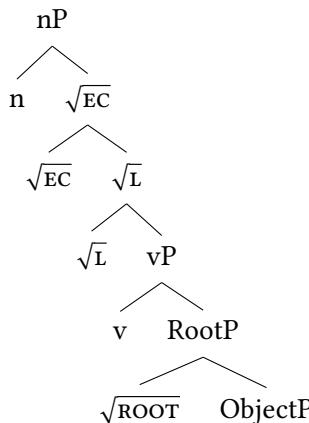


Figure 2: Lec-nominalisations

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

8 “L-participle” nominalisations

& 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 L or N/T. 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 L 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) $\sqrt{L} \leftrightarrow /n/ \backslash \quad [P A S S I V E]$
 $\qquad \qquad \qquad \leftrightarrow /l/ \quad \text{e l s e w h e r e}$

This then captures Marvin’s (2002) observation that L 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 L is merged in a functional head, we build on the analysis in Cavigani-Pots (2020) (see also Cavigani-Pots et al. 2021), who argues that semi-lexicality of some items emerges when (lexical) roots are merged/

¹⁸Another 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 L and N/T are merged in Asp⁰. 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 vPs (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), Łazarzczuk (2010), Tatevosov (2015, 2017), Mueller-Reichau (2020), Biskup (2023), Milosavljević (2023) – although their exact function varies across approaches.

¹⁹Given that the passive affixes are referred to as N/T throughout the paper, we are simplifying the Vocabulary Items and referring the reader to Marvin (2002: 92).

Simonović et al.

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 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 intermediate stance is that participial morphemes are meaningless at least in some

²⁰For 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 analysed 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

8 “L-participle” nominalisations

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 presented three families of approaches to participial morphemes. We immediately exclude the third option, i.e., the purely morphological analysis, ac-

“true” verbal participles by 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 vP 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).

²¹In Kratzer (2000) the lack of verbal inflection explains why adjectival passives miss 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.

²²Apart 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.

Simonović et al.

cording to which syntax does not play any role, as the discussed roots 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 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., vP) 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{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 vP 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.

²³The reader is also referred to Milosavljević & Arsenijević (2022), Kovačević et al. (2024) for arguments against these being purely ornamental morphemes, much like theme vowels.

8 “L-participle” nominalisations

4.3 The syntax of *lec*-nominalisations and theme vowel realisation

As discussed in 3.2.2, L-nominalisations and L-participles allow for 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 L-nominalisations only contain the vP (as they always have a theme vowel). Note that this does not hold for N/T-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, N/T-nominalisations display the same set of theme vowels as passive participles. This then also means that the N/T morpheme never functions as a true vP 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 Ø in the L-participle *vez-Ø-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 [α] and [β], 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).

$$(20) \quad TV \leftrightarrow /i/$$

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 [α].

$$(21) \quad TV_{[-\alpha]} \leftrightarrow /a/$$

Simonović et al.

Now we arrive at the vocabulary items for the two classes which have a shift to the elsewhere theme vowel in nominalisations.

- (22) $\text{TV}_{[+\alpha, -\beta]} \leftrightarrow /e/ \setminus _ [\text{VOICE}]$
 $\text{TV}_{[+\alpha, +\beta]} \leftrightarrow /\emptyset/ \setminus _ [\text{VOICE}]$
 ...

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 L-participles, exemplified by the *lec*-nominalisations in Slovenian, in order to determine the nature and position of the L-morpheme. One important point of comparison were the passive participles and the N-morpheme in nominalisations that are derived from passive participles.

We argue that the supposed L-participle nominalisations are not derived from participles in that there is no perfect containment relation between the L-participle and the *lec/lac-* nominalisations. Rather, *lec/lac*-nominalisations contain a smaller structure (*vP*) than the corresponding L-participle. This influences the spell out, and consequently, the insertion of theme vowels. In the proposed structure, L (but also N/T 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 L-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 analysis 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.

8 “L-participle” nominalisations

Abbreviations

1	first person	PASS	passive
ADJ	adjective	PFV	perfective
BCMS	Bosnian/Croatian/ Montenegrin/Serbian	PL	plural
AUX	auxiliary	PTCP	participle
F	feminine	SG	singular
INF	infinitive	Slo	Slovenian
IPFV	imperfective	SI	secondary imperfective
M	masculine	S.INF	short infinitive
N	neuter	TV	thematic vowel
NOM	nominative		

Acknowledgments

We acknowledge the financial support by the Austrian Science Fund FWF (grants I 4215 and I 6258), by the Slovenian Research and Innovation Agency ARIS (grants N6-0113, J6-4614, P6-0382 and BI-AT/23-24-021), as well as by the Austrian Federal Ministry of Education, Science and Research (BMBWF; Scientific & Technological Cooperation grant for the project *Verbal suffixes and theme vowels: Different or the same?*).

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, MA: MIT Press.

Simonović et al.

- 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](https://doi.org/10.1075/lfab.17.04ars).
- Arsenijević, Boban, Katarina Gomboc Čeh, Franc Marušić, Stefan Milosavljević, Petra Mišmaš, Jelena Simić, Marko Simonović & Rok Žaucer. in preparation. *WeSoSlav: Database of the Western South Slavic verbal system*. <https://hyperverb.ung.si/wiki/doku.php>.
- Arsenijević, Boban, Stefan Milosavljević & Marko Simonović. 2023. *Secondary imperfectivisation is reverbalisation is addition of theme vowels*. Presented at 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 jezikoslovje.
- 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](https://doi.org/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](https://doi.org/10.1023/b:ling.0000023371.15460.43).
- Borer, Hagit. 2005. *Structuring sense (volume 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](https://doi.org/10.5334/GJGL.1055).

8 “L-participle” nominalisations

- 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](https://doi.org/10.1515/probus-2020-0006).
- Cavirani-Pots, Cora. 2020. *Roots in progress: Semi-lexicity in the Dutch and Afrikaans verbal domain*. Utrecht: LOT.
- Cavirani-Pots, Cora, Marijke De Belder & Heidi Klockmann. 2021. *Semi-lexicity: 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](https://doi.org/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](https://doi.org/10.1162/002438900554343).
- Embick, David. 2004. On the structure of resultative participles in English. *Linguistic Inquiry* 35(3). 355–392. DOI: [10.1162/0024389041402634](https://doi.org/10.1162/0024389041402634).
- Embick, David & Morris Halle. 2005. On the status of stems in morphological theory. In Twan Geerts, Ivo van Ginneken & Haire Jacobs (eds.), *Romance languages and linguistic theory 2003*, 59–88. Amsterdam: John Benjamins. DOI: [10.1075/cilt.270.03emb](https://doi.org/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, MA: 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, MA: MITWPL.
- Klajn, Ivan. 2003. *Tvorba reči u savremenom srpskom jeziku. Drugi deo. Sufiksacija i konverzija* (Prilozi gramatici srpskoga jezika). Beograd, Novi Sad, 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](https://doi.org/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](https://doi.org/10.1017/S0022226723000415).

