Chapter 1

Nanosyntax: state of the art and recent developments

Pavel Caha, Karen De Clercq, Michal Starke and Guido Vanden Wyngaerd

Nanosyntax: state of the art and recent developments

Pavel Caha (Masaryk University), Karen De Clercq (CNRS/LLF/Université Paris Cité), Michal Starke (Masaryk University), and Guido Vanden Wyngaerd (KU Leuven)

1. Introduction

One of the goals of current linguistic theory is to reduce grammatical variation to the properties of lexical items. This desideratum is known in the literature as the Borer-Chomsky conjecture (Borer 1984, Baker 2008: 353). The main idea behind the conjecture is that the operations of syntax are invariant across languages, variation arising solely from the interaction of those principles with language-specific lexical entries. This locates variation to a domain where the evidence for variation is indisputable -- the lexicon.

Nanosyntax shares this desideratum with current syntactic theories. However, the implementation of this idea in Nanosyntax differs from standard theories in one crucial respect. In Nanosyntax, the terminals of syntax are not language-specific feature bundles (i.e., pre-established assemblies of features), but individual features. Those features are furthermore universal. There is thus no language-particular lexicon preceding the syntactic derivation. As a result, variation is restricted to the postsyntactic lexicon.

How can postsyntactic lexical entries trigger variation in the syntactic computation which precedes it? Starke (2014b; 2018) suggests that lexicalisation applies after every application of merge, and if it fails, the structure is handed back to syntax for last resort rescue operations — movements. This resolves the timing tension of how postsyntactic entries influence what happens in syntax.

This mechanism proposed by Starke is referred to as the *Lexicalisation Algorithm*. The Lexicalisation Algorithm regulates the process of externalization of syntactic structures. It takes the universal syntactic hierarchy or *functional sequence (fseq)* as an input and produces language-particular outputs due to its interaction with two other components of the grammar: the postsyntactic (language-particular) lexicon on the one hand and the (language-invariant) syntactic principles on the other hand. In effect, the algorithm steers the derivation in such a way that language-particular structures are formed as a function of language-particular (post-syntactic) lexicons.

Work since 2018 has been marked by two important evolutions. The first of these is the realization that lexicalisation is not only affected by the size of lexical entries, but also by their structural shape. Two lexical entries of the same size may give rise to different patterns of lexicalisation by organising the same projections into different structures. The second innovation is more fundamental, in that it involves a change to the formulation of the Lexicalisation Algorithm itself. It is therefore natural for this volume to take the Lexicalisation Algorithm as its primary focus, and to explore its empirical effects across a wide range of phenomena. By assembling many different case studies under one roof, the volume showcases the wide applicability and the explanatory power of the algorithm.

In this chapter, we discuss the basics of phrasal lexicalisation in section 2, and relate it to Cinque's (2005) account of crosslinguistic differences in ordering. Section 3 explains how root size explains allomorph distribution. Section 4 introduces Starke's (2018) Lexicalisation Algorithm, and illustrates its operation using a concrete example. Section 5 discusses the shape of lexical items, in preparation of section Section 6, which discusses the recent update of the algorithm involving subextraction.

As a final note in this section, we want to point out that this chapter does not aim at making a detailed comparison between the Lexicalisation Algorithm of Nanosyntax and related late-insertion models (like spanning approaches, Distributed Morphology and the Exoskeletal model). However, Taraldsen's chapter in the current volume compares Nanosyntax to Distributed Morphology in his analysis of suppletive verbs in Romance. We also refer the reader to Caha (2018, to appear), and Barbiers, van der Wal and Vanden Wyngaerd (to appear), for such comparisons.

2. The basics of phrasal lexicalisation

Since its beginnings, one of the tenets of Nanosyntax has been to provide a theory of the syntax-phonology interface that is direct in the sense that the output of syntax is directly translated onto the phonological representation, without any intermediate morphological structure. A crucial ingredient that allowed Nanosyntax to progress towards this goal was the concept of phrasal lexicalisation, explored by Starke since the early 2000s (Starke 2002; see also McCawley 1968, Weerman and Evers-Vermeul 2002, Neeleman and Szendroi 2007). This notion refers to a scenario where phonological interpretation is assigned to a phrasal node containing one or more terminals.

While exploring this notion, it is important to keep in mind that the expressive power of phrasal lexicalisation is inextricably linked to the constituent structure provided by syntax. We therefore inevitably introduce the notion of phrasal lexicalisation along with the specific Nanosyntactic assumptions about syntactic structures. These draw upon the work in the cartographic tradition (Cinque & Rizzi 2010a), which assumes the existence of highly structured syntactic hierarchies (the functional sequence, *fseq*), such as they were postulated in the work on the decomposition of NP (Abney 1987), CP (Rizzi 1997, 2004), IP (Cinque 1999), PP (Koopman 2000, Svenonius 2010), NegP (Zanuttini 1991, 1997; Poletto 2008), TopicP (Benincà and Poletto 2004; Frascarelli 2007; Frascarelli and Hinterhölzl 2007), as well as other domains. A final important ingredient is the theory of linear orders proposed in Cinque (2005). Let us illustrate all three components (the theory of phrasal lexicalisation, the functional sequence, and the theory of ordering) by the example of Greenberg's Universal 39, given in (1).

(1) Greenberg's Universal 39 (Greenberg 1961: 95)

Where morphemes of both number and case are present and both follow or both precede the noun base, the expression of number almost always comes between the noun base and the expression of case.

This generalisation can be illustrated by languages such as Turkish in (2) or Tagalog in (3), where the number marker (in bold) intervenes between the root and the case, no matter whether we are looking at suffixes (as in (2)) or prefixes (as in (3)).

(2) Turkish

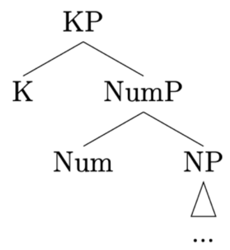
a.	el	-de		b.	el	**-ler**	-de
	hand	-LOC			hand	-PL	-LOC
	'in hand'				'in hands'		

(3) Tagalog

a.	sa	=bata		b.	sa	=**mga**	=bata
	DAT	=child			DAT	=PL	=child
	'to the child'				'to the children'		

The fact that languages tend to place number markers in between the noun and the case marker can be captured if we assume a crosslinguistically invariant hierarchy of functional projections (the functional sequence, *fseq*), where number is lower than case, and therefore, by default, closer to the noun, see (4).

(4)



Assuming a universal hierarchy like this disallows the generation of orders such as Num>K>N (with case in between number and the noun), because such an order can only arise from a hierarchy where Num is merged higher than K, violating the universal hierarchy of merge order. This ingredient of the cartographic approach is fully embraced in Nanosyntax.

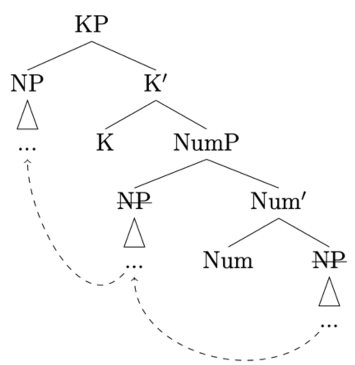
Consider now the fact that Greenberg (1961) states the generalization in (1) as a tendency, which means that there are counterexamples. As a case in point, consider the Classical Armenian example in (5).

(5)	a.	azg	-aw	b.	azg	-aw	-k’
		nation	-INS		nation	-INS	-PL
		‘with a nation’			‘with nations’		

Relevantly, as reported in a typological study by Kloudová (2020), examples like (5) only appear in a position after the noun, and never in a position before the noun (i.e., the order Num>K>N is not found). This observation falls in line with the approach pioneered by Cinque (2005) for the ordering of nouns, adjectives demonstratives and numerals (Greenberg’s Universal 20). In this domain, we also observe one order before the noun, and various different orders after the noun.

To account for this, Cinque (2005, 2009) proposed that the syntactic hierarchy is always mapped on a left-to-right linear order based on c-command, which corresponds to the invariant order before the noun. For case and number morphology, the underlying structure in (4) thus maps onto the order K>Num>N. Any other order is derived by a leftward movement of a phrasal constituent containing the noun. In Cinque's approach, Classical Armenian thus has the structure as shown in (6).[1]

(6)



In (6), instead of staying in the base position, the noun contained in the NP moves cyclically across the projections of number and case without inverting their order, yielding the surface order N>K>Num. The first movement that brings the noun across Num will be referred to as complement movement because the whole complement moves across Num. The second movement will be referred to as a 'spec-to-spec' movement because only the complex left branch inside the complement of K moves across it. As we shall shortly see, these two movement types (complement movement and spec-to-spec movement) form the backbone of the Lexicalisation Algorithm proposed by Starke (2018).

We must, however, mention at this point that while Nanosyntax adopts the geometric properties of the trees proposed in Cinque's approach, there are also some differences that become important later. Specifically, in Cinque (2005), specifiers are dominated by a label

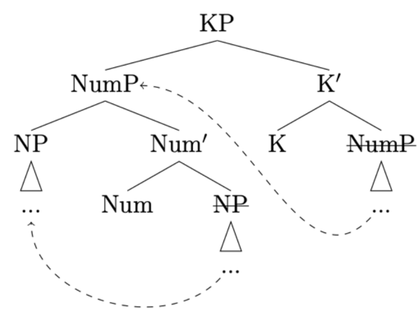
---

[1] In Cinque (2005), the NP moves to the Spec of intermediate Agr projections. We ignore this here for simplicity.

related to their sister, their movement is triggered by attracting features, and they enter in agree(ment) relationships. In Nanosyntax, phrases that sit in the same geometrical position lack these properties. They move because of 'lexicalisation-driven movement', which is a movement type distinct from feature-driven movements (to be introduced in Section 4). Furthermore, their mother node is not labelled. For now, this is merely a notational difference, but the unlabelled character of the nodes created by movement will become relevant in Section 6.

Consider now the Turkish order N>Num>K. This order would, under Cinque's (2005) approach, be derived by the sequence of two complement movements, whereby we first move the NP across Num, and then again move the whole complement across K, as in (7).

(7)



In sum, to derive the attested orders, Cinque assumes that for each contentful head that is merged into the tree, there are three different derivational options. First, it can be the case that the complement of the head stays in situ and there is no movement. Second, the specifier of the complement can move, as when the NP moves across K in Classical Armenian. Third, the whole complement of the head can move, as when the NumP moves across K in Turkish. Nanosyntax adopts this general view on the derivation, even though it changes (as we shall discuss) some of the properties of the trees, e.g., the trigger for the movement and the labelling conventions.

With the basic assumptions concerning the functional sequence and the movements clarified, let us turn to phrasal lexicalisation. The role phrasal lexicalisation plays in Nanosyntax can be illustrated by means of a comparison between Latin and Turkish in Table 1.

	TURKISH 'HOUSE'		LATIN 'GRANDFATHER'	
	SG	PL	SG	PL
NOM	ev	ev-ler	av-us	av-i
ACC	ev-i	ev-ler-i	av-um	av-os
GEN	ev-in	ev-ler-in	av-i	av-orum

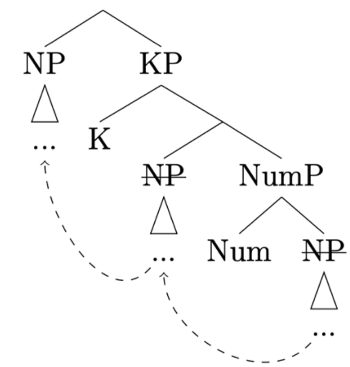
Table 1 Latin and Turkish number and Case morphemes

In Turkish, the morphology is agglutinative: the morpheme *-lAr* (with a harmonizing vowel) realizes PL, while *-i* realises accusative and *-In* genitive. In contrast, the Latin morphemes are fusional, i.e., they realise more than one grammatical feature. There is no way of segmenting the Latin morphemes into invariant plural or case markers.

