The Unpublished Manuscript

A collection of Lingbuzz papers to celebrate
Michal Starke's 50th birthday

Pavel Caha   Karen De Clercq
Guido Vanden Wyngaerd (eds.)

Contents

Introduction

Today, May 1st 2018, Michal Starke turns 50. We wanted to celebrate the occasion by compiling a small volume of papers for him that we are uploading as a gift to Lingbuzz today.

This collection of papers is not meant to be the final volume ever to honour Michal. It was not the aim to involve all people he has ever worked with or has been inspired by, but rather to put together a collection of papers with contributions from people whose work is deeply rooted in Nanosyntax, like the people who contributed to this volume.

We invited the contributors to make an empirical point, rather than a theoretical one, to presuppose knowledge of the nanosyntactic framework (such as phrasal spellout, the spellout procedure, spellout-driven movement, pointers, etc.), and to observe a page limit of 15 pages. The papers have undergone some light reviewing for the sake of coherence and clarity.

Importantly, the volume wants to be more than a birthday gift. We also want to pay tribute to Michal for the role he has played as the creator of Nanosyntax, a framework that the contributors to this volume love working in and that we feel has given us a new toolbox to approach the intricate and often gory details of natural language. In addition, we also want to thank Michal for founding Lingbuzz, and for his lasting efforts to maintain it, thus providing the linguistic community—for decades now—with an online platform that stimulates the free dissemination of ideas. One cannot but wonder whether something of the spirit of May 1968 got under the skin of that newborn baby 50 years ago.

We hope you enjoy reading this brand-new *Unpublished Manuscript*!

Happy birthday, Michal!

Pavel Caha
Karen De Clercq
Guido Vanden Wyngaerd

1 Ontological categories

Lena Baunaz & Eric Lander

UNIVERSITY OF ZURICH & UNIVERSITY OF GOTHENBURG

1.1 Introduction

This paper examines ontological categories (OC), a type of category that seems to exist in all languages. Following current trends in linguistic research, we claim that these categories belong to a closed class of functional nouns and that they make up a particular functional domain in UG. Working within the nanosyntactic approach, we claim that such a functional domain should be seen as a functional sequence (fseq) of syntactico-semantic heads.

Our paper is organized as follows. In section 1.2 we present the main notion of ontological category. We then briefly introduce our theoretical framework, the main hypothesis, as well as the methodology behind our nanosyntactic analysis. In sections 1.3 and 1.4, we present and discuss the relevant data, namely syncretism and morphological containment, in more than 39 languages belonging to 23 different language families. Section 1.5 discusses the nature of some of the counterintuitive results we arrive at, with some implications of the fseq we are proposing.

1.2 Mapping out ontological categories

1.2.1 Ontological categories: a closed class of functional nouns

A crucial topic of Kayne (2005) is that certain functional categories, like place, thing, years, color, etc. may—depending on the language—have null pronunciation while nevertheless being (universally) present in the syntax. This is the case for the functional noun hours in (1): it is overtly realized in French (1a), but not in English (1b) (where small caps indicates non-pronunciation).

(1)   What time is it?

    a.   Il est 3 *(heures).
    b.   It's 3 HOURS.

In the same vein, Cinque has proposed that headless/free relative clauses involve 'a silent external Head (of a restricted class: THING, AMOUNT, PLACE, TIME, PERSON, MANNER)' (Cinque 2008: 18). In (2) we see some more examples from English, which again does not have overt realisations of these functional nouns.

(2)   John bought [[what THING Mary wanted] (SUCH) THING]

As Cinque (2008: 18) points out, some languages do show overt realizations of these 'dummy' nouns (i.e. generic nouns meaning 'thing', 'person', etc.). This is illustrated in (3) with Lango (Nilo-Saharan (?), Noonan 1992: 220).

(3)   márô   gìn   [àmê   cámô]
        3s.like.HAB thing REL + PART 3s.eat.HAB
        'He likes what he eats'

The phenomenon is also seen in Gbe languages, Papuan languages and some Austronesian languages, among others (see Cinque 2016 for a sample). The functional nouns mentioned by Kayne (2005) and Cinque (2008) are often referred to as ontological categories in the typological literature (see Diessel 2003; Haspelmath 1997, among others).

According to Kayne (2005) and Cinque (2008), this class of nouns has limited and presumably universal membership. That is, they form a closed class of functional elements. If we take seriously the claim that there is such a closed class of functional nouns, then we can assume that they make up a particular functional domain of UG. In the cartographic framework, functional categories occupy dedicated positions in the functional sequence: this has been shown for scope-discourse properties (Rizzi 1997), for functional adverbs in the IP domain (Cinque 1999), for functional adjectives in the nominal spine (Cinque 2010), and for discourse particles in the left periphery of the clause (see Haegeman & Hill 2013; Haegeman 2014).

1.2.2 Mapping out ontological categories

Following the general nanosyntactic approach, such a functional domain should be encoded as a functional sequence of syntactico-semantic heads.[1] Our theoretical framework is Nanosyntax (Starke 2009; 2011;

---

[1]Note that functional approaches generally favor 'semantic maps' of many of the elements we are investigating (Haspelmath 1997).

Caha 2009), which dictates a much stricter linear ordering of structure. In Nanosyntax, syncretism and morphological containment are tools used to uncover the fine-grained ordering of heads in an fseq. Syncretism establishes linear ordering, while morphological containment allows us to establish a hierarchy (i.e. the hierarchical order of functional heads).

1.2.2.1 Syncretism

The nanosyntactic research program aims at systematically mapping out the universal structure of natural language. It offers a precise set of methodological tools to do this. Syncretism in particular can be used to discover the atoms of syntax and the way in which they are ordered. Syncretism can be defined as 'a surface conflation of two distinct morphosyntactic structures' (Caha 2009: 6). It thus arises when two or more distinct grammatical functions are spelled out by a single morpheme. Furthermore, syncretism is constrained, in that the phenomenon targets only adjacent cells in a paradigm. Since not every language will show every possible syncretism, a crosslinguistic approach is required, hopefully providing enough puzzle pieces for piecing together a full fseq (assuming, crucially, that ABA patterns are ruled out). We take for granted that the reader has a familiarity with the nanosyntactic theory of syncretism and morphological containment.

To take an example that is related to our domain of inquiry, Lakhota *táku* can lexicalise either an indefinite or an interrogative pronoun, or even a generic noun meaning 'thing'. We take this to be a syncretism. The lexical entry for *táku* is seen in (4):

(4)   < /táku/ ⇔ [INDF [WH [THING]]] >

We have a single lexical entry here that can, by the Superset Principle, apply in multiple syntactic environments.