Simonović et al.

- 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, CA: 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](https://doi.org/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.
- Marjanović, 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](https://doi.org/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, MA: 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. *Jezikoslovje* 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](https://doi.org/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](https://doi.org/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, MA: Massachusetts Institute of Technology. (MA thesis).

8 “L-participle” nominalisations

- Progovac, Ljiljana. 2005. *A syntax of Serbian: Clausal architecture*. Bloomington, IN: 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 extra issue). 1–15. DOI: [10.1353/jsl.2022.a923080](https://doi.org/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](https://doi.org/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](https://doi.org/10.1016/j.lingua.2007.03.008).
- Ramchand, Gillian Catriona. 2018. *Situations and syntactic structures: Rethinking auxiliaries and order in English*. Cambridge, MA: 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1418/78407).
- Simonović, Marko & Boban Arsenijević. 2020. Syntax predicts prosody: Multi-purpose morphemes in Serbo-Croatian. In Franc Marušić, Petra Mišmaš & Rok Žaucer (eds.), *Advances in formal Slavic linguistics 2017*, 277–304. Berlin: Language Science Press. DOI: [10.5281/zenodo.3764865](https://doi.org/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 multi-functionality of the Slovenian affix ov. *Linguistica* 60(1). 83–102. DOI: [10.4312/linguistica.60.1.83-102](https://doi.org/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](https://doi.org/10.16995/glossa.5809).
- Smith, Carlota. 1997. *The parameter of aspect*. Dordrecht: Kluwer.

Simonović et al.

- 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](https://doi.org/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*. Manuscript. 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, CT: 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](https://doi.org/10.5334/GJGL.751).
- Wegner, Dennis. 2019b. *The underspecification of past participles: On the identity of passive and perfect(ive) participles*. Berlin / Boston: De Gruyter.
- Wegner, Dennis. 2021. The categorial, argument structural and aspectual indeterminacy of past participles: A holistic approach. *Zeitschrift für Sprachwissenschaft* 40(2). 199–249. DOI: [10.1515/zfs-2021-2027](https://doi.org/10.1515/zfs-2021-2027).

Chapter 9

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

Carla Spellerberg

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’, *săsto* ‘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 săsto [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. [PETĀR]_F săsto mož-e da gotvi.
Petăr also able-PRS.3SG to cook
'Petăr is able to cook as well.'

¹All non-English examples in this paper are from Bulgarian, unless marked otherwise next to the example.

²In 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.

³While 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 §4 of this paper.

9 Focus-sensitive particles in Bulgarian

- c. [ANA]_F dori ſte trjabva da dojd-e.
 Ana even AUX.FUT have da 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 *az/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 §2, I briefly present the few studies that have already been conducted on Bulgarian FSPs, give an overview over 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

Carla Spellerberg

& Hartmann's (2001) and Zanon's (2023) Particle Theories. In §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 §4. §5 concludes the paper.

2 Focus-sensitive particles in Bulgarian and beyond

The aim of this section is two-fold. 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 over 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 §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)).

⁴As 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).

9 Focus-sensitive particles in Bulgarian

- (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

Carla Spellerberg

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 SÄSEDKA-TA]_F
only cat-DEF of neighbor-DEF
'only the cat [of the neighbor]_F'
- b. kotka-ta samo [na SÄSEDKA-TA]_F
cat-DEF only of neighbor-DEF
'only the cat [of the neighbor]_F'
- c. kotka-ta [na SÄSEDKA-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 §3.2 of this paper).

- (7) a. * säs 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'

9 Focus-sensitive particles in Bulgarian

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.

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 *až/čak*.⁸ According to the author, *až/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

⁵No translation of *čak* is provided here due to its intriguing semantic properties that impede a direct translation into English.

⁶Nicolova (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.

⁷Nicolova (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*Až* 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.

⁹Transliteration changed to scientific translation

Carla Spellerberg

presuppositional and can also be the direct opposite of *only*, in contrast to *even* (Tomaszewicz 2013). In contrast to *až/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 §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

9 Focus-sensitive particles in Bulgarian

- 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 §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:

- (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.’

¹⁰See 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.

Carla Spellerberg

- 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 *chi* 'only' in Vietnamese and argues that *chi* 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, independent 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)

9 Focus-sensitive particles in Bulgarian

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) analysis 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, Zanon (2023: 420) argues that *tol'ko* and the F-marked constituent form an un-

¹¹While 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?*

Carla Spellerberg

splittable complex, as shown in Figure 1.^{12,13}

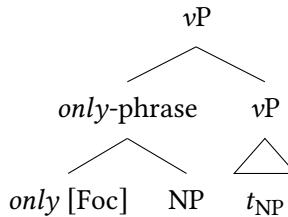


Figure 1: Structure of an *only*-phrase from Zanon (2023: ex. 7)

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)

¹²As 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.

¹³While the adjacency account correctly derives the particle placement facts for Bulgarian, the result cannot always be an unsplittable complex, as I discuss in §4 of this paper.

¹⁴A 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.

9 Focus-sensitive particles in Bulgarian

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'sju 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.

¹⁵It 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.

¹⁶Transliteration adapted to scientific transliteration in (16).

¹⁷A 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 §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.

Carla Spellerberg

- ii. # ...but I don't regret that I kissed Anastasia.
- 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 vP, 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 vP 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 ètot 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

¹⁸This analysis would run into problems under approaches that assume that the finite verb never moves out of vP 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.

¹⁹A 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 ètot 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, vP, and FP is too restrictive for Russian. Again, this problem does not affect my more permissive analysis of adjunction sites for Bulgarian FSPs.

9 Focus-sensitive particles in Bulgarian

by the existence of a functional projection FP that is able to host *tol'ko*.²⁰ As I discuss in §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 §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 (§3.1). §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 (§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 §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 sharply

²⁰Examples 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.

Carla Spellerberg

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’
(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

²¹A 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: *Pärvo da se glasuva pärvata 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.

9 Focus-sensitive particles in Bulgarian

(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 sreština-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

²²While 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 coor-

Carla Spellerberg

- (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

- (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.’

ordinated 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.

²³See Mursell (2021) for discussion.

²⁴A 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) Nova₁ e kupil [t₁ kola] (ne stara).
 new is bought car not old
 ‘He bought a new car, not an old one.’ (Tasseva-Kurtchieva & 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.

9 Focus-sensitive particles in Bulgarian

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.'
- (German; Büring & Hartmann 2001: exx. 82a, c)

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 §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

Carla Spellerberg

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 SÄSEDKA-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 kursa].
 I know only students.ACC first.GEN year.GEN
 ‘I only know the [FIRST]_F year students.’
- b. Ja znaju [studentov tol’ko PERVOGO 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 PÄRVA godina]_F.
 know-1SG only student-PL first year
 ‘I only know the [FIRST]_F year students.’

²⁵I 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.