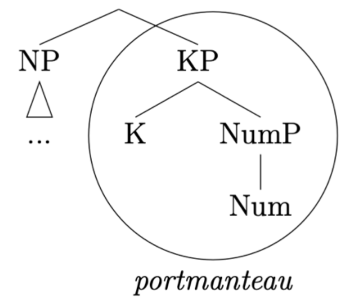
In theories where lexicalisation only targets terminal nodes, the Latin facts present a paradox. The paradox arises because to capture Greenberg's Universal 39, we must maintain that number and gender are independent projections and that their hierarchy does not vary across languages, being a part of UG. At the same time, Latin obviously uses a single morpheme to lexicalise both projections. Features associated to a single morpheme are therefore spread across several terminals, and hence a terminal-based approach cannot match the syntactic structure with the morpheme.

In Nanosyntax, the information spread across several terminals will correspond to a phrase and phrasal lexicalisation will target that constituent – no paradox. Specifically, the idea is that the noun in Latin moves cyclically to the top of the tree as shown in (8)(a). Note that since we now switch to a Nanosyntax account, we also change the labelling conventions, with the nodes dominating phrases which are similar to the traditional specifiers unlabelled. After the movement (ignoring the trace of the moved noun), we obtain a structure where K and Num form a constituent, see (8b). This constituent is lexicalised by the portmanteau morpheme.

(8) a.



b.



Phrasal lexicalisation extends the notion of constituency to a new domain: lexicalisation. When multiple terminals jointly move to a particular position (e.g., to the left of the verb in V2 languages), it is assumed that a whole constituent containing all the relevant terminals moves as one unit, i.e., as a constituent. When multiple terminals are not pronounced due to ellipsis, it is assumed that the ellipsis targets the full phrase containing these terminals (e.g., VP ellipsis). The idea in (8)(b) extends this rationale to the pronunciation of multiple terminals: when multiple terminals are realised by a single morpheme, this is because the morpheme lexicalises a constituent containing these terminals.

Summarizing, the discussion above shows that phrasal lexicalisation in combination with the types of movements utilized in Cinque (2005) allows us to capture the differences between the number of terminals and the number of morphemes. It does so without altering the underlying syntactic structure (number and case remain separate terminals) and therefore allows one to maintain the idea of a universal functional sequence. It also does not require any postsyntactic operation (like Fusion), thereby also avoiding the need for a dedicated morphological structure, distinct from syntax.

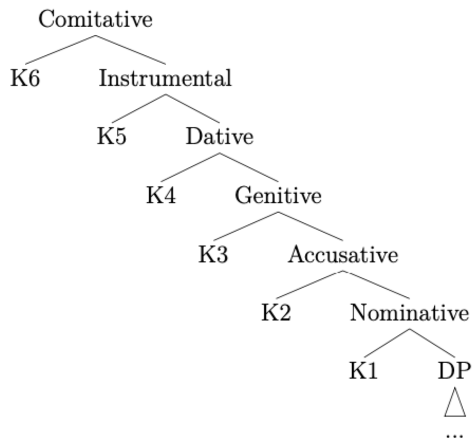
Nanosyntax generalizes this logic beyond obvious examples of fusional and portmanteau morphology, such as the Latin genitive plural -*orum* above. To see this, consider agglutinative case markers, such as the Turkish genitive -*In* or the Armenian instrumental -*aw*: they also lexicalise with a single morpheme what other languages lexicalise with more than one morpheme. In Caha's (2010) containment data (Table 2) for instance, the genitive, which is lexicalised as a single morpheme in Turkish (as we saw above) is lexicalised by two suffixes in Vlakh Romani (the accusative suffix followed by *kor*), and by a preposition together with an accusative pronoun in English. Similarly for the instrumental, which is lexicalised as a single morpheme in Armenian (see the discussion above), but by two morphemes in Budukh and German.

Case	**Language**	**Expression**		
		P	SUFF1	SUFF2
GEN	Vlakh Romani		ACC	*kor*
	English	*Of*	ACC	
DAT	Estonian		GEN	*le*
	Arabic	*Li*	GEN	
INS	Budukh		DAT	*Vn*
	German	*Mit*	DAT	
COM	Georgian		INS	*gan*
	Russian	*S*	INS	

Table 2 Case containment relations (from Caha 2010)

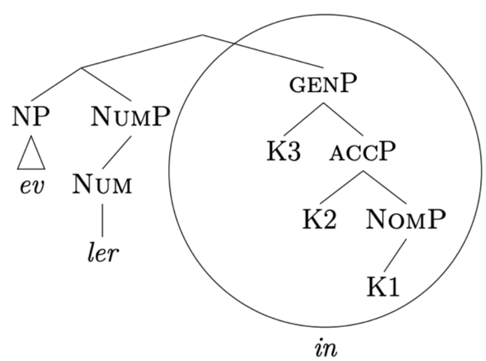
Based on this containment, syncretism patterns, and a number of other facts, Caha (2009, 2010) proposes a hierarchy of cases, such that every containment relation in Table 2 is reflected by structural containment in the functional sequence: the comitative case is built by adding a feature (K6) to the instrumental, the instrumental is built by adding a feature (K5) to the dative, and so on, see (9).

(9)



Once this decomposition is adopted, agglutinative languages with no overt case containment, like Turkish, come out similar to Latin, in that they possess a case portmanteau that lexicalises all the relevant features. This is shown in (10), where it is assumed that NP first moves across Num, and then the entire NumP cyclically moves across the case features.

(10)



For more discussion, see Caha (2009), where it is argued that phrasal lexicalisation (based on a particular matching condition and incorporating the Elsewhere Condition) derives the fact that syncretism between case markers targets adjacent projections in the *fseq*, see (11). Such statements are referred to as the *ABA generalisation (cf. Baerman et al. 2005, Bobaljik 2012, Smith et al. 2019).

(11) In the sequence *nominative – accusative – genitive – dative– instrumental – comitative,* syncretism targets only adjacent cases (Caha 2009: 10).

The study of syncretisms (as in (11)) and containment (as in Table 2) led to a rich body of work investigating the structure of the *fseq* in the nominal domain (Caha 2009; 2022; this volume; Taraldsen 2010; Starke 2017; Bergsma 2019; Kloudová 2020; Türk and Caha 2021; Janků 2022), directional expressions (Pantcheva 2011), negation (De Clercq 2013; 2020; Baunaz and Lander 2023), demonstratives (Lander 2016; Lander and Haegeman 2018), pronominals (Vanden Wyngaerd 2018), complementizers (Baunaz 2018; Baunaz and Lander 2018a; 2018c), cross-categorial syncretisms (Wiland 2019), the inflectional and derivational morphology of verbs (Taraldsen Medová and Wiland 2018; Baunaz and Lander 2019; Starke and Cortiula 2021; Cortiula 2023), degrees (Caha, De Clercq, Vanden Wyngaerd 2019, Vyshnevska 2022, Wiland 2023) or indefinites (Dekier 2021). The investigation of the *fseq* by means of the study of syncretisms and morphological and semantic containment is ongoing, and in this volume, the chapters by Dikmen and Demirok, and by Baunaz and Lander add to this investigation, too.

In some of this research, deviations from the simple rule that morphological containment and syncretism always reflect the underlying functional sequence have, however, also been noted, and they are actually predicted by the Nanosyntax approach, once all the details are considered. One area where a structure with fewer features may have more morphemes than a structure with more features are cases where structures differ ‘in the middle’ (this has been referred to as a gap). In such cases, a single morpheme potentially lexicalising a large structure will fail to do so in case there is a gap in the middle as compared to the syntactic structure, thereby preventing the morpheme from matching. In this situation, it may require multiple morphemes to lexicalise that slightly different (smaller) structure, leading to a situation where the smaller structure requires more morphemes than the bigger one. Markus (2015) explores situations like this to account for the rich morphological patterns of Hungarian middles and passive-like constructions (see also Caha, De Clercq and Vanden Wyngaerd 2023).

Surprisingly, even when structures are in a proper subset relationship, a morpheme lexicalising the bigger version is not always able to lexicalise the smaller version, i.e. it may fail to shrink. In section 5.1 below, we discuss one such case brought up by Blix (2022), the upshot of which is that Nanosyntax allows the two structures to not be in a proper subset structurally, although they are a proper subset feature-wise. This is because the same features can end up organized into different syntactic structures, thanks to movement.

As far as *ABA is concerned, section 5.2 discusses an interesting grammatical ABA case, which always takes the shape A-Ax-A. This also involves a monotonously larger structure, and integrates well into the theory, despite being a surface counterexample to the correlation with morphological containment, as well as the traditional *ABA restriction.

As a result of this, *ABA constraints on syncretism and containment relations turn out to be a useful guideline, but not an absolute criterion for the establishment of the *fseq*. We discuss this in more detail in sections 5 and 6.

3. Root size as an allomorphy trigger

This section introduces the idea that not only functional morphemes, but also roots may differ in how many grammatical features they realise, and in doing so, give rise to allomorphic variation in suffixation patterns. For example, the root *men* is lexically associated to plural, while the root *boy* is not. This is why the latter but not the former needs a plural suffix. This section provides analogous examples from the comparison of Czech, English and Armenian degree morphology.

To introduce the basic idea, consider first Czech in (12). The examples show an increasingly complex morphology as we go from the positive (*inteligentn-í* 'intelligent') to the comparative (*inteligentn-**ějš**-í* 'more intelligent') to the superlative (***nej**-inteligentn-ějš-í* 'most intelligent').

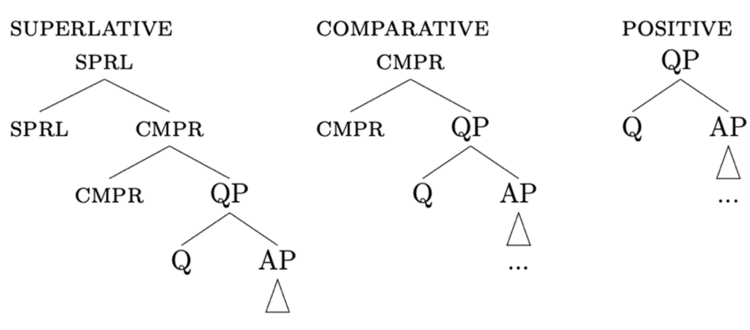
(12) a. Emm-a je inteligentn-í. POSITIVE
Emma-NOM be.3SG.PRS smart-FEM.SG
'Emma is intelligent.'

b. Emm-a je **inteligentn-ějš-í** než Pavel. CMPR
Emma-NOM be.3SG.PRS smart-CMPR-FEM.SG than Pavel
'Emma is more intelligent than Pavel.'

c. Emm-a je **nej-inteligentn-ějš-í** ze všech. SPRL
Emma-NOM be.3SG.PRS SPRL-smart-CMPR-FEM.SG from all
'Emma is the most intelligent of all.'

The increasing complexity of the morphology can be captured under the proposal (due to Bobaljik 2012) that the superlative degree contains the comparative, which is in turn based on the positive, see (13); recall also the discussion of containment in Section 2.

(13)



The positive degree as depicted in (13) is assumed to be bi-componential (enriching slightly Bobaljik's original proposal). On top of the lexical AP, the positive contains a QP, a functional category borrowed from Bresnan's (1973) and Corver's (1997) work that encodes the gradability/scalarity of the adjectives under discussion (see De Clercq et al. 2023 for discussion). The immediate motivation for it is to capture the fact that the adjective *inteligent-n-í* 'intelligent' is bi-componential in Czech (disregarding the agreement), which shows the need to assume at least two positions of exponence in the positive.

Czech contrasts with Armenian (Bobaljik 2012: 88, 91), where the form of the adjective is the same in all degrees, see (14): the adjective for 'tall' is *partsrahasag* regardless of the degree.