(5)   [INDF [WH [THING]]]   ⇒   *táku*   (indefinite pronoun)
              [WH [THING]]   ⇒   *táku*   (interrogative pronoun)
                  [THING]   ⇒   *táku*   (ontological category)

Note that German *etwas* exemplifies the same type of thing: indefinite *etwas* contains the wh-element *was*, which arguably contains the ontological category *-as* (which is also found with demonstratives like *d-as*). See (6).

(6)   German
        [*et*-[*w*-[-*as*]]]

1.2.2.2 Stem syncretism

There are cases where only part of the word is shared across cells. We call this stem syncretism. In Lakhota, for instance, the 'circumstantial stem' marking non-specificity has two forms, *to-* and *tu-* Ingham (2003: 52).

(7) Lakhota stem syncretism

THING	PERSON	PLACE	MANNER	AMOUNT	TIME
táku	**tú**-wa	**tu**-ktél	**tó**-kheškhe	**tó**-nakeča	**tó**-ha

Even though both *tu-* and *to-* appear to be indefinite/interrogative markers, the lexical packaging of these two morphemes should (obviously) be slightly different, just like the lexical packaging of *-kheskhe* and *-wa* should be. For the sake of argument we could imagine that they are structurally partitioned as in (8).

(8)

INDF	WH	… φ …	OC
to-		-kheškhe, -nakeča, etc.	
tu-			-wa, -ktél

At this point, one might wonder whether stem syncretism tells us anything about OCs, but actually it does (or can). We can view stem syncretism in terms of 'selection': some OC morphemes are large enough to fit with the smaller stem *to-*. Some are too small and must take the bigger stem *tu-*. So *tu-* will never co-occur with *-kheškhe* or *-nakeča*, just like *to-* will never co-occur with *-wa* or *-ktél*. Hence it could be argued that the size of the stem determines the size of the co-occurring morpheme (and vice versa). Thus if elements in a paradigm share a particular stem, we can assume that these elements must be (about) the same size and thus in the same area of the fseq (if not right next to each other). Since stem syncretism involves potential complications beneath the surface that are not always easy to disambiguate, however, we will avoid using stem syncretism where possible when putting together our fseq of ontological categories.

1.2.2.3 Getting under the surface

Languages are not always as transparent as Lango when it comes to overtly realizing ontological categories as 'dummy' nouns. Much more commonly, ontological categories are buried deep within indefinite or *wh*-pronouns. Nevertheless we can study them in a methodologically systematic way. Whatever it is that makes something an interrogative or an indefinite pronoun (some kind of WH-features or INDF features, respectively), by keeping this variable constant (e.g. comparing only

interrogatives across the board) we are in fact factoring out these extraneous features and getting to the ontological categories at the core of these words. More concretely, whatever it is that makes *what* and *who* different, it is crucially not the WH-feature.

Our work is in line with current work in nanosyntax on the functional structure of *wh*-words in Scandinavian languages, as elaborated by Vangsnes (2013), although we develop a macro-comparative/typological perspective involving more languages. We also note that Vangsnes argues that there are two functional sequences for *wh*-words: a D-related sequence, and a P-related sequence. Our results suggest that only one sequence is needed.

1.2.3 Summary and the data

Our working hypothesis is thus that a crosslinguistic collection of attested syncretisms in the domain of OCs should reveal the linear order of functional layers relevant to OCs and that morphological containment will decide in which hierarchical order the functional layers are directed. In the following sections we show that we do in fact find both syncretism and containment in the domain of ontological categories.

With English as a starting point, we note that it quite straightforwardly distinguishes only seven interrogative categories, to which we link 7 OCs (9a-g). Other languages distinguish an additional category, 'amount', as in French *combien,* or German *wieviel,* meaning 'how much/how many'.[2]

(9)   a.   what   THING
      b.   who   PERSON
      c.   where   PLACE
      d.   how   MANNER
      e.   when   TIME
      f.   which   FORM (=Selection/Determiner)
      g.   why   REASON
      h.   combien   AMOUNT (=quantity)

Cysouw (2004) shows that 5% of the world's languages show a PERSON/THING syncretism, 90% lexicalize the interrogative category of PLACE , 60% lexicalize the interrogative category of FORM (or 'selection'), 40%

---

[2]A reviewer wonders whether we could add (THE) FACT as a universal ontological category. The reviewer notes that it appears to be universally present in factive clauses, and can, in some languages be left non-overt (as in English *I regret {the fact/Ø} that the reviewer is right,* for instance). Nouns like (THE) FACT have not been considered in our study, mostly because our empirical material centers on indefinite and interrogative pronouns, while (THE) FACT is clearly definite.

lexicalize the interrogative category of MANNER, 40% lexicalize the interrogative category of quantity, and 40% lexicalize the interrogative category of TIME. The lexeme for REASON is almost universally derived from THING; for this and other complicating reasons, we did not include it in our study for now. The ontological categories which we look at are thus the following:

(10) a. what   THING
      b. who   PERSON
      c. where   PLACE
      d. how   MANNER
      e. how much/many   AMOUNT
      f. when   TIME
      g. which   FORM

We looked at 39 languages belonging to 23 language families, listed here.

(11) Arawá (Paumari), Arawakan (Perené Asheninka, Asheninka Campa, Terena, Bare, Warekana, Amuecha), Aymaran (Jaqaru), Austronesian (Muna, Tukang Besi, Rapanui), Barbacoan (Awa Pit), Eskimo-Aleut (Greenlandic), Finno-Ugric (Finnish, Hungarian), Indo-European (Danish, Polish, Germanic, Latvian, Lithuanian, Greek), Kwa (Krachi), Kuliak (Ik), Kwa (Krachi), Muran (Pirahã), (West) Papuan (Maybrat), Quechuan (Imbabura Quechua), Ramu-Lower Sepik (Yimas), Sino-Tibetan (Dumi), Siouan (Lakhota), Thai-Kadai (Thai), Tucanoan (Barasano), Turkic (Azerbaijani), Uto-Aztecan (Pipil, Southern Paiute), non-Pama-Nyungan languages (Wardaman; Gooniyandi, Yawuru), Yanomaman (Sanumá)

Our main sources are a 2004 handout by Cysouw containing a great deal of cross-linguistic data, as well as Haspelmath (1997), which focuses on indefinite pronouns in a wide variety of genetically/areally unrelated language families.