9 Focus-sensitive particles in Bulgarian

- 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, such as [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

Carla Spellerberg

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 §[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 vP, CP, and FP. Please note that [Zanon's](#) analysis

9 Focus-sensitive particles in Bulgarian

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 étot fil'm.
 watched Q only Ivan this movie
 intended: 'if only [Ivan]_F watched this movie.'
 - b. posmotrel li Ivan tol'ko [ÉTOT]_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 vP, 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.

- (34) Az ne znaj-a...
 1SG NEG know-1SG
 'I don't know...'

²⁶A 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 [ÉTOT]_F student étot 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.

Carla Spellerberg

- 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 §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 §4, this optional step would have to be constrained by discourse-level constraints instead of being driven by a syntactic feature.

9 *Focus-sensitive particles in Bulgarian*

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

²⁹The 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.

³⁰I thank Željko Bošković for discussion of this point.

Carla Spellerberg

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 §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 (35b) and (35c).

(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

9 *Focus-sensitive particles in Bulgarian*

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 sǎsto 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 following (simplified) trees show how cases such as (39) and (40), respectively, could be represented.

³¹A 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.

³²A 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

Carla Spellerberg

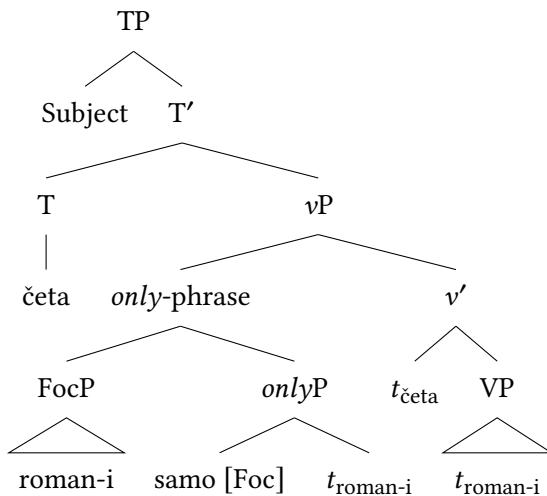


Figure 2: Low movement out of the *only*-phrase

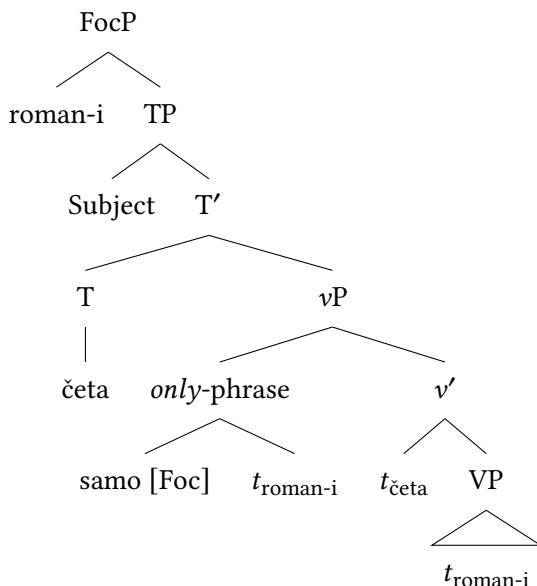


Figure 3: High movement out of the *only*-phrase

9 Focus-sensitive particles in Bulgarian

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.

³³A 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.

Carla Spellerberg

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

³⁴A 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.

9 Focus-sensitive particles in Bulgarian

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 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	REFL	reflexive
DEF	definite	SG	singular
FUT	future	Q	question
GEN	genitive		

Acknowledgments

I would like to express my thanks to my main consultants, Joana Emilova Djadjeva, Iverina Ivanova, Nelli Kerezova, and Evelina Parvanova, for their judgments and discussion of the data. Nothing of this would have been possible without them. I am also thankful to the audiences at Sinfonija 15, FDSL-15, and the syntax colloquium at Goethe University/Frankfurt for feedback and comments. The first draft of this paper was substantially improved was the comments made by two

Carla Spellerberg

reviewers, whom I would like to thank for their reviews. Additionally, I thank Rakesh 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](https://doi.org/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](https://doi.org/10.1023/A:1010653115493).
- Chomsky, Noam. 1986. *Barriers* (Linguistic Inquiry Monograph 13). Cambridge, MA: MIT Press.
- Erlewine, Michael Yoshitaka. 2014. *Movement out of focus*. Cambridge, MA: 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](https://doi.org/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 Reisch (eds.), *Particles in German, English, and beyond*, 323–354. Amsterdam: John Benjamins. DOI: [10.1075/slcs.224.12erl](https://doi.org/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](https://doi.org/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. *Working Papers in Linguistics* 12. 107–128.

9 Focus-sensitive particles in Bulgarian

- 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](https://doi.org/10.1007/s11049-008-9060-2).
- Mursell, Johannes. 2021. *The syntax of information-structural agreement* (Linguistik Aktuell/Linguistics Today 268). Amsterdam & Philadelphia: John Benjamins.
- Nicolova, Ruselina. 2000. Über den Kontrastfokus mit Partikeln im Bulgarischen. In Gerhard 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: [0.1007/s11049-004-0766-5](https://doi.org/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](https://doi.org/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](https://doi.org/10.3765/sp.11.6).
- Sudhoff, Stefan. 2010. *Focus particles in German: Syntax, prosody, and information structure* (Linguistik Aktuell/Linguistics Today 151). Amsterdam & Philadelphia: 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 morphology in honor of Catherine Rudin*, 287–312. Bloomington, IN: 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: Proceedings 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](https://doi.org/10.1111/synt.12202).
- Tomaszewicz, Barbara. 2013. *Až/čak*: the scalar opposite of scalar only. In Uwe Junghanns, Dorothee Fehrman, Denisa Lenertová & Hagen Pitsch (eds.), *Formal description of Slavic languages: The ninth conference. Proceedings of FDSDL 9, Göttingen 2011*, 301–323. Frankfurt am Main: Peter Lang.

Carla Spellerberg

- 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](https://doi.org/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, MI: Michigan Slavic Publications. DOI: [10.17863/CAM.101872](https://doi.org/10.17863/CAM.101872).

Chapter 10

The Western South Slavic verbal suffix *-nV/-ne* is a diminutive affix with a theme vowel

Ema Šarkl^a, Marko Simonović^b, Stefan Milosavljević^b &
Boban Arsenijević^b

^aUniversity of Nova Gorica, ZRC SAZU, University of Ljubljana ^bUniversity 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 \emptyset/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/-nⁱ* (with a floating vowel) and the theme vowel

Štarkl, Simonović, Milosavljević & Arsenijević

\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
blink-nV-INF
'blink' | b. | to(n)-nu-ti
sink-nV-INF
'sink' | (BCMS) |
| (2) | a. | mežik-ni-ti
blink-nV-INF
'blink' | b. | to(n)-ni-ti
sink-nV-INF
'sink' | (Slovenian) |

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-∅-ti	vez-e-mo	vez-∅-en	'embroider'
Slovenian	ves-∅-ti	vez-e-mo	vez-∅-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 §2, we critically assess previous accounts of $-nV/-ne$. §3 provides a quantitative description of this sequence in BCMS and Slovenian. In §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 §5. §6 concludes the paper.