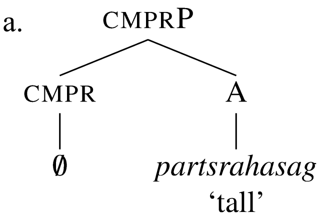
(14) a. Artak-ə partsrahasag e. POSITIVE
Artak-DEF tall be.3SG.PRS
'Artak is tall.'

b. Artak-ə Bagrat-e-n **partsrahasag** e. CMPR
Artak-DEF Bagrat-ABL-DEF tall be.3SG.PRS
'Artak is taller than Bagrat.'

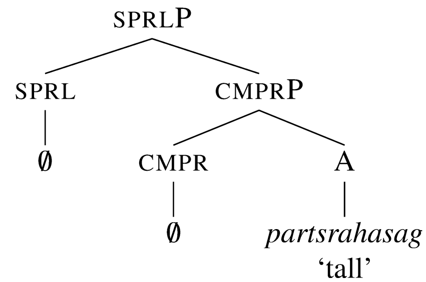
c. Artak-ə amen-e-n **partsrahasag** e. SPRL
Artak-DEF all-ABL-DEF tall be.3SG.PRS
'Artak is the tallest.' (lit. 'Artak is taller than all.')

Assuming that the underlying functional sequence is invariant, a terminal-only insertion would force one into assuming multiple zeroes, as in (15).

(15) a.

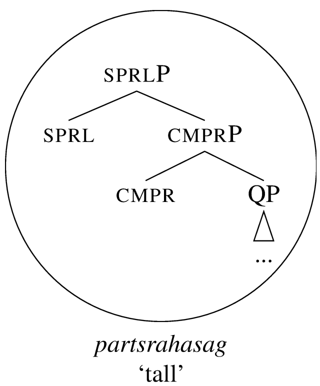


b.

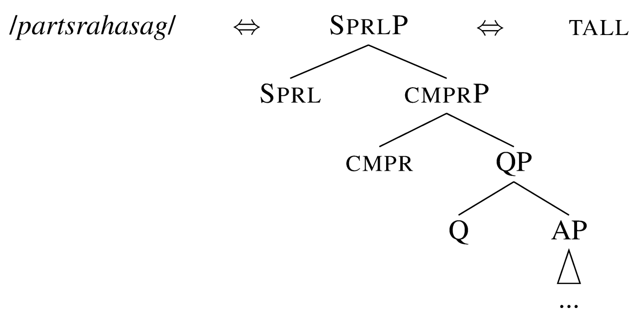


However, if one adopts phrasal lexicalisation, then one could say that Armenian roots like *partsrahasag* 'tall' can lexicalise all features relevant to the adjectival root (QP), as well as the comparative and the superlative, as in (16).

(16)



At this point we need to say something about how lexical items are stored in the lexicon. The general assumption is that lexical items are stored (memorised) links between representations belonging to different modules. For example, the depiction in (16) corresponds to the syntactic representation of the superlative. In the lexical entry (17), the syntactic tree is linked to a phonological representation (between slanted brackets) and conceptual information (in small capitals). In essence, the entry says that if syntax builds a structure like the one in (17), this structure can be linked to the phonological representation */partsrahasag/* and the concept TALL.

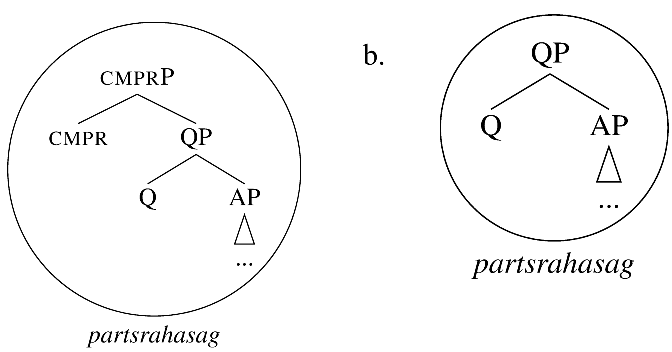
(17)


To determine whether a lexical entry matches a syntactic structure, Nanosyntax relies on identity. Specifically, if the constituent built by syntax is identical to a constituent stored in the lexicon, the entry of that constituent can be used to 'lexicalise' the syntactic representation, i.e., it can be linked to a particular phonological and/or conceptual representation. This condition on matching is given in (18).

(18) *Condition on matching*

A lexically stored constituent L matches a syntactic phrase S iff S is identical to L.

The condition makes it possible for the lexical item in (17) to lexicalise either of the syntactic structures in (19), i.e. those of the comparative and the positive. This is so because any subconstituent of a lexically stored tree is itself stored in the lexicon, and therefore both the CMPRP constituent of (19)(a) and the QP constituent of (19)(b) find a matching constituent in the entry (17).

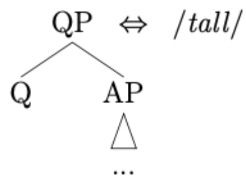
(19) a. b.


Phrasal lexicalisation, understood as identity matching, thus derives the syncretism between the three degrees of comparison that we observe in Armenian without any additional assumptions.

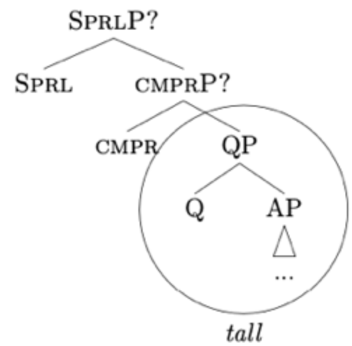
The matching condition (18) is an unpacked version of the matching condition used in Starke (2009): 'A lexically stored tree matches a syntactic node iff the lexically stored tree contains the syntactic node'. This expands to 'A lexically stored tree matches a syntactic node iff the lexically stored tree contains a constituent identical to the syntactic node'. This formulation further implicitly assumes that we look at the root constituent of the lexical tree, and because of that it has to explicitly allow matching by constituents that it 'contains'. Removing that implicit stipulation allows a simplified formulation: 'A lexically stored tree matches a syntactic node iff the lexically stored tree is identical to the syntactic node', i.e. (18). The goal of the unpacking is to make identity requirement on matching explicit. The matching condition (18) -- just like the original formulation -- derives the so-called Superset effect, meaning that a lexical entry can lexicalise any phrase that it contains. Because of this, the matching condition is sometimes referred to as the Superset Principle (also in the subsequent chapters). However, it is important to realise that the Superset Principle is not an independent principle in addition to the matching condition.

Let us now turn to English, which, unlike Armenian, requires additional morphology in the comparative and superlative (*tall* – *tall*-***er*** – *tall*-***est***). If the lexical tree for *tall* is as in (20)(a) (omitting the conceptual information), we can understand why this is so: the lexical tree of *tall* is too small to lexicalize the degree features CMPR and SPRL, as (20)(b) shows. It can only realize the features of AP and QP. In effect, the root size of *tall* determines that the adjective will need help from other lexical items to lexicalize CMPR and SPRL.

(20) a. lexical entry

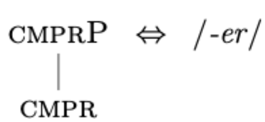


b. syntax

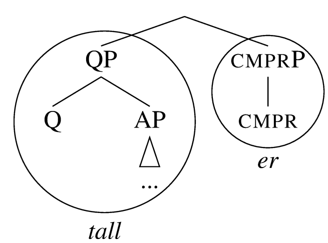


To capture the morphology of the forms *tall-er* and *tall-est*, we need two more assumptions. The first assumption concerns the morphosyntactic structure, namely, we need to assume that the QP in English moves across CMPR by complement movement, yielding the structure in (21)(b). In this structure, the CMPR feature is lexicalised as a part of the CMPRP, using the entry in (21)(a).

(21) a. lexical tree



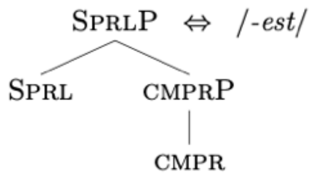
b. syntax



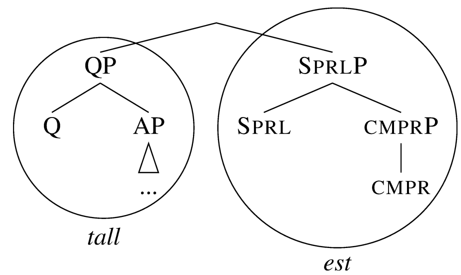
In the syntactic tree (21b), we assume the same types of movements as introduced in Section 2. In addition, traces are omitted, and the nodes created by movement are not labelled (see the discussion in section 2).

In (22) we show an analogous analysis for the superlative.

(22) a. lexical tree



b. syntax



Note that as a consequence of the matching condition (18), the superlative *-est* with the lexical entry as in (22)(a) is a candidate for the lexicalisation of the comparative in (21)(b) because the CMPRP in (21b) is identical to a subtree of (22)(a). The reason why *-er* nevertheless appears in the comparative is that the lexical tree of *-er* in (21)(a) is a better match for the structure in (21)(b) as it has fewer superfluous features compared to (22)(a). In sum, when there are multiple candidates, they compete, and the winner is the lexical entry that is the best match with the syntactic structure. This competition principle is an instance of the Elsewhere Condition, going back to Kiparsky (1973):[2]

(23) In case two rules, R1 and R2, can apply in an environment E, R1 takes precedence over R2 if it applies in a proper subset of environments compared to R2.

Let us now come back to the pattern of the Czech adjective *inteligent-n-í* 'intelligent,' with a comparative form *inteligent-n-**ějš**-í*, and a superlative ***nej**-inteligent-n-ějš-í*. Czech differs from English on two counts. First, unlike in English, the superlative marker does not replace

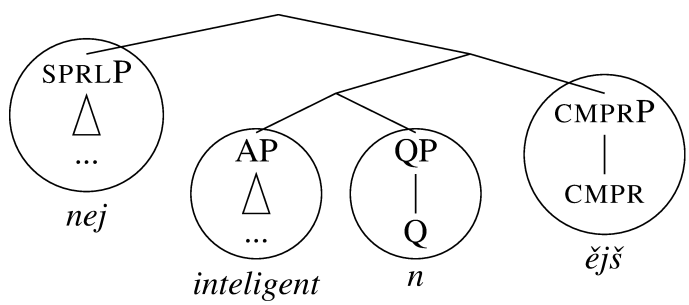
---

[2] The chapter by Balsemin and Pinzin discusses an interesting case where the winner is determined by phonology, rather than by the Elsewhere Principle.

('override') the comparative, but rather attaches on top of the comparative. Intuitively, this is because the Czech morpheme *nej*- only lexicalises the feature SPRL, i.e. it is smaller than the English -*est*. This will require -*ejš*- to realize the CMPR feature in the superlative.

The second difference is that the positive in Czech is bimorphemic, consisting of the root and an adjectiviser -*n.* We may thus informally represent the situation in Czech as in (24).

(24)



Here the root only lexicalises the AP, which moves across Q by complement movement, and the adjectiviser lexicalises the QP. The whole QP then moves again across CMPR, and the CMPRP is lexicalised by the comparative -*ějš*. The lexical entry of *nej*- is unable to lexicalise the CMPR head realised by -*ějš*, and so we see the superlative prefix attaching to the comparative, which is unlike what happens with the English -*est*.

Summarising, this section so far showed that varying root size across languages (i.e., whether the root lexicalises AP, QP, or SPRLP) correlates with whether we find affixal material on the adjective in the positive, comparative and/or superlative. This can be captured if the root always lexicalises whichever features its lexical entry allows it to, and the remaining features are lexicalised by affixes or independent words.

It is important to observe that this type of variation is not only found across languages, but also within languages. For example, the English positive-comparative pair *bad* ~ *worse* has a suppletive root and a zero allomorph of the comparative. In Nanosyntax, the most natural analysis is one where there is no zero, rather *worse* has the same comparative structure as the

Armenian *partsrahasag* 'tall,' i.e., containing no movement, recall (19)(a). This differs from *tall-er*, which requires complement movement, recall (21)(b). If that is so, we end up in a situation where different adjectives require different structures within the same language, i.e., English has both (19)(a) (no movement) and (21b) (complement movement), but for different roots.