1.3 Attested Syncretisms

1.3.1 The data

First off, there are languages with no syncretism at all, like Polish (Indo-European), Hungarian (Finno-Ugric) and Thai (Thai-Kadai, Smyth 2002. This is shown in table 1.1.

There are also languages with two OCs showing syncretism: in Lithuanian (Indo-European, Baltic, Haspelmath 1997: 327), *kàs* means

6

Table 1.1: No syncretisms

	THING	PERSON	PLACE	MANNER	AMOUNT	TIME
Polish	co	kto	gdzie	jak	ile	kiedy
Hungarian	mi	ki	hol	hogy(an)	hány	mikor
Thai	aray	khray	thîi nǎy	yaŋŋay	thâwrày	mûarày

either 'what' (THING) or 'who' (PERSON), indicating a THING/PERSON syncretism. The same syncretism is found in the interrogative/indefinite paradigm of Yawuru (Nyulnyulan, Western Australia, McGregor 2004: 128), Paumari (Arawá, Brazil, Chapman & Derbyshire 1990: 203-216), Terena (Arawakan, Eastlack 1968: 7-8), Bare (Arawakan, Aikhenvald 1995: 25, and Warekena (Arawakan, Aikhenvald 1998: 261, 325-326) (table 1.2).

Table 1.2: THING-PERSON syncretism

	THING	PERSON	PLACE	MANNER	AMOUNT	TIME
Lithuanian	kàs	kàs	kur	kaip	kiek	kada
Yawuru	yangki	yangki	jana	janala(-kaja)	nganyja	bana
Paumari	nahina	nahina	hana	niha	nihafori	nihaforija
Terena	kuti	kuti	na	?	na	namo
Bare	ne	ne	awati	ika	ikabe	ikabure
Warekena	iʃi	iʃi	da-	iʃ(i)alema	iperi	yumirehe

Next, there is crosslinguistic evidence that the categories PERSON and PLACE are closely related, as shown in Awa Pit (Barbacoan, Curnow 2006: 225) with full syncretism, and in Lakhota (Siouan, Rood & Taylor 1996: 451, 457; Ingham 2003: 51-53) with stem-syncretism (see table 1.3).

Table 1.3: PERSON-PLACE syncretism

	THING	PERSON	PLACE	MANNER	AMOUNT	TIME
Awa Pit	shi	mɨn	mɨn=	mizha	yawa	mizhaka mizhuta
Lakhota	táku	tú-wa	tu-ktél	tó-(khe)škhe	tó-na(keča) to-haN'yaN	tó-hạ (realized) to-hạl (unrealized)

Syncretism involving PLACE and MANNER is attested in Tukang Besi (Austronesian, Donohue 1999: 105) and Amuecha (Arawakan, Wise 1986: 573) (table 1.4).

Table 1.4: PLACE-MANNER syncretism

	THING	PERSON	PLACE	MANNER	AMOUNT	TIME
Tukang Besi	paira	ie'ei (ie')emai	'umpa	'umpa	sapaira 'how much' (price)	kehia (FUT) (d)ehia (PST)
Amuecha	es	eseša	ez	ez	?	?

There is a MANNER/AMOUNT syncretism found in Wardaman (Isolate, Western Australia, Merlan 1994: §7.5) (table 1.5).

Table 1.5: MANNER-AMOUNT syncretism

	THING	PERSON	PLACE	MANNER	AMOUNT	TIME
Wardaman	ngamanda	yinggiya	guda	gun.garr (-ma)	gun.garr (-ma)	nyangurlang

Finally, an AMOUNT/TIME 'when, what quantity' syncretism is instantiated by Hup (Nadahup, Brazil, Colombia, and Venezuela, Epps 2008: 778) (table 1.6).

Table 1.6: AMOUNT-TIME syncretism

	THING	PERSON	PLACE	MANNER	AMOUNT	TIME
Hup	hɨ́-n'ɨ̆h	ʔŭy	hɨ́t	hɨ́p	hɨ́-m'ǽ	hɨ́-m'ǽ

In Modern Greek (Indo-European, Hellenic, Roussou 2016) there is evidence for a PERSON/PLACE/MANNER syncretism in the locative interrogative *pú*, which can be interpreted as 'to whom', 'where', or 'how'. Roussou (2016: 6) writes: '(…) interrogative *pu* primarily has a locative reading, as [(12a)], but can be interpreted as a manner adverbial (from which x, did you understand it from x) [(12b)], or stand for an indirect object, as in [(12c)] (on the latter, see Michelioudakis 2012.'). This is shown in table 1.7.

(12) a. Pú   pas?
         where go-2SG
         'Where are you going?'
      b. Pú   to katalaves?
         where it understood.2SG
         'How did you understand this?'
      c. Pú to edhoses?
         where it gave.2SG
         'Who did you give it to?'   (Roussou 2016: (12))

Table 1.7: PERSON-PLACE-MANNER syncretism

	THING	PERSON	PLACE	MANNER	AMOUNT	TIME
Modern Greek	-ti	pu	pu	pu	póso	póte
				pos	'how much'	

Lakhota (Siouan, United States, Rood & Taylor 1996: 451, 457; Ingham 2003: 51-53) has stem syncretism of MANNER/AMOUNT/TIME, whereas Sanuma (Isolate, Brazil, Borgman 1990: 66-72) has full syncretism of these categories (table 1.8).

Table 1.8: MANNER-AMOUNT-TIME syncretism

	THING	PERSON	PLACE	MANNER	AMOUNT	TIME
Lakhota	táku	**tú**-wa	**tu**-ktél	**tó**-(khe)škhe	**tó**-na(keča) **to**-haN'yaN	**tó**-hạ (realized) **to**-hạ́l (unrealized)
Sanuma	?	witi	witi na	wi na	wi na	wi na

In addition to these two-cell syncretisms, some languages show widespread (even bordering on total) syncretism, as in Barasano (Tucanoan, Colombia, W. Jones & P. Jones 1991: 31), Perené Asheninca (Arawakan, Reed & Payne 1986: 330) and Asheninka Campa (Arawakan, Peru, Reed & Payne 1986: 328-329; Givón 2001: 304, citing D. Payne p.c.) (table 1.9).[3]

Table 1.9: Widespread syncretism

	THING	PERSON	PLACE	MANNER	AMOUNT	TIME
Barasano	yẽ	yĩb...	dõ	dõ	dõ	dõ
Perené Asheninka	paita	ninka	t^{sh}ika	t^{sh}ika	t^{sh}ika	t^{sh}ika
Asheninka	tsika	tsika	tsika	tsika	tsika	tsika-paita

We also predict that 'bare' ontological categories (generic/functional/light/dummy nouns) should also be able to show syncretism (see section 1.2.2.3). Actually, syncretism here is quite rare since these light nouns are usually grammaticalized from distinct lexical nouns meaning 'thing', 'person', 'place', etc. Nevertheless, Rapanui (Austronesian) is an exception with a PERSON/THING syncretism in *me'e*: 'Headless relatives are not found. Instead a dummy head is used: *me'e* 'thing or person', *hora*

---

[3]For Asheninka *tsika-paita*, Cysouw (2004: 2, (2e)) glosses *paita* as TEMP'.