10 *The Western South Slavic -nV/-ne is a diminutive affix with a theme vowel*

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, Nesson 2013, Sokolova 2015 for Russian, Nesson 2012 for Old-Church Slavonic, Bacz 2012 for Polish).

- | | | | | | |
|-----|----|---|----|--|-------------|
| (3) | a. | mah-nu-ti
wave-nV-INF
'wave once' | b. | mah-a-ti
wave-TV-INF
'wave repeatedly' | (BCMS) |
| (4) | a. | mah-ni-ti
wave-nV-INF
'wave once' | b. | mah-a-ti
wave-TV-INF
'wave repeatedly' | (Slovenian) |

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 §3.2).¹ NATURAL PERFECTIVES 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)

¹An 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 on treating these two *-nV/-ne* classes as containing the same suffix; see also Taraldsen Medová & Wiland (2019) (presented in §2.2.2) for a unified syntactic analysis of this suffix within a nanosyntactic framework.

Štarkl, Simonović, Milosavljević & Arsenijević

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; translation ours)
 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 §2.1, after which we provide a detailed discussion of previous formal analyses of *-nV/-ne* in §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, Vidović 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 (§2.2.1), and those arguing that *-nV/-ne* is decomposable into a suffix and a theme vowel – BIMORPHEMIC ANALYSES (§2.2.2).

10 The Western South Slavic *-nV/-ne* is a diminutive affix with a theme vowel

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.’

Štarkl, Simonović, Milosavljević & Arsenijević

- ‘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 (2020, 2023a,b), *-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.

10 The Western South Slavic *-nV/-ne* is a diminutive affix with a theme vowel

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 §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 & Sherkin-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 & Sherkin-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.,

Štarkl, Simonović, Milosavljević & Arsenijević

on nouns, the number markers are heads, 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/specifyed 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 vP-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 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

²On 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.

10 The Western South Slavic *-nV/-ne* is a diminutive affix with a theme vowel

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

Łazarczyk (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). Łazarczyk (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*.

Štarkl, Simonović, Milosavljević & Arsenijević

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 §3.1 and then present the quantitative data on the sequence *-nV/-ne* in §3.2. In §3.3 we summarize the discussion and findings so far to prepare the ground for our morpho-phonological (§4) and syntactic/semantic analysis (§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*, Arsenijević et al. 2022). 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), supplemented by additional data available at ([link to be provided at a later date](#)).

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.

10 *The Western South Slavic -nV/-ne is a diminutive affix with a theme vowel*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 §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 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

³Out 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.

⁴The 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.

⁵Out of 9 imperfective verbs in BCMS, 7 are degree achievements, and 2 are lexicalized states.

Štarkl, Simonović, Milosavljević & Arsenijević

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

Table 4: Imperfective counterparts of unprefixed perfective verbs

Simple PFV <i>-nV</i> 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 <i>-t-</i> counterpart (trep-nu-ti – trep-ta-ti ‘blink’)	11 (13.41%)	0 (0%)
an IPFV <i>-k-</i> counterpart (tres-nu-ti – tres-ka-ti ‘snap’)	24 (29.27%)	0 (0%)
IPFV SI counterpart, without preserving <i>-nV</i> (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 <i>-nV</i> (sva-nu-ti – sva-nj-ava-ti ‘dawn’)	4 (4.88%)	2 (9.52%)

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

Out of the 3 imperfective verbs in Slovenian, 1 is a degree achievement, and 2 are lexicalized states.

⁶The 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.’

10 *The Western South Slavic -nV/-ne is a diminutive affix with a theme vowel*

Table 5: Imperfective counterparts of prefixed perfective verbs

Prefixed PFV -nV verbs in <i>WeSoSlaV</i> 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%)

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 -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 §5. The other fact, i.e. the compatibility of -nV/-ne with secondary imperfectivizing suffixes in at least some

⁷The 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).

⁸In 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.

Štarkl, Simonović, Milosavljević & Arsenijević

verbs, corroborates our claim that the two suffixes are not in complementary distribution, contrary to much previous work (see §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 §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 §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)/*-nⁱ* (Slovenian), which selects the theme vowel *∅/e*.

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)/*-nⁱ* (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. 2022). The question remains how to treat the vowels *u* and *i* that appear next to the consonant

⁹See 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.

10 *The Western South Slavic -nV/-ne is a diminutive affix with a theme vowel*

-*n* when it is combined with the \emptyset -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 don't 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*- \emptyset -ti → maxnuti, *maxnti (BCMS)
max-*nⁱ*- \emptyset -ti → maxniti, *maxnti (Slovenian)
wave-nV-TV-INF
- b. max-*n^u*- \emptyset -l-a → maxnula, *maxnla (BCMS)
max-*nⁱ*- \emptyset -l-a → maxnila, *maxnla (Slovenian)
wave-nV-TV-PST-F
- c. max-*n^u*-*e*-mo → maxnemo, *maxnuemo (BCMS)
max-*nⁱ*-*e*-mo → maxnemo, *maxniemo (Slovenian)
wave-nV-TV-PRS.1PL
- d. max-*n^u*-*i*-mo → maxnimo, *maxnuimo (BCMS)
max-*nⁱ*-*i*-mo → maxnimo, *maxniimo (Slovenian)
wave-nV-TV.IMP-1.PL

An important argument for adding -*nV/-ne* verbs to the \emptyset/e class lies in the fact that the forms in (9) (as well as the rest of the paradigm) feature the endings typical of \emptyset/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 \emptyset/e class would become *u/e* in BCMS and *i/e* in Slovenian.

Štarkl, Simonović, Milosavljević & Arsenijević

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-Ø-ti* → *pasti*, **paduti*, **paditi* ‘fall.INF’ or *griz-Ø-ti* → *gristi*, **grizuti*, **griziti* ‘bite.INF’.

As mentioned above, adding *-nV/-ne* verbs to the \emptyset/e class does appear to face some potential empirical issues. In both languages, the passive participle of *-nV/-ne* verbs diverges from most \emptyset/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 \emptyset/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. ukrad-∅-l-a | ukrad-∅-en
 steal-TV-PST-F | steal-TV-PASS.PTCP

b. ugriz-∅-l-a | ugriz-∅-en
 bite-TV-PST-F | bite-TV-PASS.PTCP

(11) dir-nu-∅-l-a | dir-nu-∅-t, *dir-nu-∅-en
 touch-nV-TV-PST-F | touch-nV-TV-PASS.PTCP, touch-nV-TV-PASS.PTCP

As it turns out, the \emptyset/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.

¹⁰This allomorph shows up in several other environments in the classes \emptyset/e and a/a . As shown in Bešlin (2023), its conditioning is at least partially lexical.