Similarly, there are adjectives in Czech that do not require an adjectivizer, and their positive is monomorphemic, e.g., *chytr-ý* 'smart' with the comparative *chytř-ejš-í* 'smarter'. Their structure must be therefore like the one of *tall* ~ *tall-er*, where the root lexicalizes the whole QP. Once again, this means that there is no movement of AP across Q with these adjectives, and we are once again led to conclude that different roots require different structures.

The conclusion that different roots within the same language yield different structures neatly captures the differences between classes of adjectives within a language, and, by extension, differences between different root classes of any type. In this volume, such differences play a crucial role in the contributions by Bergsma, Don, Merkuur and Smith, by Caha and Taraldsen Medová, by Natvig, Putnam and Wilson, by Taraldsen, and by Vyshnevska.

When there are multiple ways of lexicalising the same set of features (e.g., *worse* vs. **badd-er*), the system of lexicalisation is set up in such a way that a preference for one type of lexicalisation against another type follows from the system. More concretely, Starke (2018) proposes that the movements which give rise to different structures are not triggered by syntactic features, but by the need to create a structure that can be lexicalized, i.e., externalised at the interface. The theory of how movement (and structure building in general) interacts with lexicalisation is the main content of the Lexicalisation Algorithm, which is in turn the main focus of this book. The Lexicalisation Algorithm interacts with lexical items in a way that different lexical items give rise to different structures within the same language. The algorithm also correctly determines the preference for a particular type of lexicalisation (e.g. *worse* is preferred over **badd-er*). We provide more detail in the next section.

4. The Lexicalisation Algorithm step-by-step

The first important assumption is that features are merged one by one, and that whenever a new feature is merged, lexicalisation must first succeed before the next feature can be merged. To this end, the structure created by Merge is matched against the existing lexical entries of the particular language. When a matching item is found for the newly formed structure, the derivation may continue by Merge F. However, when no matching item is found, syntax tries to rescue the phrase marker by performing left-branch movement (akin to spec-to-spec movement in traditional theories) and lexicalising the result. If that fails, complement movement is attempted.

These steps are given in (25), and they quite straightforwardly correspond to Cinque's (2005) parameters of variation (no movement in (25)(a), spec movement in (25)(b), complement movement in (25)(c). The ordering of the options sets preferences for particular structures: lexicalisation without movement, whenever possible, is preferrable to movements (*worse* is preferred to *badd-er*).

(25) a. Merge F and lexicalize.

b. If fail, try a spec-to-spec movement and lexicalize.

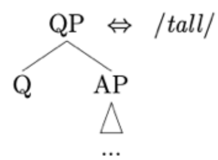
c. If fail, try a movement of the complement of the newly inserted feature and lexicalize.

(adapted from Starke 2018)

Let us now use this algorithm to see how English *tall* differs from the Czech *inteligent-n-í* 'intelligent' in derivational terms. The syntactic derivation starts by assembling AP, see (26)(a) (we assume that the AP is internally complex, as in Vanden Wyngaerd et al. 2020). After this, the lexicon is checked, and since *tall* (repeated for convenience in (26)(b)) contains the AP, it is inserted, see (26)(c).[3]

(26) a.

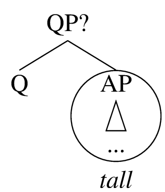
b.



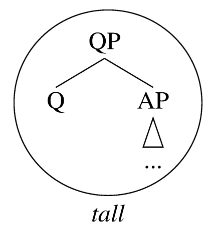
c.

Since matching succeeded without any movement, the gradability feature Q can be merged, yielding (27)(a). (26)(b) is a perfect match and hence can lexicalize the structure, shown in (27)(b). This is the correct lexicalisation of the positive degree.

(27) a.



b.



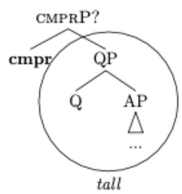
If deriving the positive degree was intended, the derivation would terminate (or merge the adjective with a noun). If the comparative is intended, syntax merges CMPR in the next step, see (28a). The lexicon is checked, but there are no lexical items matching this structure: the lexical

---

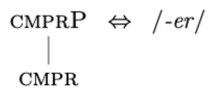
[3] See Caha et al. (2019); Vanden Wyngaerd et al. (2021) for a discussion of how a particular root is selected among those that match. One idea is that roots do not compete in terms of the Elsewhere Condition, and one inserts the root that "one wants to talk about". This can be implemented by proposing that the conceptual representation restricts the candidate set of competing items, so that the root *tall* is the only root in the competition set (all other roots are irrelevant, since they are linked to the wrong concept).

tree for *tall* in (26)(b) does not contain the syntactic tree (28)(a). The lexical entries for *-er* and *-est* (repeated in (28)(b-c)) have CMPR, but neither of them matches the structure in (28a).

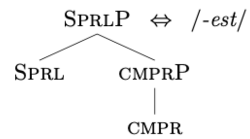
(28) a.



b.

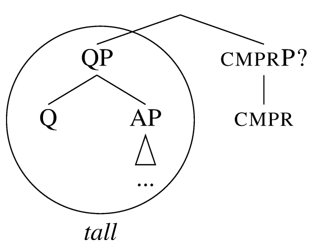


c.

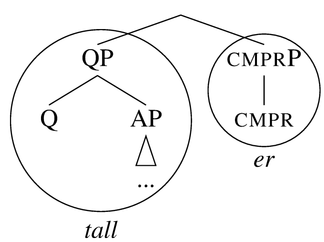


Since no match is found, the syntactic tree (28)(a) cannot be externalised. This leads to rescue movements, following the Lexicalisation Algorithm in (25). The first movement to be tried is moving the spec of the complement, recall (25)(b). However, the complement of CMPR has no spec in (28)(a), so syntax proceeds directly to complement movement (moving QP out of CMPRP), yielding (29)(a). In (29)(a), CMPRP can be lexicalized by the lexical item (28)(b), see (29)(b).

(29) a.



b.

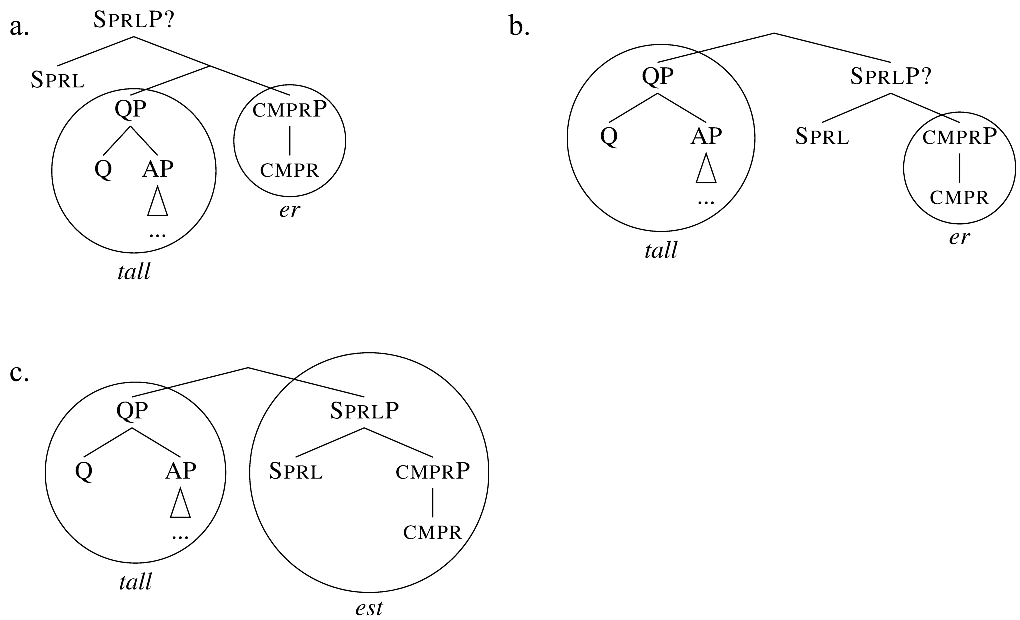


Summarising, movement is necessary if the constituent created by Merge does not find an immediate match in the lexicon. This movement is not a feature-driven, but it applies for the sake of lexicalisation: it is a lexicalisation-driven movement.

Because of this, the algorithm easily handles the fact that different roots interact differently with the derivation: for instance, if the root *worse* (capable of pronouncing (28)(a)) was compatible with the intended conceptual structure, (28)(a) would have been successfully lexicalised, and no movement would have been triggered.

Going back to our derivation of *tall*, consider now how the superlative is derived. First, syntax merges the feature SPRL to (29)(b), yielding (30)(a).

(30)

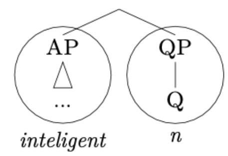


Since lexicalisation is rigidly cyclic, the lexicon is checked again, but there is no matching lexical item for the structure in (30)(a). Therefore, syntax tries moving the specifier (recall (25)(b)), yielding the tree (30)(b). After QP moves out, the SPRLP on the right branch can be lexicalized by the lexical item in (28)(c), overriding the previous lexicalisation, i.e. *-er*, and leading to the structure in (30)(c).

Let us now turn to Czech. In this language, some adjectives (e.g., *chytr-ý* 'smart') are rather similar to *tall* and do not require any extra morpheme in the positive. The morphology of such adjectives can be captured by attributing to them a root of the size QP, like the English *tall.* However, recall that the adjective *inteligent-n-í* 'intelligent' requires a special adjectiviser in the positive, namely *-n*. We can model this by attributing the size AP to the adjectival root *inteligent-* and letting *-n* lexicalise QP, as shown in the lexical entries in (31)(a-b).

(31) a. AP ⇔ */inteligent/*
△
...

b. QP ⇔ */n/*
|
Q

c. 

Since (31)(a) cannot lexicalise the entire structure of the positive, AP has to move out of QP, yielding (31)(c), where the adjectiviser is inserted as the lexicalisation of the QP.

Summarising, we have now seen how the lexical items that a language has at its disposal ‘decide’ which of the three different movement scenarios takes place. Given the appropriate lexical item, Merge can be followed straightaway by lexicalisation without movement as in (27)(b), but it can also be followed spec-to-spec movement (30)(b), or complement movement (31)(c). These are still the same movement types as proposed in Cinque (2005), but which of these applies is not dependent on ‘strong features,’ EPP, OCC and the like, but it depends on the lexical items available in each particular language, in full conformity with the Borer-Chomsky conjecture.

5. The shape of lexical items

In this section, we focus on two apparent paradoxes for the theory introduced so far, namely *ABA violations and inverse marking systems. Systems of inverse marking have the property that a larger structure is lexicalized by less morphemes than a smaller one. A hypothetical case would be one were the positive degree of an adjective required a morpheme in addition to the root (as in the Czech *intelligent-n-i*), and the comparative of that same adjective would be lexicalized by the root alone (as in the case of Armenian is *partsrahasag* ‘taller’ discussed above). The lexicalisation algorithm would appear to rule out such a hypothetical situation: if the lexicon contains a root of size CMPRP, then it is never expected to give rise to movement (and consequent suffixation) in the positive degree. We discuss such a pattern from the domain of singular and plural marking in Kipsigis in section 5.1 below. The other paradox is the existence of *ABA patterns in the domain of verbal inflection, which we turn to in section 5.2

We show that both these phenomena can be accounted for once it is admitted that lexical items may have a more complex internal structure than acknowledged so far. So far, all the left branches inside lexical items only contained terminals, i.e. they were of the shape [X [Y [Z]]]. However, if we moved YP across X, the resulting tree would be of the shape [[Y [Z]] X]. The latter is a well-formed syntactic object, and it is therefore to be expected that lexical trees of this shape can be also stored (Starke 2014a).