'time', *kona* 'place', *aŋa* 'action" (Du Feu 1996: 47, cited in Cinque 2016: fn. 7).

1.3.2 Linear order

Again, syncretism reflects structural adjacency, revealing which syntactic heads are merged next to each other in the functional sequence. Some crucial syncretisms from the point of view of linear ordering have been put in table 1.10.

Table 1.10: Order according to the syncretisms

	THING	PERSON	PLACE	MANNER	AMOUNT	TIME
Lithuanian	kàs	kàs	kur	kaip	kiek kada	
Awa Pit	shi	mɨn	mɨn=	mizha	yawa	mizhaka mizhuta
Tukang Besi	paira	ie'ei	'umpa	'umpa	sapaira	kehia, dehia
Warademan	ngamanda	yinggiya	guda	gun.garr (-ma)	gun.garr (-ma)	nyangurlang
Hup	hǐ-n'ĭh	ʔǔy	hǐt	hǐp	hǐ-m'ǽ	hǐ-m'ǽ

The patterns observed require a linear order of heads such that the functional layer THING is next to PERSON, which is next to PLACE, which is next to MANNER, which is next to AMOUNT, which is next to TIME.

(13)   THING | PERSON | PLACE | MANNER | AMOUNT | TIME

(13) predicts when syncretisms are possible: two non-adjacent OCs in a paradigm should never be syncretic, i.e. THING and PLACE should never be syncretic to the exclusion of PERSON for instance. This is simply the *ABA theorem of Bobaljik (2007; 2012); Caha (2009), and others.

1.4 Morphological containment

The question we turn to now is which order in (14) is the correct one.

(14)   a.   THING > PERSON > PLACE > MANNER > AMOUNT > TIME
        b.   TIME > AMOUNT > MANNER > PLACE > PERSON > THING

As we will see in the next section, attested cases of morphological containment clearly indicate that (14b) is correct.

1.4.1 PERSON contains THING: PERSON > THING

In Muna (Austronesian, Indonesia, Van den Berg 1989: §8.6.2) and Amuecha (Arawakan, Wise 1986: 573), we have a clear case of THING being contained within PERSON (suggesting that PERSON is bigger than THING and thus higher up in the fseq).

(15)

	THING	PERSON
Muna	**hae**	la-**hae**
Amuecha	**es**	**es**eša

(16)   TIME > AMOUNT > MANNER > PLACE > **PERSON** > **THING**

1.4.2 PLACE contains PERSON: PLACE > PERSON

Second we see in Sanumá (Borgman 1990: 67,70) and Pipil (Campbell 1985: 114) that PLACE contains PERSON, suggesting that PLACE is bigger than PERSON.

(17)

	PERSON	PLACE
Sanumá	**witi**	**witi** ha
Pipil	**ka:**	**ka:**n

(18)   TIME > AMOUNT > MANNER > **PLACE** > **PERSON** > THING

1.4.3 MANNER contains PLACE: MANNER > PLACE

Danish (Indo-European), Muna (Van den Berg 1989: §8.6.2), and Southern Paiute (Uto-Aztecan, Sapir 1930: 209) illustrate cases where MANNER contains PLACE, suggesting that MANNER is bigger than PLACE.

(19)

	PLACE	MANNER
Danish	**hvor**	**hvor**dan
Muna	hamai	**peda hamai**
Southern Paiute	**my**	**my**ja

(20)   TIME > AMOUNT > **MANNER** > **PLACE** > PERSON > THING

1.4.4 AMOUNT contains MANNER: AMOUNT > MANNER

AMOUNT contains MANNER in Bare (Arawakan, Aikhenvald 1995: 25), German (Indo-European), and Gooiyandi (McGregor 2004: 128)). That is, AMOUNT is structurally bigger than MANNER.

(21)

	MANNER	AMOUNT
Bare	**ika**	**ika**be
German	**wie**	**wie**viel
Gooniyandi	**yiniga**	**yiniga**ngarri

(22)   TIME > **AMOUNT** > **MANNER** > PLACE > PERSON > THING

1.4.5   TIME contains AMOUNT: TIME > AMOUNT

AMOUNT is contained within TIME in Maybrat ((West) Papuan, Dol 1999: 118), Terena (Arawakan, Eastlack 1968: 7-8), Jaqaru (Aymaran, Hardman 2000: 33), and German.

(23)

	AMOUNT	TIME
Maybrat	**tiya**	ti**tiya**
Terena	**na**	**na**mo
Jaqaru	**ayka**	**ayka**p"a
German	**wieviel**	**wieviel** (Uhr)

(24)   **TIME** > **AMOUNT** > MANNER > PLACE > PERSON > THING

1.4.6   TIME > AMOUNT > MANNER

Paumari (Arawá, Brazil, Chapman & Derbyshire 1990: 203-216) shows a nice case of MANNER being contained within AMOUNT and TIME, and AMOUNT contained within TIME.

(25)

	MANNER	AMOUNT	TIME
Paumari	**niha**	**niha-fori**	**niha-fori**-ja

(26)   **TIME** > **AMOUNT** > **MANNER** > PLACE > PERSON > THING

1.4.7   Morphological containment is not constrained by adjacency

The nanosyntactic account of syncretism relies on syncretism being constrained by adjacency. Morphological containment, on the other hand, is not constrained by adjacency. As shown in Azerbaijani (Turkic, see Cysouw 2004), both AMOUNT and TIME contain THING, but MANNER, PLACE and PERSON do not.