10 The Western South Slavic *-nV/-ne* is a diminutive affix with a theme vowel

- (12) a. obu-∅-l-a | obuv-∅-en,
 put.shoes.on-TV-PST-F | put.shoes.on-TV-PASS.PTCP,
 ?obu-∅-t
 put.shoes.on-TV-PASS.PTCP
- b. natʃu-∅-l-a | natʃu-∅-t, natʃuv-∅-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 stress¹¹), as illustrated in (13) for the verbs *ukrasti* ‘steal’ and *gristi* ‘bite’. 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. u'krad-∅-l-a | u'krad-∅-en
 steal-TV-PST-F | steal-TV-PASS.PTCP
- b. 'griz-∅-l-a | 'griz-∅-en
 bite-TV-PST-F | bite-TV-PASS.PTCP
- (14) napix-nⁱ-∅-en → na'pixnjen, *napixnien
 inflate-nV-TV-PASS.PTCP

We suggest that the passive participle morpheme is actually *-jen*, with a floating *j*. This hypothesis is supported by the fact that in the *∅/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 *-nⁱ* morpheme and the passive participle ending *-jen* have floating segments (which in addition have the same features), there is

¹¹The 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.

Štarkl, Simonović, Milosavljević & Arsenijević

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. | pre'nes-∅-l-a
transfer-TV-PST-F | prene' ^j -∅-en
 transfer-TV-PASS.PTCP |
| | b. | pre'rast-∅-l-a
grow.overTV-PST-F | pre'raſt ^j -∅-en
 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 ∅/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)/*-nⁱ* (Slovenian), which we argue is a diminutive suffix. We start by showing the special status of *-nV* among suffixes in §5.1: its perfective nature, its possibility to participate in suffix stacking, and the theme vowel it combines with. In §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 modelling and formal semantic description are provided in §5.3 and §5.4, respectively. §5.5 brings a discussion on how the suffix *-nV* fits the broader picture of suffixes in Western South Slavic. Finally, in §5.6 we compare our analysis to the previous approaches to the suffix *-nV* and outline the advantages of our analysis.

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

¹²The 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 ∅/e class. This is expected on our account because in the former class both the original theme vowel (*i*) and the morpheme *(^j)en* favour palatalization.

10 The Western South Slavic *-nV/-ne* is a diminutive affix with a theme vowel

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)
J AUX begun COMP strike-SUFF-PRS.3.SG
'Jan began to strike.'
- b. Jan štrajk-uj-e dva sata.
J 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)
J AUX begun COMP shout-SUFF-PRS.3.SG
Intended: 'Jan began to shout.'
- b. * Jan vik-n-e dva sata.
J 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)
J strike-SUFF-SUFF-TV-PST.M
- b. * Ovas je stas-av-av-a-o. (BCMS)
oat grow-SUFF-SUFF-TV-PST.M
- c. Dan je sva-n^u-av-a-o [svajavao]. (BCMS)
day dawn-SUFF-SUFF-TV-PST.M
'The day was dawning.'

Štarkl, Simonović, Milosavljević & Arsenijević

- 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)
J 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 West South Slavic (WSS) 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 *Ø/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-Ø-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-Ø-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.

10 The Western South Slavic *-nV/-ne* is a diminutive affix with a theme vowel

- (21) a. i. lav
 lion
 ‘lion’

ii. lav-ić lav-č-e lav-č-ić
 lion-DIM lion-DIM-INFL lion-DIM-DIM
 ‘little lion’

b. i. smed-e
 brown
 ‘brown’

ii. smed-ast-o smed-(i)k-av-o
 brown-DIM.ADJ-INFL brown-DIM-ADJ-INFL
 smed-(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.

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 |
| | $\sqrt{\text{SINK}}\text{-nV-INF}$ | $\sqrt{\text{ROT}}\text{-nV-INF}$ | $\sqrt{\text{WORRY}}\text{-nV-INF}$ | $\sqrt{\text{DAWN}}\text{-nV-INF}$ |
| | 'sink' | 'rot' | 'worry' | 'dawn' |

Štarkl, Simonović, Milosavljević & Arsenijević

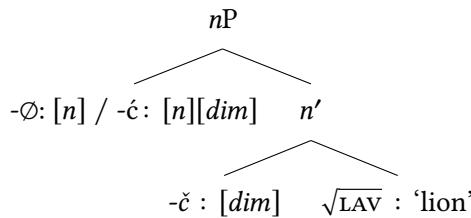


Figure 1: Syntactic representation of (double) diminutive nouns

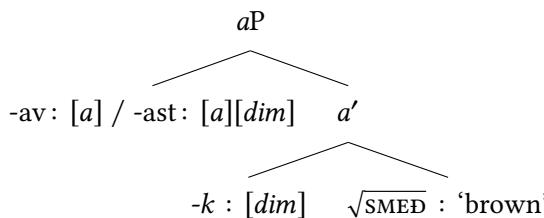


Figure 2: Syntactic representation of (double) diminutive adjectives

- | | | | | |
|------|---------------------------------|-------------------------------|------------------------------|-------------------------------|
| (23) | greb-nu-ti | spav-nu-ti | skok-nu-ti | kuc-nu-ti |
| | $\sqrt{\text{SCRATCH}}$ -nV-INF | $\sqrt{\text{SLEEP}}$ -nV-INF | $\sqrt{\text{JUMP}}$ -nV-INF | $\sqrt{\text{KNOCK}}$ -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 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).

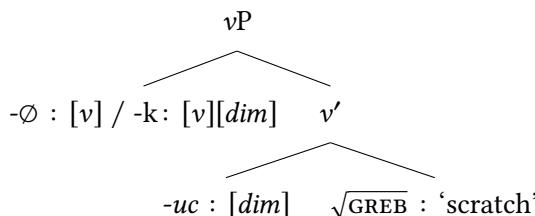


Figure 3: Syntactic representation of (double) diminutive verbs

10 *The Western South Slavic -nV/-ne is a diminutive affix with a theme vowel*

(24)	greb-uc-nu-ti ✓SCRATCH-DIM-nV-INF 'scratch a little'	spav-uc-nu-ti ✓SLEEP-DIM-nV-INF 'sleep a little'	(BCMS)
	prd-uc-nu-ti ✓FART-DIM-nV-INF 'fart a little'	??kuc-uc-nu-ti ✓KNOCK-DIM-nV-INF '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 ✓STEP-TV-INF	stop-ic-a-ti ✓STEP-DIM-TV-INF	stop-ic-ni-∅-ti ✓STEP-DIM-nV-TV-INF	(Slovenian)
			'make a step' 'make little steps/ make steps a little' 'make one little step'	

However, verb diminution is common in child-oriented 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 exponented by a theme vowel.