Lexically stored trees of this latter type can be referred to as trees with a Complex Left Branch (CLB). The existence of such trees introduces another type of variation in addition to root size. In the example just given, the size of the lexical entries is exactly the same, i.e. the exact same features X, Y, and Z are involved, but the way they are hierarchically organized in the lexical entries is different. The exploration of the possibilities of lexical items with CLBs constitutes the first of the two recent innovations we mentioned in the introduction (the other being subextraction, to which we turn in section 6). It turns out that next to root size, root shape can be put to use to account for certain lexicalisation patterns that are empirically attested, and that would otherwise be unexplained.

5.1. Inverse marking

This section shows that CLB lexical items provide a natural explanation for systems of inverse marking, which raise a paradox stemming from containment (Blix 2022). An example of such a system is the pattern of number marking in Kipsigis (Kalenjin). Table 3 summarizes the core of the system of nominal classification that Blix (2022) discusses.

	root	SG	PL	
a.	kipaw	kipaw	kipaw-**tiin**	‘rhino’
b.	peel	peel-**yaan**	peel	‘elephant’
c.	pata	pata-**yaan**	pat-**een**	‘duck’

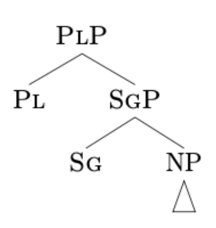
Table 3 Endo-Marakwet (Kalenjin), modified after Blix (2022) via Kouneli (2019, 2021)

The table shows that some nouns -- like *kipaw* ‘rhino’ on row (a) -- show a common pattern where the plural is marked with respect to the singular. Row (b) shows a case where the singular is marked. This type of marking is sometimes referred to as singulative, and the whole pattern can be summarised under the notion of an ‘inverse’: the idea is that each root has a default reference, which is either singular (*kipaw* ‘rhino’) or plural (*peel* ‘elephant’), and the inverse is the marking used when the actual number is the inverse of the default reference. Row (c) shows that there are also nouns where both the singular and plural are marked.

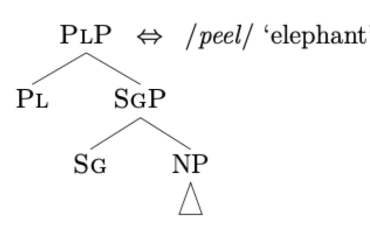
The potential paradox arises because it is impossible to represent both (a) and (b) patterns in terms of morphosyntactic containment because structural containment is asymmetric (i.e., if the structure of the plural properly contains the structure of the singular, it cannot be that at the same time, the singular structure properly contains the plural structure).

To state the problem in terms of the Lexicalisation Algorithm, let us adopt Blix’ functional sequence as in (32)(a) where the plural contains the singular (as in *chair* ~ *chair-s*, or as in row (a) of Table 3). Given that *peel* ‘elephant’ on row (b) has no marking in the plural, we would expect that it has an entry as in (32)(b).

(32) a.



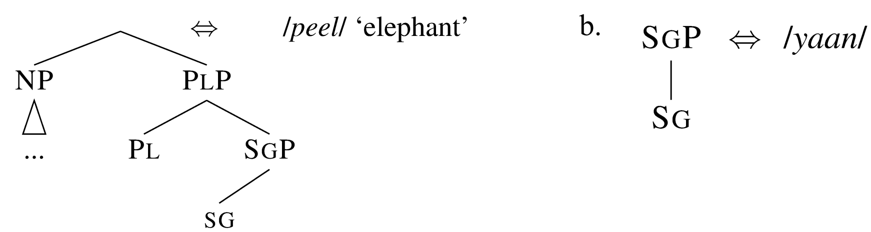
b.



But if that is so, then we would expect that *peel* can also lexicalise the singular number (corresponding to SGP), which is contained inside the lexical tree of *peel*, according to (32)(a).

It turns out that the functional sequence (32)(a) can be maintained even in the light of the facts surrounding the behaviour of the noun *peel* 'elephant.' This can be achieved if the lexical entry of *peel* looks as in (33)(a), containing a complex left branch, namely the NP. Such a tree arises in the course of the derivation when the NP cyclically moves from below SG (where it originates as per (32)(a)) to the position where we see it in (33)(a). Since lexical items are links between well-formed syntactic structures and phonology, and since (33)(a) links a well-formed structure to phonology, such a lexical item is fully in line with the expectations of the theory. Observe that in terms of their size, i.e. the number of features they contain, (32)(b) and (33)(a) contain exactly the same features, but arranged in a different shape.

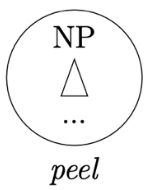
(33) a. b.



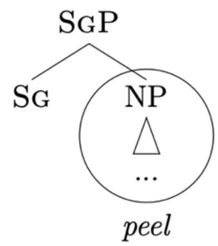
Observe the oblique line connecting SGP with its daughter SG in (33)(a), which serves to visually indicate the extraction site of the NP. It is notationally equivalent to the straight line in (33)(b), and we shall use both notations in what follows. The lexical item in (33)(b) represents the singular marker *yaan*. This entry is consistent with the pattern seen on row (c) of Table 3, where we see a root that has a marker both for the singular and the plural. Since this root always requires a number marker, we analyse it as lexicalising just an NP, and that is why *yaan* is needed in the singular.

Let us now observe how the markers in (33) interact with the derivation of the noun *peel* 'elephant.' The first step in the derivation is to assemble the NP, which is lexicalised as *peel*, since NP is contained in the entry (33)(a), namely on the left branch. Since matching succeeded, the SG feature is merged, see (34)(b).

(34) a.



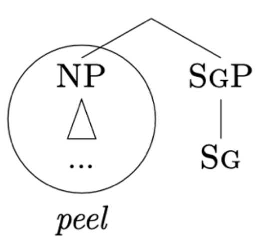
b.



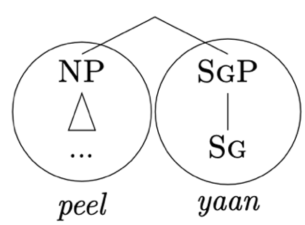
The initial problem with (34)(b) was that when we postulated the entry for *peel* as in (32)(b), the expectation was that the structure (34)(b) will be lexicalised by *peel*, yielding the incorrect unmarked singular. However, after we have modified the entry for *peel* to include a complex left branch, the tree in (34)(b) is actually not contained in the lexical entry (33)(a): the entry contains no constituent that is identical to (34)(b). As a consequence, there is no match with (34)(b) and the Lexicalisation Algorithm resorts to rescue movements.

Spec-to-spec movement is inapplicable, complement movement yields (35)(a). This structure leads to the correct lexicalisation *peel-yaan* 'elephant-SG', as in (35)(b).

(35) a.

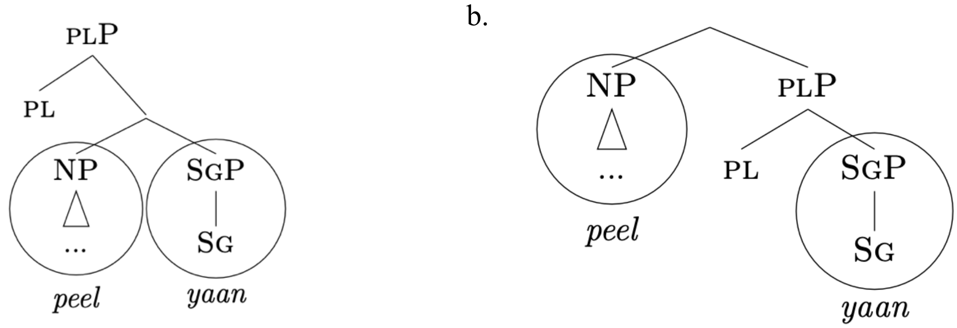


b.



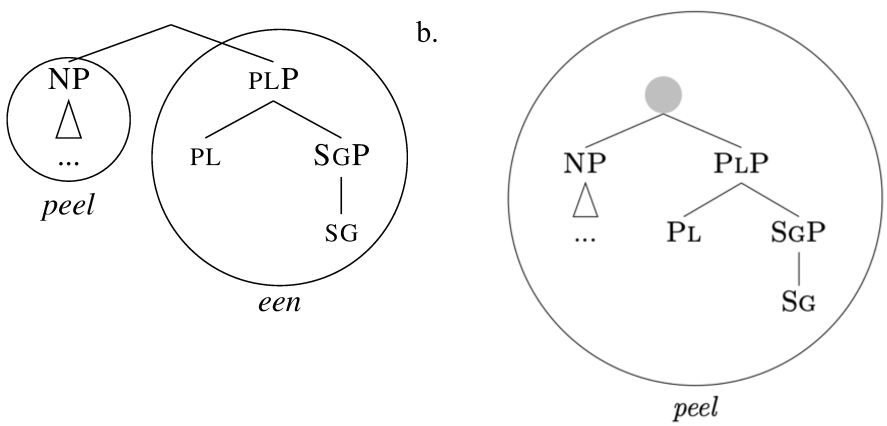
When PL is merged, as in (36)(a), there is no match. Spec-to-Spec movement is therefore tried, yielding (36)(b).

(36) a. b.



There are two ways to lexicalise (36)(b). One option is to use the plural marker *een*, as in (37)(a). This marker appears in the plural of the noun *pata* 'duck' on row (c) of Table 3. Since *pata* 'duck' is just the lexicalisation of an NP, *een* is the realisation of the remaining features.

(37) a. b.



If the lexical entry for *peel* was of the size NP, this would have been the correct result. However, the lexical entry of *peel* is more complex, in that it matches the topmost constituent of the whole structure (marked by a grey circle in (37)(b)), which means that *peel* can actually lexicalise the whole structure. This lexicalisation will therefore override the lexicalisation as in (37)(a).

In sum, once we acknowledge the existence of lexical items with a complex left branch, which are an inherent option in the Nanosyntax theory, we are able to capture 'inverse' patterns of marking, where structures which are more complex in terms of the meaning ingredients have simpler marking (i.e. less morphemes) than less complex structures.

5.2. *ABA patterns

Another important feature of CLB entries is that they allow for the generation of a special type of *ABA pattern, represented as A-Ax-A patterns by Blix (2022). Such a pattern is attested in a dialect of Friulian (spoken in Tualis) studied by Cortiula (2023). In this dialect, the first person is unmarked and syncretic with the third person (*bat* 'I hit/he hits'), while the second adds a suffix to the root (*bat-s* 'you hit'). When the persons are ordered 1-2-3, as in the table, this gives rise to an A-Ax-A pattern, as shown in the Friulian column of Table 4.

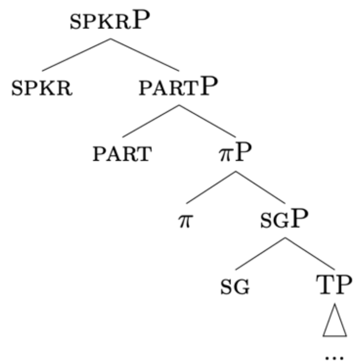
		FRIULIAN *bati* 'hit'	**BULGARIAN** *četa* 'read'
SG	1	bat	čet-ox
	2	bat-s	čet-e
	3	bat	čet-e

Table 4 Present tense of *bati* 'hit' in Tualis Friulian and aorist of *četa* 'read' in Bulgarian

Before we show how CLB trees allow for the generation of this pattern, let us say why Cortiula adopts the ordering 1-2-3, rather than 2-1-3, which would make the Friulian paradigm trivially compatible with *ABA. The reason is that in other languages, of which Bulgarian in the second column of Table 4 is an instance, we find a syncretism between second and third person, which are incompatible with the 2-1-3 order. The syncretism between second and third person is in fact more common than the one between first and third person (even though the latter is far from being rare either; see Baerman et al. 2005; Baerman and Brown 2013; Ackema and Neeleman 2013). Cortiula (2023) therefore accepts the *ABA property of Friulian at face value.