(27)

	THING	…	AMOUNT	TIME
Azerbaijani	**nə**	…	**nə** kədər	**nə** vaxt

(28)   **TIME** > **AMOUNT** > MANNER > PLACE > PERSON > **THING**

Similar cases are exemplified here with MANNER containing THING in Dumi (Sino-Tibetan, Van Driem 1993) and Yimas (Foley 1991: 114-115), and PLACE containing THING in Pirahã Everett (1986: 239-245) and Greenlandic (Sadock 1984).

(29)

	THING	...	MANNER
Dumi	**mwo**	...	**mwi** :ho
Yimas	**wara**	...	**wara**tnti, **wara**wal

(30)   TIME > AMOUNT > **MANNER** > PLACE > PERSON > **THING**

(31)

	THING	...	PLACE
Pirahã	**gó**	...	**gó**ó
Greenlandic	**su-**	...	**su**mi

(32)   TIME > AMOUNT > MANNER > **PLACE** > PERSON > **THING**

In Danish (Indo-European), moreover, PLACE is contained within AMOUNT and TIME.

(33)

	PLACE	...	AMOUNT
Danish	**hvor**	...	**hvor** meget

(34)   TIME > **AMOUNT** > MANNER > **PLACE** > PERSON > THING

(35)

	PLACE	...	TIME
Danish	**hvor**	...	**hvor**når

(36)   **TIME** > AMOUNT > MANNER > **PLACE** > PERSON > THING

TIME contains THING in Imbabura Quechua (Quechuan, Cole 1982: 16-20), and it contains MANNER in Bare (Aikhenvald 1995: 25) and Awa Pit Curnow (2006: 225):

(37)

	THING	...	TIME
Imbabura Quechua	**ima**	...	**ima** ura

(38)   **TIME** > AMOUNT > MANNER > PLACE > PERSON > **THING**

(39)

	MANNER	...	TIME
Bare	**ika**	...	**ika**bure
Awa Pit	**mizha**	...	**mizha**ka

(40)   **TIME** > AMOUNT > **MANNER** > PLACE > PERSON > THING

Finally, Muna (Austronesian, Indonesia, Van den Berg 1989: §8.6.2) and Warekena (Awarakan, Aikhenvald 1998: 261, 325-326) are cases where THING is contained within PERSON, MANNER, and AMOUNT, but not within PLACE (table 1.11).

13

Table 1.11: THING contained in PERSON, MANNER, AMOUNT, TIME

	THING	PERSON	PLACE	MANNER	AMOUNT	TIME
Muna	**hae**	la-**hae**	hamai	peda **hae** peda hamai (peda 'like')	se-**hae**	nefiemo (PST) indefie (PST) naefie (FUT)
Warekena	**iʃi**	**iʃi**	da-	**iʃ(i)alema**	iperi	yumirehe

(41)   TIME > **AMOUNT** > **MANNER** > PLACE > **PERSON** > **THING**

Our empirical generalisations with regard to both syncretism and containment are captured by the functional sequence in (42), made up of (at least) six 'OC' heads.

(42)   [$_{\text{TIME}}$ OC6 [$_{\text{AMOUNT}}$ OC5 [$_{\text{MANNER}}$ OC4 [$_{\text{PLACE}}$ OC3 [$_{\text{PERSON}}$ OC2 [$_{\text{THING}}$ OC1]]]]]]

We take (42) to be universal. The fseq captures the possible syncretisms in terms of adjacency of functional layers, while also straightforwardly capturing the attested containment relations (of which there are various kinds attested, but all of them consistent with the hierarchy given here).

1.5   Conclusion

In this study we have shown that ontological categories can be ordered in a nanosyntactic fseq, with clear generalizations to be had regarding both syncretism and morphological containment. The facts are captured in a single fseq that we take to be universal (see (42) above).

Before closing, we would like to point out that the ontological categories uncovered here have to be thought of as very small bits of structure, compared to larger lexical nouns like English *thing, person, place* (etc.). In Baunaz & Lander (2018) we actually propose that *what* can be decomposed into at least two morphemes: *wh-at*, with *-at* corresponding to what we there call the nominal core, a semantically bleached, non-referential functional element which can be found in certain nominal environments. Nominal cores, moreover, come in different flavors (e.g. *n*FORM, *n*BODY, *n*THING, *n*PLACE, etc.) and must be distinguished from lexical nouns (in that cores are invariable, while lexical nouns are not). If this is correct, then THING isn't actually silent in the English example in (2) at all, but rather is overtly realized in *wh-**at*** (see Baunaz & Lander's 2018 *n*THING). Semi-lexical nouns like *-thing* or *-body* in *nothing, somebody*, etc., furthermore, are slightly bigger than *-at* (but not as big as a full lexical noun). It seems likely to us that the fine-grained difference

in structural size between ontological category *-at* (THING), semi-lexical *-thing,* and fully lexical *thing* might be similar to the difference between light verb *v* (e.g. GET) as in *I'm getting sleepy* and the main verb *get* meaning 'receive' as in *I'm getting a motorcycle for my birthday.*[4]

As a final note, we would like to point out that our results may be considered counterintuitive in some respects. While we leave aside, for now, our ideas about the semantic import of each of the individual heads in our fseq, we would like to make a more general point about expectations, that is, what we expect to find when doing research or thinking about language. As one reviewer for a conference commented on our work: 'Do we expect to find, for example, words for 'how' built off of a word for 'who'? I bet we won't find that.' This is a perfect example of why we should not always follow our common-sense intuitions when mapping out functional structure, because in fact Serbo-Croatian shows exactly this kind of containment: *(t)ko* 'who' is contained within *kako* 'how, in what way' (we can see the same thing in *ni-ko* 'no one' and *ni-kako* 'by no means'). In other words:

(43)   [ko] = PERSON
         [ka- [ko]] = MANNER

Sometimes it is important to put our intuitions aside and follow the data.

References

Aikhenvald, Alexandra. 1995. *Bare.* Vol. 100 (Languages of the World/Materials). Munich: Lincom Europa.

Aikhenvald, Alexandra. 1998. Warekena. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian languages,* vol. 4, 225–440. Berlin: Mouton de Gruyter.

Baunaz, Lena & Eric Lander. 2018. Deconstructing categories syncretic with the nominal complementizer. *Glossa: a journal of general linguistics* 3(1). 31. 1–27.

Bobaljik, Jonathan. 2007. On comparative suppletion. Ms. U of Connecticut.

Bobaljik, Jonathan. 2012. *Universals in comparative morphology*. Cambridge, MA: MIT Press.

Borgman, Donald. 1990. Sanuma. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian languages,* vol. 2, 15–248. Berlin: Mouton de Gruyter.