- | | | | | | |
|------|---|--------------------------------|--|------|-------------|
| (26) | a. čič-a-ti
✓SIT-TV-INF | čič-k-a-ti
✓SIT-DIM-TV-INF | čič-ni-∅-ti
✓SIT-nV-TV-INF REFL | (se) | (Slovenian) |
| | | | 'sit' 'sit in a small way' 'sit down' | | |
| | b. čič-k-ni-∅-ti
✓SIT-DIM-nV-TV-INF REFL | (se) | | | |
| | | | 'sit in a small way' | | |
| | c. * čič-n(i)-k-a-ti
✓SIT-nV-DIM-TV-INF REFL | (se) | | | |
| (27) | a. cap-a-ti
✓DRIP-TV-INF | cap-k-a-ti
✓DRIP-DIM-TV-INF | cap-lj-a-ti
✓DRIP-DIM-TV-INF | (se) | (Slovenian) |
| | | | 'take steps' 'take little steps/step a little' 'take little steps/step a little' | | |

Štarkl, Simonović, Milosavljević & Arsenijević

- 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. hopa-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
 $\sqrt{\text{HOP-nV-DIM-TV-INF}}$ $\sqrt{\text{HOP-nV-DIM-TV-INF}}$

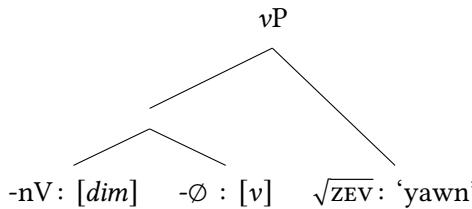
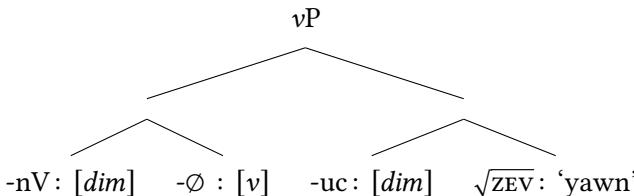
We take the similarity of the position in the words between the diminutive suffix *-lj* and the suffix *-nV* (i.e. the fact that they both can precede another verbal suffix or follow another diminutive suffix) and their complementary distribution in Slovenian as additional evidence for *-nV* combining 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)
 $\sqrt{\text{YAWN}}\text{-nV-TV-INF}$ $\sqrt{\text{YAWN}}\text{-DIM-nV-TV-INF}$
 ‘yawn a little’ ‘yawn a little’

Subsequent head movement derives the surface order.

10 The Western South Slavic *-nV/-ne* is a diminutive affix with a theme vowelFigure 4: Syntactic representation of the verb *zevnuti* in (29)Figure 5: Syntactic representation of the verb *zevucnuti* in (29)

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

Štarkl, Simonović, Milosavljević & Arsenijević

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 $M(x)$ 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. In 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* '√run-TV-INF' gets the interpretation of a small (i.e. atomic) instance of running.¹⁴

- (32) a. $[dim] := \lambda x[M(x) < \text{STD}]$

¹³Here 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 assume with Milosavljević (2023b) 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.

¹⁴An 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 WSS 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 and *tresk-ni-ti*

10 The Western South Slavic *-nV/-ne* is a diminutive affix with a theme vowel

- b. $\sqrt{\text{TRK}} := \lambda x[\llbracket \sqrt{\text{TRK}} \rrbracket(x)]$
by predicate modification:
 $\llbracket [\text{dim}] \sqrt{\text{TRK}} \rrbracket := \lambda x[\llbracket \sqrt{\text{TRK}} \rrbracket(x) \wedge M(x) < \text{STD}]$
- c. $[v] := \lambda x[\text{EVENT}(x) \wedge \text{NEAT}(x)]$
by predicate modification:
 $\llbracket [\text{dim}][v] \rrbracket := \lambda x[\text{EVENT}(x) \wedge \text{NEAT}(x) \wedge M(x) < \text{STD}]$
- d. by predicate modification with the root:
 $\llbracket [[\text{dim}][v]] \sqrt{\text{TRK}} \rrbracket := \lambda x[\text{EVENT}(x) \wedge \text{NEAT}(x) \wedge M(x) < \text{STD} \wedge \llbracket \sqrt{\text{TRK}} \rrbracket(x)]$

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)
	$\sqrt{\text{LIGHT}}$ -DIM-TV-INF	$\sqrt{\text{WHITE}}$ -DIM-TV-INF	
	'emit light a little'	'be white a little'	
	svir-uc-a-ti	šet-uc-a-ti	
	$\sqrt{\text{PLAY}}$ -DIM-TV-INF	$\sqrt{\text{WALK}}$ -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 WSS (and possibly more generally Slavic). 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

in Slovenian, with the respective word *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 counter-argument 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).

Štarkl, Simonović, Milosavljević & Arsenijević

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 biaspectral) verbal suffixes, traditionally associated with some aspectual projection. The last type of suffixes has been analyzed in Simonović et al. (2023), 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. Together with Svenonius (2004) and Biskup (2020, 2023a,b), it relates the suffix with the verbal category. With Kwapiszewski (2020), our analysis attributes the suffix specification of properties of quantity (the unit of counting), and with Arsenijević (2006), it associates it with diminutivity. Finally, with Armoškaitė & Sherina-Lieber (2008), we associate the suffix with the unit of counting, and with Łazorczyk (2010), Taraldsen Medová & Wiland (2019), 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. PERFECTIVE DELIMITATIVES 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).

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

¹⁵Due 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.

10 The Western South Slavic *-nV/-ne* is a diminutive affix with a theme vowel

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 born 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 \emptyset/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 on our approach since the suffix does not target the AspP, but a lower head (i.e. on 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 (\emptyset/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.

Štarkl, Simonović, Milosavljević & Arsenijević

Abbreviations

1	first person	NOM	nominative
3	third person	PASS	passive
ACC	accusative	PL	plural
AUX	auxiliary	PREF	prefix
ADJ	adjective	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

This research was funded by the Austrian Science Fund FWF (grants I 4215 and I 6258), the Austrian Federal Ministry of Education, Science and Research (BMBWF); Scientific & Technological Cooperation grant for the project *Verbal suffixes and theme vowels: Different or the same?*), as well as the Slovenian Research And Innovation Agency (ARIS; grant P6-0382).

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, MI: 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, IL: 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: Jon Benjamins. DOI: [10.1075/lfab.17.04ars](https://doi.org/10.1075/lfab.17.04ars).