The specific decomposition of the agreement features she works with is given in (38) (see Béjar 2003, Blix 2021a, Starke and Cortiula 2021).

(38)



Let us start unpacking this structure by pointing out that it is essentially an extended version of the proposal by Pollock (1988), who proposed that above tense inflection, there is a dedicated position for agreement features. In (38), each agreement feature is given an independent projection, in accordance with the general tenets of the Nanosyntax approach. Let us now introduce the ingredients.

Immediately above TP, we see the singular projection SG. Plural would require the addition of PL, as discussed in the previous section, but we ignore this here for simplicity. Above number are person features. If only the feature π was present, we would be looking at a default person, i.e., the third person. The second person has the additional feature PART(icipant), and first person is semantically most complex, also including the feature SPKR. With respect to this decomposition, the paradigm found in Friulian represents a *ABA paradigm.

The fact that the form *bat* can lexicalise the first-person structure suggests that it is able to lexicalise all the features present in (38). The fact that it can lexicalise the third-person structure is then consistent with the matching condition (18). What is unexpected is that the root *bat* needs the suffix *-s* in the second person: by standard logic of constituent matching, the root should be able to lexicalise PARTP on its own. In fact, this is the same logic that otherwise bars ABA patterns.

In the form of a lexicalisation table, the pattern looks as in Table 5.

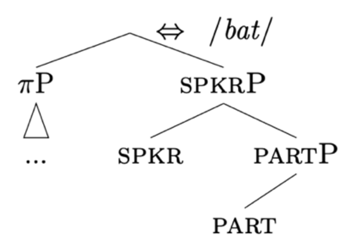
	INDP	T	#	p	PART	SPKR
1SG	bat				*	
2SG	bat				s	
3SG	bat					

Table 5 Lexicalisation table of the present tense singular in TF

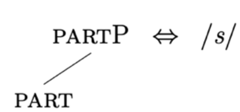
The problem posed by ABA patterns can be formulated as one of nonshrinkability: a large lexical item A cannot shrink to the size of a smaller constituent apparently contained in it, but instead requires another lexical item B, even though A can shrink to a size that is below that of B. We indicate this nonshrinkability by means of the asterisk in Table 5. Applied to the case at hand, the root *bat* can lexicalise a constituent as large as SPKRP (1SG). It can also lexicalise the subconstituent πP of 3SG, but it cannot lexicalise the subconstituent PARTP (2SG), since it needs the suffix *-s* to lexicalise the feature PART. The inability of the root to lexicalise PARTP is what the asterisk in Table 5 indicates.

Lexical entries with CLBs have exactly the properties that are needed to derive such patterns. Cortiula proposes the lexical entry in (39)(a) for the root *bat,* and the one in in (39)(b) for the 2nd person suffix *-s*.

(39) a.



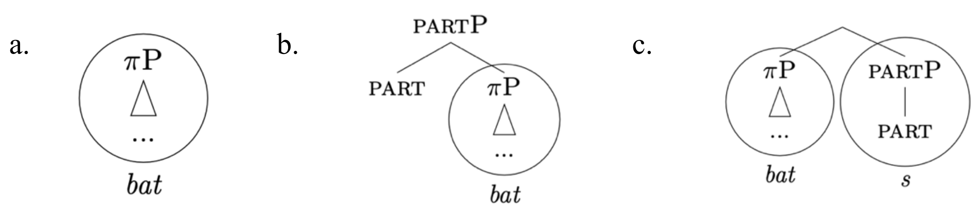
b.



This lexical entry with the CLB can lexicalise the most complex structure (SPKRP), but it does not contain a subconstituent that is identical to the structure of a 2SG: the PARTP in (39)(a) only dominates PART, and not the rest of the features that would be needed for the root to lexicalise the full 2SG form of the root. On the other hand, the entry does contain a subconstituent as a CLB, namely πP, which corresponds with the least complex form (3SG). In the lexicalisation table (Table 5), the feature that is immediately to the left of the asterisk corresponds with the size of the CLB.

Let us now go through the derivation stepwise. The derivation first takes several steps to construct the πP, corresponding to the third person. Since πP is contained in the entry of *bat* in (39)(a) as a CLB, the third person is lexicalised as *bat*, as in (40)(a). When the PART feature is merged, as in (40)(b), there is no match. Therefore, lexicalisation-driven movement is required. Spec-to-Spec movement does not apply as there is no such phrase to be moved, so that complement movement applies, yielding (40)(c), which is lexicalised as *bat-s*.

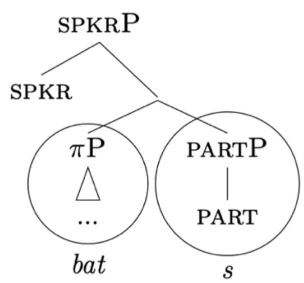
(40)



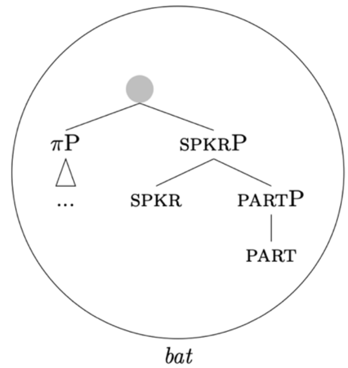
Note that (40)(c) cannot be lexicalised as *bat*, since there is no constituent in (39)(a) that is identical to (40)(c).

To derive the first person, the SPKR feature is merged, as in (41)(a). Since there is no match, spec-to-spec movement is triggered, yielding (41)(b). This constituent is identical to the lexical tree of *bat*, and the first person is therefore lexicalised as *bat*.

(41) a.



b.



In sum, once lexical entries with CLBs are admitted into the framework (and it would be hard to rule them out, since they are a theoretically expected type of an object), they open the possibility for generating attested patterns of inverse marking and ABA patterns. The theoretical relevance of lexical entries with CLBs has been explored by Blix (2021b). Their empirical relevance has been demonstrated in the domain of Czech declension classes by Janků (2022), that of Friulian verb classes (Cortiula 2023), and the morphology of Czech adjectives (Caha, De Clercq and Vanden Wyngaerd 2023a;b). In this volume, CLB entries are explored in the contributions by Bergsma, Don, Merkuur and Smith, by Caha, by Caha and Taraldsen Medová, by Kasenov, and by Vyshnevska.

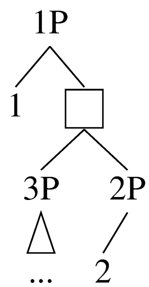
6. Simplifying the lexicalisation algorithm

Starke (2022) suggests a new way of thinking about rescue movements, with the side-effect that a new type of rescue movement becomes available: subextractions. The core of the idea is that gradual pied-piping is itself a rescue strategy. Evacuation then always targets a single node, but that node can pied-pipe gradually larger constituents in the quest for a successful lexicalisation.[4]

[4] Targeting a single node is parallel to Cinque's notion that only NP moves, possibly pied-piping its containing phrases, but adding the idea that pied-piping is a gradual last-resort option.

Let's unpack that by looking at a typical derivation such as (42).

(42)

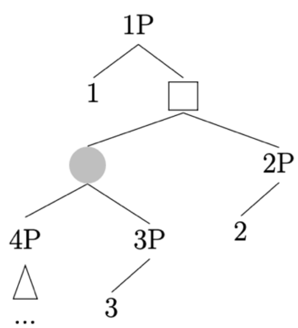


Here 3P has evacuated out of 2P, and the square node lexicalised successfully. Now feature 1 has been merged and the constituent cannot lexicalise as it is. Starke's reformulation of the lexicalisation algorithm always evacuates 3P in this configuration. That leads to a spec-to-spec evacuation. If this evacuation fails to yield a lexicalisation, 3P retries by pied-piping its mother node, i.e. the square constituent. This formulation derives both the existence of complement movement, without the need to stipulate that there is such an option and the fact that it comes after spec-to-spec evacuation, without having to explicitly list that order.

Why 3P though? Here there are two formulations, one in Starke's original 2022 seminar and one in Caha's follow up 2023 seminar. In both formulations, only labelled nodes can be targeted by movement. In Starke's proposal, the target is the labelled constituent closest to 1 inside the previous merge-cycle – i.e. inside the material that was added to 2P in order to create the square node. Within that material, 3P is the closest labelled node. In Caha's formulation, the same effect is obtained by restricting the target to the closest labelled non-remnant constituent. In (42) for instance, 3P is the only non-remnant node since 2P is a remnant in the sense that material has been extracted out of it (see Caha and Taraldsen Medová, this volume; Caha, this volume, for more details).

As a side-effect, this new way of thinking about evacuation movements allows for a new type of rescue movement: subextraction. Consider the tree in (43), a stage of the derivation where 4P has evacuated out of 3P, the resulting circled node has evacuated out of 2P, and the feature 1 is then merged.

(43)



Since only labelled nodes may be evacuation targets, 2P, 3P and 4P could in principle be evacuated, but the round or square nodes could not. In this situation, both Starke's and Caha's formulation have the effect that 4P ends up being the only candidate for evacuation. This then yields a new evacuation type: subextraction from within a "Spec". If the evacuation of 4P fails to create a lexicalisable structure, the mother node of 4P, marked by the gray circle, would be pied-piped, yielding a traditional spec-to-spec evacuation. If that in turn fails to create a lexicalisable structure, another degree of pied-piping is attempted, evacuating the entire square node – yielding a traditional complement movement. The new formulation thus deduces both the types of evacuation movements, and their relative ordering, while predicting new types of evacuation movement: sub-extraction, sub-sub-extraction, etc.[5]

---

[5] The first indication in the Nanosyntax literature that such derivations may be needed appeared in Pantcheva (2011), in a rather different derivational system. Subextractions were also argued for explicitly by Wiland (2019), where subextraction represents a separate step of the algorithm, ordered after complement movement. Subextractions are also used in Cinque's (2005) work.

On the empirical side, this algorithm allows us to preserve previous analyses, because in addition to subextraction (the extraction of 4P in 43), it also allows for spec-movement (i.e., moving the gray circle) and complement movement (i.e., moving the node marked by the square). Moreover, it does so in the same order, always moving the spec before moving the complement. This means that analyses that relied on these two movement types are preserved. However, in addition, the algorithm gives us new analytical options that were not possible before.

Let's show this with a small case study of the Friulian conjugation, using the following formulation of the algorithm for concreteness:

(44) (a) Merge F and lexicalise

(b) If fail, evacuate the closest labelled non-remnant constituent and lexicalise

(c) If fail, evacuate the immediately dominating constituent and lexicalise (recursive)

The Friulian conjugation of *bati* 'to hit' and *finî* 'to finish' are shown in Table 6.[6]

		FRIULIAN	
		fini 'finish'	*bati* 'hit'
SG	1	fin-iʃ	bat
	2	fin-iʃ-s	bat-s

[6] The table ignores certain phonological processes which convert the underlying form of the 2SG into the surface form *finis.* See Cortiula (2023) for discussion.

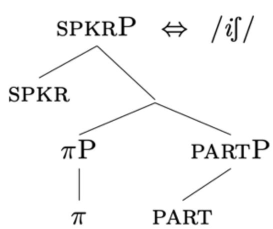
	3	fin-iʃ	bat

Table 6 Present tense singular of *bati* ‘hit’ and *finî* ‘finish in Tualis Friulian

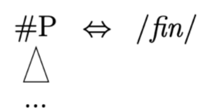
The table shows that the two verb classes have exactly the same pattern: we see the bare stem in the first and third persons, and we see the suffix *-s* in the second person. The difference is that the stem of *bati* ‘hit’ is monomorphemic and corresponds to the bare root, while *fini* ‘finish’ is complex and its stem is composed of two pieces, *fin* and *-iʃ*.

Since these are obviously two instantiations of the same abstract pattern, we will adopt the same analysis for the suffix *-iʃ* as for the root *bat*. Recall that the lexical entry for *bat* (39)(a) above contained a CLB, and the same is true for the lexical entry for *-iʃ* , as shown in (45)(a).