---

[4]We would like to thank a reviewer for comments on this topic.

Caha, Pavel. 2009. *The nanosyntax of case*. Tromsø: University of Tromsø dissertation.

Campbell, Lyle. 1985. *The Pipil language of El Salvador*. Vol. 1 (Mouton Grammar Library). Berlin: Mouton de Gruyter.

Chapman, Shirley & Desmond Derbyshire. 1990. Paumari. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian languages*, vol. 3, 161–354. Berlin: Mouton de Gruyter.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: a cross-linguistic perspective*. Oxford: Oxford University Press.

Cinque, Guglielmo. 2008. More on the indefinite character of the head of restrictive relatives. *Rivista di Grammatical Generativa* 33. 3–24.

Cinque, Guglielmo. 2010. *The syntax of adjectives: a comparative study*. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 2016. On the double-headed analysis of 'headless' relative clauses. Ms. University of Venice.

Cole, Peter. 1982. *Imbabura Quechua*. Vol. 5 (Lingua Descriptive Studies). Amsterdam: North Holland Publishing Company.

Curnow, Timothy. 2006. La interrogación y la negación en awa pit. *Amerindia* 29/30. 219–234.

Cysouw, Michael. 2004. Interrogative words: an exercise in lexical typology. Presentation at 'Bantu grammar: description and theory workshop 3', ZAS Berlin.

Diessel, Holger. 2003. The relationship between demonstratives and interrogatives. *Studies in Language* 27(3). 635–655.

Dol, Philomena Hedwig. 1999. *A grammar of Maybrat: a language of the bird's head, Irian Jaya, Indonesia*. Leiden: Universiteit Leiden dissertation.

Donohue, Mark. 1999. *A grammar of Tukang Besi*. Berlin: Mouton de Gruyter.

Du Feu, Veronica. 1996. *Rapa nui*. London: Routledge.

Eastlack, Charles. 1968. Terena (Arawakan) pronouns. *International Journal of American Linguistics* 34(1). 1–8.

Epps, Patience. 2008. *A grammar of Hup*. Vol. 43 (Mouton Grammar Library). Berlin: Mouton de Gruyter.

Everett, Daniel. 1986. Pirahã. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian languages*, vol. 1, 200–325. Berlin: Mouton de Gruyter.

Foley, Willam A. 1991. *The Yimas language of New Guinea*. Stanford, CA: Stanford University Press.

Givón, Talmy. 2001. *Syntax: an introduction*. Vol. 2. Amsterdam: Benjamins.

Haegeman, Liliane. 2014. West Flemish verb-based discourse markers and the articulation of the speech act layer. *Studia Linguistica* 68(1). 116–139.

Haegeman, Liliane & Virginia Hill. 2013. The syntacticization of discourse. In Raffaella Folli, Christina Sevdali & Robert Truswell (eds.), *Syntax and its limits*, 370–389. Oxford University Press.

Hardman, Martha. 2000. *Jaqaru*. Vol. 183 (Languages of the World/Materials). Munich: Lincom Europa.

Haspelmath, Martin. 1997. *Indefinite pronouns* (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.

Ingham, Bruce. 2003. *Lakota*. Vol. 426 (Languages of the World/Materials). Munich: Lincom Europa.

Jones, Wendell & Paula Jones. 1991. *Barasano syntax*. Vol. 2 (Studies in the Languages of Colombia). Arlington, TX: Summer Institute of Linguistics.

Kayne, Richard. 2005. *Movement and silence*. Oxford: Oxford University Press.

McGregor, William. 2004. *The languages of the Kimberley, Western Australia*. London: Routledge.

Merlan, Francesca C. 1994. *A grammar of Wardaman: a language of the Northern territory of Australia*. Vol. 11 (Mouton Grammar Library). Berlin: Mouton de Gruyter.

Michelioudakis, Dimitris. 2012. *Dative arguments and abstract case in Greek*. University of Cambridge dissertation.

Noonan, Michael. 1992. *A grammar of Lango*. Vol. 7 (Mouton Grammar Library). Berlin: Mouton de Gruyter.

Reed, Judy & David L. Payne. 1986. Asheninca (Campa) pronominals. In Ursula Wiesemann (ed.), *Pronominal systems*, vol. 5 (Continuum), 323–331. Tübingen: Narr.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar. Handbook in generative syntax*, 281–337. Dordrecht: Kluwer.

Rood, David & Allan Taylor. 1996. Sketch of Lakhota, a Siouan language. In Ives Goddard (ed.), *Handbook of North American indians*, vol. 17 (Languages), 440–482. Washington, DC: Smithsonian Institution.

Roussou, Anna. 2016. Complement clauses: case and argumenthood. Ms. University of Patras.

Sadock, Jerrold. 1984. West Greenlandic. In William S. Chisholm (ed.), *Interrogativity: a colloquium on the grammar, typology and pragmatics of questions in seven diverse languages*, vol. 4 (Typological Studies in Language), 189–214. John Benjamins.

Sapir, Edward. 1930. Southern Paiute, a Shoshonean language. In William Bright (ed.), *Southern Paiute and Ute linguistics and ethnography*, vol. 10 (Collected works of Edward Sapir). Berlin: Mouton de Gruyter.

Smyth, David. 2002. *Thai: an essential grammar*. London: Routledge.

Starke, Michal. 2009. Nanosyntax: a short primer to a new approach to language. *Nordlyd* 36. 1–6.

Starke, Michal. 2011. Towards an elegant solution to language variation: variation reduces to the size of lexically stored trees. Ms., Tromsø University.

Van den Berg, René. 1989. *A grammar of the Muna language*. Dordrecht: Foris.

Van Driem, George. 1993. *A grammar of Dumi*. Vol. 10 (Mouton Grammar Library). Berlin: Mouton de Gruyter.

Vangsnes, Øystein. 2013. Syncretism and functional expansion in Germanic wh-expressions. In Muriel Norde, Alexander Lenz & Karin Beijering (eds.), *Language sciences: special issue on current trends in grammaticalization research*, vol. 36, 47–65. Dordrecht: Elsevier.

Wise, Mary Ruth. 1986. Grammatical characteristics of PreAndine Arawakan languages of Peru. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian languages*, vol. 1, 567–642. Berlin: Mouton de Gruyter.