10 The Western South Slavic -nV/-ne is a diminutive affix with a theme vowel

- 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](https://doi.org/10.5281/zenodo.10123627).
- Arsenijević, Boban, Katarina Gomboc Čeh, Franc Lanko Marušić, Stefan Milosavljević, Petra Mišmaš, Jelena Simić, Marko Simonović & Rok Žaucer. 2022. *WeSoLaV: Database of the Western South Slavic verbal system*. Database. In preparation. <https://hyper-verb.ung.si/wiki/doku.php>.
- 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](https://doi.org/10.1162/LING_a_00095).
- Arsenijević, Boban, Stefan Milosavljević & Marko Simonović. 2023. Secondary imperfectivisation is reverbalisation is addition of theme vowels. Presented at 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.
- 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](https://doi.org/10.3765/plsa.v8i1.5499).
- Biskup, Petr. 2019. *Prepositions, case and verbal prefixes: The case of Slavic* (Linguistik Aktuell/Linguistics Today 255). Amsterdam & Philadelphia: John Benjamins Publishing Company. DOI: [10.1075/la.255](https://doi.org/10.1075/la.255).
- Biskup, Petr. 2020. An Agree analysis of the morphological aspect in Slavic. Manuscript, University of Leipzig. <https://ling.auf.net/lingbuzz/006032>.
- 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](https://doi.org/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, Frankfurt am Main: Peter Lang.
- 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](https://doi.org/10.1007/1-4020-3033-9_3).

Štarkl, Simonović, Milosavljević & Arsenijević

- Borer, Hagit. 2005b. *The normal course of events: Structuring sense vol. II*. Oxford: Oxford University Press. DOI: [10.1093/acprof:oso/9780199263929.001.0001](https://doi.org/10.1093/acprof:oso/9780199263929.001.0001).
- 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, MA: 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1023/A:1009751908064).
- Farris-Timble, 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](https://doi.org/10.1075/sl.31.3.04jan).

10 The Western South Slavic -nV/-ne is a diminutive affix with a theme vowel

- 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, MI: 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, CA: CSLI Publications.
- Klimek-Jankowska, Dorota, Anna Czypionka, Wojciech Witkowski & Joanna Bła- szczak. 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.4148/biyclc.v6i0.1579).
- Łazorczyk, Agnieszka Agata. 2010. *Decomposing Slavic aspect: The role of aspectual morphology in Polish and other Slavic languages*. Los Angeles, CA: 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, Sweden: Association for Computational Linguistics. DOI: [10.3115/v1/W14-0405](https://doi.org/10.3115/v1/W14-0405).
- Logar-Berginc, Nataša, Simon Krek, Tomaž Erjavec, Miha Grčar, Peter Halozan & Simon Šuster. 2012. *Gigafida. corpus*. <http://www.gigafida.net>.
- 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](https://doi.org/10.1080/00806760903175417).

Štarkl, Simonović, Milosavljević & Arsenijević

- Markman, Vita G. 2008. On Slavic semelfactives and secondary imperfectives: Implications for the split ‘AsP’. In *Proceedings of the 31st Annual Penn Linguistics Colloquium* (University of Pennsylvania Working Papers in Linguistics 14(1)), 255–268. Philadelphia, PA: Penn Graduate Linguistics Society. <https://repository.upenn.edu/handle/20.500.14332/44670>.
- Mathieu, Éric. 2012. Flavors of division. *Linguistic Inquiry* 43(4). 650–679. DOI: [10.1162/ling_a_00110](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.16995/glossa.8535).
- Milosavljević, Stefan, Petra Mišmaš, Marko Simonović, Boban Arsenijević, Kata-Rina 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 *nq*-verbs and linguistic profiling. *Russian Linguistics* 36(3). 285–303. DOI: [10.1007/s11185-012-9093-3](https://doi.org/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>.

10 The Western South Slavic -nV/-ne is a diminutive affix with a theme vowel

- 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](https://doi.org/10.1007/s10828-010-9040-x).
- Pietroski, Paul M. 2005. *Events and semantic architecture*. Oxford: Oxford University Press.
- Progrovac, Ljiljana. 2005. *A syntax of Serbian: Clausal architecture*. Bloomington, IN: Slavica Publishers.
- Rijkhoff, Jan N.M. 1991. Nominal aspect. *Journal of Semantics* 8(4). 291–309. DOI: [10.1093/jos/8.4.291](https://doi.org/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 / Philadelphia: John Benjamins Publishing Company. DOI: [10.1075/la.110.04rot](https://doi.org/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 / New York: Walter de Gruyter. DOI: [10.1515/9783110925449.175](https://doi.org/10.1515/9783110925449.175).
- Sauerland, Uli. 2003. A new semantics for number. In Robert B. Young & Yiping Zhou (eds.), *SALT 13: Proceedings of the 13th Semantics and Linguistic Theory Conference*, 258–275. Ithaca, NY: CLC Publications. DOI: [10.3765/salt.v13i0.2898](https://doi.org/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 semanticheskaja zona: Tipologija sistem i scenarii diakhronicheskogo*

Štarkl, Simonović, Milosavljević & Arsenijević

- razvitija. *Sbornik Statej, V. mezhunarodnoj konferencii komissii po aspektologii mezhunarodnogo 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](https://doi.org/10.7557/12.213).
- 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](https://doi.org/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](https://doi.org/10.1007/s11049-018-9434-z).
- Toporišić, Jože. 2000. *Slovenska slovnica*. Maribor: Založba Obzorja.
- Townsend, Charles E. 1968. *Russian word-formation*. McGraw-Hill.
- Vidović 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](https://doi.org/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: Studies in honor of Henk van Riemsdijk*, 669–678. Berlin / New York: Mouton de Gruyter. DOI: [10.1515/9783110892994.669](https://doi.org/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](https://doi.org/10.1556/2062.2017.64.4.4).

Chapter 11

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. “non-clitic”. 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

Zanon

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 licensor with a (better studied) *nibud*-indefinite. Neither is possible in past episodic declaratives like (3).

- | | | |
|-----|---|--|
| (3) | * Včera kto-nibud' kto umer. | |
| | yesterday who- <i>nibud</i> ' 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 non-veridical 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.

¹Data 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 the 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.

11 *Wh-indefinites in Russian*

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 & Fălăuș 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)

²But see Šimík (2009) for arguments that wh-elements in MECs are (Hamblin) indefinites, after all.

³Belyaev & 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.

Zanon

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 eto 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 (§2.1). But there are multiple ways to bypass this requirement: by introducing an extrapropositional (epistemic) speech act adverb (§2.2), by integrating an explicit scale whose value is set to be “less than” the alternatives (§2.3), or by embedding the indefinite in the context of “high” negation (§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 §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.

Zanon

- (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 licensor must be merged in a position presumably related to the (split) C-domain, which houses interrogative, conditional and epistemic operators. But

11 *Wh-indefinites in Russian*

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) s delaet 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 poučavstvuet.
 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

Zanon

bias, may salvage the configuration (as demonstrated by the contrast between (13) and (14b)).

- (15) [CP *O_P_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 licensors 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 §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

⁴Modal verbs are frequently ambiguous between the two (see, e.g., Papafragou 2006 and references therein).

11 *Wh-indefinites in Russian*

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 (*vrijad li*) to neutral (*možet (byt')*) to strong (*očevidno, dolžno byt'*). As elements of epistemic/evidential/inferential flavor, they form a natural class.

- (20) a. (?) {Vrijad 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.'

Zanon

- 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 byť / očevídnō}, obidel ee kto (raz pláčet).
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 §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.'