(45) a.



b.



We shall now show that this morpheme is able to lexicalise the first and the third person features, while failing to lexicalise second person features. As in the case of the root *bat*, this is because its shape prevents it from shrinking to the size of PARTP (=2SG). However, since it contains a CLB of size πP, it is able to shrink down to πP, i.e. 3SG.

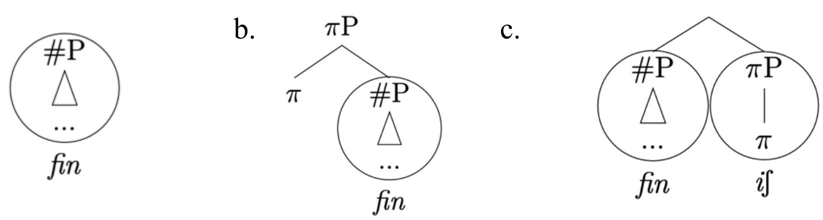
Simplifying Cortiula’s account, we posit the entry for the root *fin-* ‘finish’ as in (45)(b). This entry can only lexicalise features up to #P. Unlike *bat* ‘hit,’ *fin* cannot lexicalise πP, which is why *-iʃ* has to appear. The lexicalisation table corresponding to this analysis is provided in Table 7. The table shows (with an asterisk on the line of 1SG) the special shape of the lexical entry for *-iʃ*, which can shrink to the size of its left branch (πP), but not to the size of PARTP.

	TP	#	π	PART	SPKR
1SG	fin		-iʃ	*	
2SG	fin		-iʃ	s	
3SG	fin		-iʃ		

Table 7 Lexicalisation table of TF singular present tense of *iʃ*-verbs

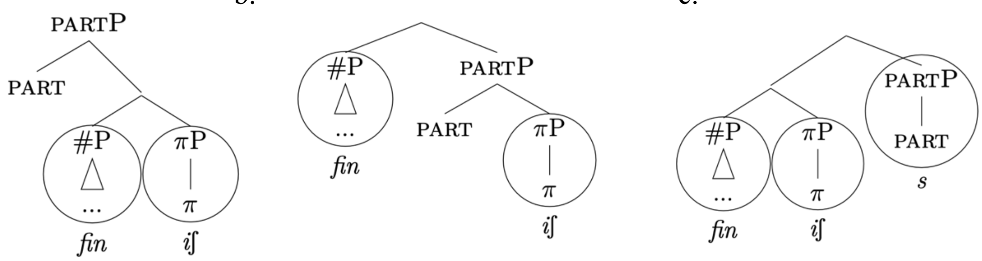
The derivation starts by assembling #P, see (46)(a). When πP is merged, as in (46)(b), *fin* is no longer a match. Following the Lexicalisation Algorithm, we must subextract the node which is closest to the root node, and which is at the same time labelled and non-remnant. In (46)(b), this is the #P node. After #P is evacuated from πP, we get the correct third person *fin-iʃ*, see (46)(c).

(46) a. b. c.



The derivation continues by merging PART, see (47)(a). Since there is no match, we must evacuate the closest non-remnant labelled node. This cannot be the complement of PART (since it is unlabelled), and it cannot be πP (since it is remnant), so it must be #P. Evacuating #P gives us (47)(b), but there is no match for PARTP, since the entry for -*iʃ* does not contain this constituent.

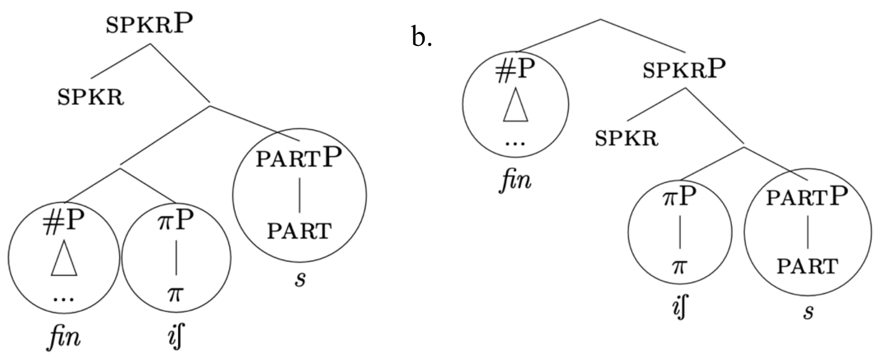
(47) a. b. c.



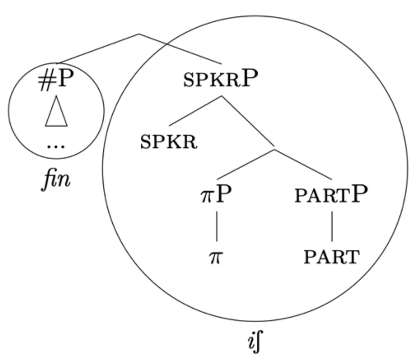
As a result, we revert to (47)(a), and we pied-pipe the node immediately dominating #P, which is the complement of PART. Evacuation of this node yields (47)(c), which can be successfully lexicalised, yielding the correct second person form.

Notice that so far, the subextracting algorithm gives the exact same steps as the standard algorithm: we first moved the spec, and then the complement. The difference between the algorithms surfaces when we introduce the speaker feature, see (48)(a).

(48) a. b.



c.



Since there is no match for (48)(a), we need to move the closest non-remnant labelled constituent. Looking at (48)(a), this is not the spec of the complement, which has no label. Hence, at this point, the difference between the old and the new algorithm emerges, because the subextracting algorithm leads to the evacuation of #P. This is shown in (48)(b). We see that this movement strands the πP in between SPKR and PART, which is precisely the shape of the entry for -*iʃ*, which has a remnant πP stranded between these two features. As a result, the whole SPKRP can be lexicalised as -*iʃ* (48)(c). This derives the ABA-pattern displayed by the -*iʃ*-suffix in Tualis Friulian.

For further explorations of the revised algorithm, we refer the reader to Kasenov (2023), as well as Caha and Taraldsen Medová (this volume) and Caha (this volume).

7. Conclusion

This chapter outlines the basic assumptions of Nanosyntax and discusses two recent evolutions pertaining to the lexicalisation algorithm. The first concerns the possibilities offered by lexical items with Complex Left Branches (CLBs), which considerably extend the empirical coverage of the nanosyntactic framework. The second evolution is a new formulation of the Lexicalisation Algorithm itself, which simplifies it and as a side-effect adds subextractions as a tye of rescue movement.

Finally, we need to point out some important ground that we did not cover. The lexicalisation algorithm in Starke's (2018) version also contains the possibility of backtracking, as well as the formation of a complex left branch in a new workspace. Backtracking can be invoked when the derivation has reached a dead end, and allows it to revert to the previous cycle and try the next option in the algorithm (e.g. movement instead of direct lexicalisation). Complex left branches merged in a new workspace surface as prefixes, or as other material that precedes the root (auxiliaries, complementisers, adverbs, etc). For reasons of space, we have not touched upon these options allowed by the lexicalisation algorithm. We refer the reader to the existing literature where they have been explored in detail (e.g. Janků 2022 for backtracking derivations, and De Clercq 2020 for prefixes).

Another topic that we have not covered is that of feature-driven movement, i.e. the syntactic movement of a phrase that has already been lexicalized in an earlier stage of the derivation. Such movements will need to be incorporated into the lexicalisation algorithm, by allowing the existing derivation to be searched for an already existing lexicalisation of F, either before merge-f or after a failure of merge-f to lexicalise (see De Clercq 2019; 2020). We are confident that future work will explore this option in much greater detail than our summary treatment in the present context.

8. References

Abels, Klaus, and Ad Neeleman. 2012. "Linear asymmetries and the LCA." *Syntax* 15:25–74.

Abney, Stephen. 1987. "The English noun phrase in its sentential aspect." PhD diss., MIT.

Ackema, Peter, and Ad Neeleman. 2013. "Person Features and Syncretism." *Natural Language & Linguistic Theory* 31:901–950.

Baerman, Matthew, and Dunstan Brown. 2013. "Syncretism in Verbal Person/Number Marking." In *The World Atlas of Language Structures Online,* edited by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/ chapter/29.

Baerman, Matthew, Dunstan Brown, and Greville Corbett. 2005. *The syntax-morphology interface: A study of syncretism.* Cambridge: Cambridge University Press.

Baker, Mark. 2008. "The macroparameter in a microparametric world." In *The Limits of Syntactic Variation,* edited by Theresa Biberauer, 351–373. Amsterdam: John Benjamins.

Barbiers, Sjef, Guido Vanden Wyngaerd, and Jenneke Van der Wal. To appear. "Theoretical approaches to comparative syntax." In *The Cambridge Handbook of Comparative Syntax,* edited by Sjef Barbiers, Norbert Corver, and Maria Polinsky. Cambridge: Cambridge University Press.

Baunaz, Lena. 2018. "Decomposing Complementizers: The Fseq of French, Modern Greek, Serbo-Croatian and Bulgarian Complementizers." In *Exploring Nanosyntax,* edited by Lena Baunaz, Karen De Clercq, Liliane Haegeman, and Eric Lander, 149–179. Oxford: Oxford University Press.

Baunaz, Lena, and Eric Lander. 2018a. "Deconstructing categories syncretic with the nominal complementizer." *Glossa: a journal of general linguistics* 3 (1): 31.1–27.

Baunaz, Lena, and Eric Lander. 2018b. "Nanosyntax: the basics." In *Exploring Nanosyntax,* edited by Lena Baunaz, Karen De Clercq, Liliane Haegeman, and Eric Lander, 3–56. Oxford: Oxford University Press.

Baunaz, Lena, and Eric Lander. 2018c. "Syncretisms with the Nominal Complementizer." *Studia Linguistica* 72 (3): 537–570.

Baunaz, Lena, and Eric Lander. 2019. "Subjunctive selection in French vs. Balkan." In *Mapping Linguistic Data. Essays in honour of Liliane Haegeman,* edited by Metin Bağrıaçık, Anne Breitbarth, and Karen De Clercq, 35–49. Ghent: Ghent University.

Baunaz, Lena, and Eric Lander. 2023. "Modal negators in Romeyka Greek and the negative functional sequence." *RGG* 45 (1): 1–25.

Béjar, Susana. 2003. "Phi-Syntax: A Theory of Agreement." PhD diss., University of Toronto.

Benincà, Paola, and Cecilia Poletto. 2004. "Topic, Focus, and V2: Defining the CP Sublayers." In *The Structure of CP and IP,* edited by Luigi Rizzi, 52–75. Oxford: Oxford University Press.

Bergsma, Fenna. 2019. "Mismatches in free relatives–grafting nanosyntactic trees." *Glossa* 4 (1.119): 1–37.

Bergsma, Fenna. 2021. "Case competition in headless relatives." PhD diss., Goethe-Universität.

Blake, Barry. 1994. *Case.* Cambridge: Cambridge University Press.

Blix, Hagen. 2021a. "Phrasal spellout and partial overwrite: On an alternative to backtracking." *Glossa* 6 (1): 62.1–17.

Blix, Hagen. 2021b. "Spans in South Caucasian agreement: Revisiting the pieces of inflection." *Natural Language & Linguistic Theory* 39:1–55.

Blix, Hagen. 2022. "Interface Legibility and Nominal Classification: A Nanosyntactic Account of Kipsigis Singulatives." *Glossa: a journal of general linguistics* 7 (1): 1–46.

Bobaljik, Jonathan. 2012. *Universals In Comparative Morphology.* Cambridge, MA: MIT Press.

Borer, Hagit. 1984. *Parametric syntax.* Dordrecht: Foris.

Bresnan, Joan. 1973. "Syntax of the comparative clause construction in English." *Linguistic Inquiry* 4:275–343.