2 Syncretism as Merge F: the Nanosyntax of case ten years on

Pavel Caha

Masarykova Univerzita, Brno

2.1 Introduction

In this paper, I summarise the research program followed in Caha (2009), along with its main hypotheses and results. I further review some of the problems pointed out in the subsequent literature concerning the syncretism of datives (Harðarson 2016) and present a solution to these problems proposed in Starke (2017). The solution consists in enriching the original case hierarchy with a new type of a dative and a new type of an accusative case. Starting from there, I note a curious possibility made available by Starke's solution, which allows for the existence of a language with a surface case hierarchy NOM—GEN—ACC—DAT—etc. I argue that this surface hierarchy is needed in order to deal with syncretism and containment in several Saami languages.

2.2 Syncretism as Merge F

The idea that syntactic structures are built incrementally by putting smaller pieces together is one of the fundamental properties of syntactic derivations (Chomsky 1995). In the traditional (well-trodden) cases, the addition of a head F is always reflected by the addition of a visible morpheme, as in (1).

(1)
```
        FP
       ╱ ╲
      F   XP
      |   ╱ ╲
      α   ...
```
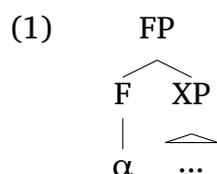
α in such structures could easily be taken for a determiner or a preposi-

tion being added to a noun phrase; or for an infinitival marker added to a bare verb phrase, etc.

In Nanosyntax (Starke 2009), the very same operation Merge F can be reflected in at least two additional ways: by syncretism and by override. The crucial ingredient in obtaining this result is phrasal spell out, where a potential match between the syntactic tree and the lexical item is determined by the Superset Principle, given in (2).

(2)   *The Superset Principle* (Starke 2009):
    A lexically stored tree matches a syntactic node iff the lexically stored tree contains the syntactic node.

To see how phrasal spell out increases the number of ways in which Merge F is reflected on the surface, suppose that a language has the lexical entry given in (3):

(3)   FP $\Leftrightarrow$ /α/
    F   XP
        X Y

When syntax builds an XP, the XP can be spelled out by (3), since XP is contained in it.

(4)   XP $\Rightarrow$ /α/
    X Y

After merge F applies to XP, yielding FP, we can still spell out by α, since FP is contained in (3) as well.

(5)   FP $\Rightarrow$ /α/
    F   XP
        X Y

The result is that two different structures—XP and FP (immediately containing XP)—are pronounced the same, a situation which I will refer to as 'syncretism' (i.e., between XP and FP).

It is fair to say that traditional models are also able to incorporate the fact that the output of Merge F is pronounced the same as the input to this operation through the use of zero spell out of F (an option systematically explored and exploited in recent work by Richard Kayne). What the standard model would have trouble capturing is if Merge F was reflected simply by changing the previous spell out completely, something which

became an option under the phrasal spell out model. So suppose that a language has the following two lexical entries:

(6)　　XP $\Leftrightarrow$ /β/
　　　　△
　　　X　Y

(7)　　FP　　$\Leftrightarrow$ /γ/
　　　　△
　　　F　XP
　　　　　△
　　　　X　Y

When syntax builds XP, both β and γ are candidates for spell out (they both contain XP). But β wins because it is a better match:

(8)　　XP $\Rightarrow$ /β/
　　　　△
　　　X　Y

When Merge F applies, spell out targets FP, inserting γ and overriding β:

(9)　　FP　　$\Rightarrow$ /γ/
　　　　△
　　　F　XP
　　　　　△
　　　　X　Y

The three different ways in which the surface form may reflect the operation Merge F are summarised in (10). On the first line, we have an XP spelled out uniformly as α. The second row gives the three options of how Merging F to XP may be reflected on the surface.

(10)

The three outputs of merge F			
XP	α	α	α
Merge F to XP	α + β	α	β
	stacking	syncretism	override

An idea explored in Caha (2009) is that all aspects of case paradigms can be reduced to the three scenarios in (10), so that ultimately paradigms become just a by-product of the fundamental operation Merge F. Under this view, Merge F creates individual cases by merging together privative case features, such that, for instance, XP corresponds to NOM, FP to ACC etc., with each new case formed by adding a feature to the previous one in a hierarchy of cases.

　　Additional differences among languages come about as a result of positioning the noun with respect to the markers in ways outlined in Cinque (2005). For instance, the stacking scenario can (among others) yield a prepositional marker β governing the case α, if the complement of F does not move, as in (19). But stacking can also yield a complex case suffix (when XP moves across F). See Caha (2009; 2011) for details.

21

2.3 *ABA and the Case hierarchy

One of the predictions of the model is related to its inability to generate the so-called *ABA pattern (Bobaljik 2007, Starke 2009). 'ABA' refers to a type of an overriding pattern, given in (11), where β first overrides α only to be overriden by the same α after the next step of Merge. The '*' before the ABA means that this pattern is impossible to derive, and hence that it should be unattested, accidental homophony aside.

(11)

*The *ABA*	
XP	α
Merge F to XP ⇒ FP	β
Merge F2 to FP ⇒ F2P	α

The reasoning behind the restriction is well known, but for clarity, let me give it here again. In order for α to spell out F2P, as it does in the last row of (11), its entry must be as in (12a). The entry of β, which is in the second row, spells out FP only, so its entry is as in (12b). In fact, both β and α match FP, but β wins over α because it is a better match (it has fewer superfluous features). Now when it comes to the first row of the table (11), the spell out of XP, β must win over α again, yielding a full paradigm of the shape α (for F2P), β (for FP), β (for XP). The ABA pattern is underivable.

(12)   a.   F2P   ⇔ /α/   b.   FP   ⇔ /β/

$$
\begin{array}{ccc}
& \text{F2P} & \Leftrightarrow /\alpha/ \\
\text{F2} & & \text{FP} \\
& \text{F} & \text{XP} \\
& & \text{X} \quad \text{Y}
\end{array}
\qquad
\begin{array}{cc}
\text{FP} & \Leftrightarrow /\beta/ \\
\text{F} & \text{XP} \\
& \text{X} \quad \text{Y}
\end{array}
$$
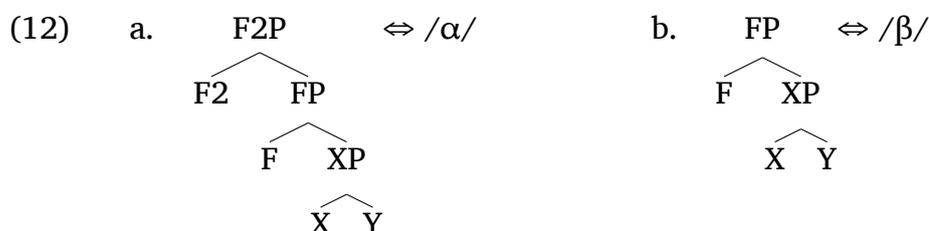
Caha (2009) combined the *ABA prediction with a strongly Cartographic position (Cinque & Rizzi 2008) which admits no variation in the kind and order of Fs that various languages have at their disposal. If that is so, and every language indeed merges the same Fs in exactly the same order, then case paradigms of all languages unfold in a uniform fashion (from XP to FP, then to F2P, and so on), differing only in how Merge F is reflected at each step. When Merge F is reflected by syncretism, then XP can be syncretic with FP, FP with F2P, and so on. Now since XP, FP, F2P etc. correspond to individual cases (understood abstractly through their syntactic/semantic function), the model predicts that all languages will show exactly the same *ABA constraint on case syncretism.