⁵As an aside, *nibud'*-indefinites are perfect in (22) (though somewhat awkward in (23)).

11 *Wh-indefinites in Russian*

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 licensors), 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 pokože}, 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’

Zanon

- 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 licensors 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)

11 *Wh-indefinites in Russian*

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'*.

⁶The 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'*.

Zanon

- (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, ...’
- ✓/? Nadejus’, čto [tam] kogo vstretit.
hope that there who.INDF meets
‘I hope he meets somebody there (at least).’
 - ✓/? 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’ (Giannakidou 2007). Giannakidou argues that *esto* carries a negative existential presupposition 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 licensing environments, wh-indefinites are possible only in the presence of a scalar element which supplies a (contextual) bottom-of-the-scale condition. If so, iterative contexts with frequency adverbs such as (30) likewise comply with this “less than” requirement: wh-indefinites are only possible with negative frequency adverbs 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 licensing 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 behind that.SBJV who.INDF who-nibud’ ni.who NEG sbežal dorogoju, exali na konjax dva monaxa.
ran.away 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 molodaia*. 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.

Zanon

- (34) (Smotri / Bojus',) kak by {kto / ⟨✓kto-nibud'⟩ / ⟨✓nikto⟩}
 watch.out.IMP / fear how SBJV who.INDF who-nibud' ni.who
 telefon ne stibril v takoj tolpe!
 phone NEG 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, LEST-clauses (31) and apprehensive subjunctives (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), alleged to be devoid of polarity reversing semantics despite the compulsory realization 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 evaluative 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) [… [NegP1 NEG … [TP … [NegP2 NEG [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 classified 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.NEG). Neither construction tolerates NCIs.

⁷Contrary 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.

11 Wh-indefinites in Russian

- (36) Uvjažneš po samye stupitsy i zagoraеš, poka {kто /
 stuck to very hubs and tan until who.INDF
 ⟨✓kто-nibud’⟩ / ⟨ *nikto⟩} ne vytaščit.
 ‘who-nibud’ ni.who NEG will.pull.out
 ‘Your hubs get stuck and you hang out until somebody pulls you out.’
 (O. Efremov. *Rybak primor’ja*. 2003)
- (37) Xuj {кто / kto-nibud’ / ⟨ *nikto⟩} prišel.
 X.NEG who.INDF 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
LEST-clauses	✓	✓
apprehensive subjunctives	✓	✓
<i>until</i> -clauses	✗	✓
X-NEG	✗	✓

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

Zanon

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 [Act⁰ •][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 vytáčit.
 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).

⁸I refer the reader to Padučeva (2007), Pereltsvaig (2008) for discussion of Russian *nibud'*-indefinites in quantified contexts.

Zanon

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 ékzamen?
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].'

This locality constraint is intuitively logical. While both local and distant licensors require full morphological specification, the medial one enables the spellout 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 licensor. 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

⁹Reviewer 2 points out that the first clause of this biconditional is falsified by Czech (see Šimkovič 2010) and perhaps Slovenian (Mišmaš 2017).

11 *Wh-indefinites in Russian*

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 nakravyval doždiček.
over Krakow drizzled rain
'The rain was drizzling over Krakow.'
 - b. Nad Krakovom DOŽDIČEK nakravyval.
over Krakow rain drizzled
- (45) a. Nad Krakovom | nakravyval | doždiček.
LH | LH | HL (\`)
 - b. Nad Krakovom | DOŽDIČEK nakravyval.
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.

¹⁰This representation is borrowed from King (1993).

Zanon

- (46) Pojdeme kuda-nibud'. (Type 1 with indefinites)
 lets.go where-*nibud'* (↖)
 '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 (↖)
 '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'* (↖)
 'Somebody must've died.'
 b. Dolžno byt', | kto-nibud' umer.
 must be who-*nibud'* died (↖)

11 *Wh-indefinites in Russian*

Finally, consider (50). The sentence is parsed into two π Ps. 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 π P-initial position to the left of the element realizing default declarative prosody.

- (50) a. Dolžno byt', |prixodil kto.
 HL (↖↓)
 b. ?*Dolžno byt', |kto prixodil.
 HL (↖↓)
 c. *Dolžno byt', |KTO 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 (↖↓)
 b. Možet, kto umer.
 maybe who.INDF died (↖↓)
- (52) ?*/??Možet byt', |kto prišel.
 may be who.INDF came (↖↓)
 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?
 *

Zanon

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.'

11 *Wh-indefinites in Russian*

- 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.

Zanon

- (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, *fu* ‘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. * fu w min bta-ʃref 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 ʃam t-rooh-o ʃa-l-masbah?
 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 dependent 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 quantified elements in Russian, the baseline assumption would situate wh-indefinites

¹¹See 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 §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

¹²Reviewer 1 points out that this property renders them rather akin to weak pronouns in the sense of [Cardinaletti & Starke \(1999\)](#)

Zanon

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
INDEF	indefinite	Q	question marker
IPFV	imperfective	REL	relative
M	masculine	SBJV	subjunctive
NEG	negation	SG	singular

Acknowledgments

I am grateful to the audience of FDSL 15 as well as the two reviewers for helpful comments and suggestions.

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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1002/9781118358733.wbsyncom119).

11 *Wh-indefinites in Russian*

- 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](https://doi.org/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](https://doi.org/10.1016/S0024-3841(02)00146-8).
- Caponigro, Ivano & Anamaria Fălăuș. 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1007/s11049-009-9069-1).

Zanon

- 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](https://doi.org/10.1075/cilt.91.07far).
- Fitzgibbons, Natalia. 2010. *Licensers and meanings: Structural properties of dependent indefinites*. Storrs, CT: 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](https://doi.org/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](https://doi.org/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 Veltman*, 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 of 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](https://doi.org/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](https://doi.org/10.3765/salt.v18i0.2510).
- Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.

11 *Wh-indefinites in Russian*

- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1002/9781118584552.ch8).
- Lin, Jo-Wang. 2020. Correlatives. *The Wiley Blackwell Companion to Semantics*. 1–25. DOI: [10.1002/9781118788516.sem062](https://doi.org/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](https://doi.org/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>.

Zanon

- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1016/j.lingua.2005.05.009).
- Paperno, Denis. 2012. *Semantics and syntax of non-standard coordination*. Los Angeles, CA: 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, MA: 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](https://doi.org/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, MI: 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, MI: 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, MI: 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](https://doi.org/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](https://doi.org/10.1162/ling_a_00345).

11 *Wh-indefinites in Russian*

- 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.3765/salt.v15i0.2921).
- Yokoyama, Olga T. 1987. *Discourse and word order*. Amsterdam/Philadelphia: John Benjamins.
- Yoon, Suwon. 2011. “*Not*” in the Mood: The syntax, semantics, and pragmatics of Evaluative Negation. Chicago, IL: The University of Chicago. (Doctoral dissertation).

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 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 the linguistics more broadly.