Caha, Pavel. 2009. "The Nanosyntax of Case." PhD diss., University of Tromsø.

Caha, Pavel. 2010. "The Parameters of Case Marking and Spell-Out Driven Movement." In *Linguistic Variation Yearbook,* edited by Jeroen Van Craenenbroeck, vol. 10. John Benjamins.

Caha, Pavel. 2018. "Notes on insertion in Distributed Morphology and Nanosyntax." In *Exploring Nanosyntax,* edited by Lena Baunaz, Karen De Clercq, Liliane Haegeman, and Eric Lander, 57–87. Oxford: Oxford University Press.

Caha, Pavel. 2019. "Case competition in Nanosyntax. A study of numerals in Ossetic and Russian." Ms. Masaryk U.

Caha, Pavel. 2023. "Tsez spatial cases." Seminar at Masaryk U.

Caha, Pavel. This volume. "A local analysis of an apparent non-local allomorphy in Tamil: a perspective from Rutul." In *Nanosyntax and the Lexicalisation Algorithm,* edited by Pavel Caha, Karen De Clercq, and Guido Vanden Wyngaerd. Oxford University Press.

Caha, Pavel. To appear. "Nanosyntax: some key features." In *The Cambridge Handbook of Distributed Morphology,* edited by Artemis Alexiadou, Ruth Kramer, Alexc Marantz, and Isabel Oltra-Massuet. Cambridge: Cambridge University Press.

Caha, Pavel, Karen De Clercq, and Guido Vanden Wyngaerd. 2019. "The Fine Structure of the Comparative." *Studia Linguistica* 73 (3): 470–521.

Caha, Pavel, Karen De Clercq, and Guido Vanden Wyngaerd. 2023a. "Allomorphy without context specification: A case study of Czech *n*." NELS, Göttingen.

Caha, Pavel, Karen De Clercq, and Guido Vanden Wyngaerd. 2023b. "Diminutive adjectives in Czech as evidence for a rich internal structure of gradable adjectives." Olomouc Linguistics Colloquium.

Caha, Pavel, and Lucie Taraldsen Medová. This volume. "Syncretism and markedness paradoxes in the Russian demonstrative declension." In *Nanosyntax and the Lexicalisation Algorithm,* edited by Pavel Caha, Karen De Clercq, and Guido Vanden Wyngaerd. Oxford University Press.

Chomsky, Noam. 1995. *The Minimalist Program.* Cambridge, Massachusetts: MIT Press.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-linguistic Perspective.* Oxford: Oxford University Press.

Cinque, Guglielmo. 2005. "Deriving Greenberg's Universal 20 and Its Exceptions." *Linguistic Inquiry* 36 (3): 315–332.

Cinque, Guglielmo, and Luigi Rizzi. 2010. "The Cartography of Syntactic Structures." In *The Oxford Handbook of Linguistic Analysis,* edited by Bernd Heine and Heiko Narrog, 51–65. Oxford: Oxford University Press.

Cortiula, Maria. 2023. "The Nanosyntax of Friulian Verbs. An Analysis of the Present and Past in Tualis Friulian." PhD diss., Masaryk University.

Corver, Norbert. 1997. "Much-Support as a Last Resort." *Linguistic Inquiry* 28:119–164.

De Clercq, Karen. 2013. "A Unified Syntax of Negation." PhD dissertation, University of Ghent.

De Clercq, Karen. 2019. "French negation, the Superset Principle and feature conservation." In *Cycles in Language Change,* edited by Miriam Bouzouita, Anne Breitbarth, Lieven Danckaert, and Elisabeth Witzenhausen, 199–227. Oxford: Oxford University Press.

De Clercq, Karen. 2020. *The Morphosyntax of Negative Markers. A Nanosyntactic Account.* Mouton de Gruyter.

De Clercq, Karen, Pavel Caha, Michal Starke, and Guido Vanden Wyngaerd. 2023. "Degree morphology." In *The Wiley Blackwell Companion to Morphology,* edited by Peter Ackema, Eulàlia Bonet, Sabrina Bendjaballah, and Antonio Fábregas. Oxford: Blackwell. https://doi.org/ 10.1002/9781119693604.morphcom024.

De Clercq, Karen, and Guido Vanden Wyngaerd. 2017. "*ABA revisited: evidence from Czech and Latin degree morphology." *Glossa* 2 (1): 69: 1–32.

Dekier, Jakub. 2021. "Morphosyntax of specific and non-specific indefinite markers." *Glossa* 6 (1.70): 1–33.

Frascarelli, Mara. 2007. "Subjects, topics and the interpretation of referential *pro*." *Natural Language & Linguistic Theory* 25 (4): 691–734.

Frascarelli, Mara, and Roland Hinterhölzl. 2007. "Types of topics in German and Italian." In *On Information Structure, Meaning and Form: Generalizations across languages,* edited by Kerstin Schwabe and Susanne Winkler, 87–116. Amsterdam: John Benjamins.

Greenberg, Joseph. 1961. "Some universals of grammar with particular reference to the order of meaningful elements." In *Universals of language,* edited by Joseph Greenberg. Cambridge, Massachusetts: MIT Press.

Janků, Lucie. 2022. "The Nanosyntax of Czech Nominal Declension." PhD diss., Masaryk University.

Kasenov, Daniar. to appear. "ABA in Russian adjectives, subextraction, and Nanosyntax." In *Advances in Formal Slavic Linguistics 2022.* Language Science Press.

Kayne, Richard. 1994. *The Antisymmetry of Syntax.* Cambridge, MA: MIT Press.

Kiparsky, Paul. 1973. "'Elsewhere' in phonology." In *A Festschrift for Morris Halle,* edited by Stephen Anderson and Paul Kiparsky, 93–106. New York: Holt, Rinehart & Winston.

Kloudová, Veronika. 2020. *The Interaction of Functional Morphemes inside the Nominal Phrase.* München: Lincom.

Koopman, Hilda. 2000. *The syntax of specifiers and heads.* London: Routledge.

Kouneli, Maria. 2019. "The syntax of number and modification: An investigation of the Kipsigis DP." PhD diss., NYU.

Kouneli, Maria. 2021. "Number-based noun classification: The view from Kipsigis." *Natural Language & Linguistic Theory* 39 (4): 1195–1251.

Márkus, Andrea. 2015. "Taming the Hungarian (in)transitivity zoo." PhD diss., University of Tromsø.

McCawley, James. 1968. "Lexical insertion in a transformational grammar without Deep Structure." *CLS* 4:71–80.

Neeleman, Ad, and Kriszta Szendroi. 2007. "Radical Pro Drop and the Morphology of Pronouns." *Linguistic Inquiry* 38 (4): 671–714.

Pantcheva, Marina. 2011. "Decomposing Path: The Nanosyntax of Directional Expressions." PhD diss., University of Tromsø.

Poletto, Cecilia. 2008. "On negative doubling." *Quaderni di Lavoro ASlt* 8:57–84.

Rizzi, Luigi. 1997. "The Fine Structure of the Left Periphery." In *Elements of Grammar. Handbook in Generative Syntax,* edited by Liliane Haegeman, 281–337. Dordrecht: Kluwer.

Rizzi, Luigi. 2004. "Locality and Left Periphery." In *Structures and Beyond,* edited by Adriana Belletti, 223–251. Oxford University Press.

Starke, Michal. 2002. "The day syntax ate morphology." EGG class lectures, Novi Sad.

Starke, Michal. 2009. "Nanosyntax: A Short Primer to a New Approach to Language." *Nordlyd* 36:1–6.

Starke, Michal. 2014a. "Cleaning up the lexicon." *Linguistic Analysis* 39:245–256.

Starke, Michal. 2014b. "Towards Elegant Parameters: Language Variation reduces to the Size of Lexically-Stored Trees." In *Linguistic Variation in the Minimalist Framework,* edited by M. Carme Picallo, 140–152. Oxford: Oxford University Press.

Starke, Michal. 2017. "Resolving (DAT = ACC)≠ GEN." *Glossa* 2 (1): 104.1–8.

Starke, Michal. 2018. "Complex Left Branches, Spellout, and Prefixes." In *Exploring Nanosyntax,* edited by Lena Baunaz, Karen De Clercq, Liliane Haegeman, and Eric Lander, 239–249. Oxford: Oxford University Press.

Starke, Michal. 2021. "Online Nanoseminar on Verb Morphology." Brno, Czech Republic.

Starke, Michal, and Maria Cortiula. 2021. "Bearing the Brazilian Cross." *Revista Virtual de Estudos da Linguagem - ReVEL* 19 (18): 9–35.

Svenonius, Peter. 2010. "Spatial P in English." In *Mapping spatial PPs,* edited by Guglielmo Cinque and Luigi Rizzi, 127–160. The cartography of syntactic structures 6. Oxford: Oxford University Press.

Taraldsen, Knut Tarald. 2019. "An introduction to Nanosyntax." *Linguistics Vanguard* 5: 1-11.

Taraldsen, Knut Tarald, Lucie Taraldsen Medová, and David Langa. 2018. “Class prefixes as specifiers in Southern Bantu.” *Natural Language & Linguistic Theory,* 1–56.

Taraldsen Medová, Lucie, and Bartosz Wiland. 2018. “Functional sequence zones and Slavic L>T>N participles.” In *Exploring Nanosyntax,* edited by Lena Baunaz, Karen De Clercq, Liliane Haegeman, and Eric Lander, 305–328. Oxford: Oxford University Press.

Türk, Utku. 2020. “Tackling the Augmentative Puzzle in Turkish.” In *Proceedings of the Workshop on Turkic and Languages in Contact with Turkic,* edited by Bilge Palaz and Larson Stromdahl, 5:155–162.

Türk, Utku, and Pavel Caha. 2021. “Nanosyntactic Analysis of Turkish Case System.” In *Proceedings of the Workshop on Turkic and Languages in Contact with Turkic,* edited by Songül Gündoğdu, Sahar Taghipour, and Andrew Peters, 6:1–15. 5051.

Vanden Wyngaerd, Guido. 2018. “The feature structure of pronouns: a probe into multidimensional paradigms.” In *Exploring Nanosyntax,* edited by Lena Baunaz, Karen De Clercq, Liliane Haegeman, and Eric Lander, 277–304. Oxford: Oxford University Press.

Vanden Wyngaerd, Guido, Karen De Clercq, and Pavel Caha. 2022. “A Nanosyntactic Approach to Dutch Deadjectival Verbs.” *Linguistics in the Netherlands* 39:245–267.

Vanden Wyngaerd, Guido, Michal Starke, Karen De Clercq, and Pavel Caha. 2020. “How to be positive.” *Glossa* 5 (1): 23. 1–34.

Vyshnevska, Anastasiia. 2022. “There is more to the Morphology of -š/-iš: the Comparative in Ukrainian.” *Linguistics in the Netherlands* 39:263–278.

Wągiel, Marcin, and Pavel Caha. 2021. “Complex simplex numerals.” *Acta Linguistica Academica* 68 (4): 470–515.

Weerman, Fred, and Jacqueline Evers-Vermeul. 2002. “Pronouns and case.” *Lingua* 112:301–338.

Wiland, Bartosz. 2019. *The spell-out algorithm and lexicalisation patterns. Slavic verbs and complementizers.* Berlin: Language Science Press.

Wiland, Bartosz. 2023. “Polish comparative adjectives and adverbs.” Ms. Adam Mickiewicz University, Poznań.

Zanuttini, Raffaella. 1991. “Syntactic Properties of Sentential Negation: A Comparative Study of Romance Languages.” PhD diss., University of Pennsylvania.

Zanuttini, Raffaella. 1997. *Negation and Clausal Structure.* Oxford: Oxford University Press.

Ziková, Markéta. 2016. “Supletivismus.” In *Nový encyklopedický slovník češtiny,* edited by Petr Karlík, Marek Nekula, and Jana Pleskalová, 1795–1796. Praha: Nakladatelství Lidové noviny.