Drawing on previous research (see in particular McCreight & Chvany

1991, Plank 1991 and Johnston 1996), Caha (2009) proposed the generalisation/hypothesis in (13):[1]

(13)   Universal (Case) Contiguity (Caha 2009):
    a.   Non-accidental case syncretism targets contiguous regions in a sequence invariant across languages.
    b.   The Case sequence: NOM—ACC—GEN—DAT—INS—COM

The first part of the Case sequence (13b) is based on an observation by Baerman, Brown & Corbett (2005), who found in their 200 language sample that first of all, the syncretism of core cases is common, and that second of all, if one of the core cases is syncretic with an oblique, it is (as a strong tendency) the marked core case (i.e., ACC in NOM/ACC languages). This finding provides a solid background to the claim that there is indeed a cross-linguistically stable order of cases with NOM at one end, followed by ACC and then the oblique cases.

The general consensus in the subsequent literature is that this part of the generalisation should indeed be maintained (Harðarson 2016, Zompì 2017, McFadden 2018, Smith et al. 2018). If correct, then this part of the proposal is a success story for the idea that case paradigms at their core are nothing else but a specific manifestation of Merge F, and moreover, that there is no variation in the inventory and ordering of Fs across various languages.

Another stable part of the hierarchy (as far as I am aware) is the peripheral location of INS and COM, which rarely cause any trouble (though see Zompì 2017 for a potential problem in Latin). The order of the remaining two obliques (GEN and DAT) has, however, proven to be a rather difficult topic (Harðarson 2016, Starke 2017, Van Baal & Don 2018). The goal of this paper is to contribute to this area by a couple of new observations from Ugro-Finnic languages.

The reason for proposing the ordering ACC—GEN—DAT in (13b) is the type of pattern found in Russian (as well as other languages), where the ACC is syncretic with GEN, as shown in table 2.1.[2]

Under the hypothesis that syncretism reflects merge F, NOM and ACC must be related by this operation (14a), because they are syncretic in the paradigm 'window'. Further, since ACC is the same as GEN in the paradigm 'teacher,' GEN must be related to ACC via Merge F (14b). DAT

---

[1]On purpose, the statement leaves out spatial cases which rise complications that need a separate treatment (see Caha 2017).

[2]As a general strategy here and also in what follows, I only give paradigm fragments that illustrate efficiently and non-redundantly the paradigm structure of a given language. It goes without saying that as far as I am aware, there are no additional syncretisms of the relevant cases beyond what is displayed, and to the extent that there are, they have been argued to be accidental in the references given.

Table 2.1: Syncretism in Russian (McCreight & Chvany 1991)

	window, SG.	teacher, PL.	both, M.IN.	book, SG.	100
NOM	okn-o	učitel-ja	dv-a	knig-a	st-o
ACC	okn-o	učitel-ej	dv-a	knig-u	st-o
GEN	okn-a	učitel-ej	dv-ux	knig-y	st-a
PREP	okn-e	učitel-jax	dv-ux	knig-e	st-a
DAT	okn-u	učitel-am	dv-um	knig-e	st-a
INS	okn-om	učitel-ami	dv-umja	knig-oj	st-a

must be placed outside of this constituent, as shown in (14c).

(14)   a.   ACC
            ⌒
          F   NOM
              ⌒
              ...

       b.   GEN
            ⌒
          F2   ACC
              ⌒
            F   NOM
                ⌒
                ...

       c.   DAT
              ⌒
          F3   GEN
               ⌒
             F2   ACC
                 ⌒
               F   NOM
                   ⌒
                   ...
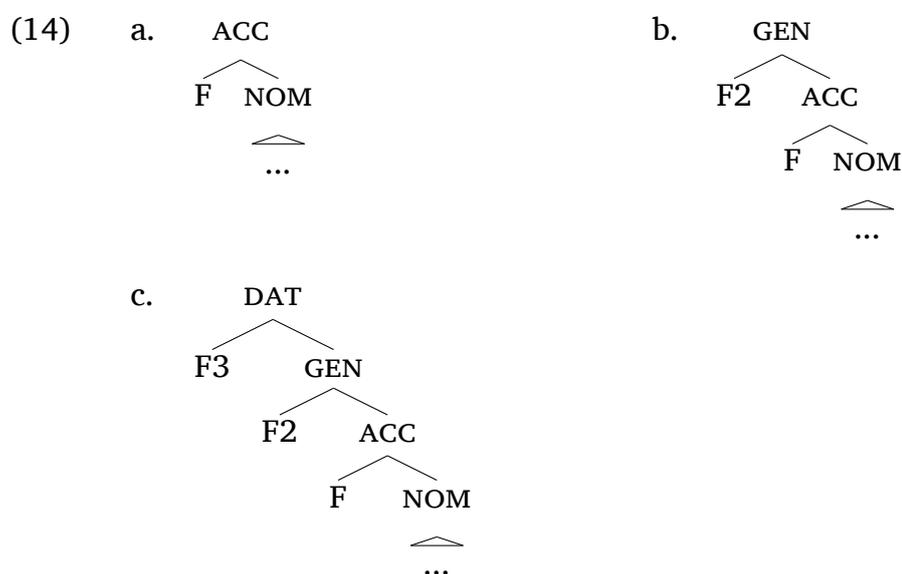
I am ignoring here the other cases on purpose, since the focus is on the four core cases depicted in (14c).

2.4   Datives in the way

However, as Harðarson (2016) points out, Icelandic provides similar type of evidence for the contradictory conclusion that DAT must be placed in between the ACC and GEN. Table 2.2 shows that when DAT is right after ACC, syncretism respects the *ABA.

As in Russian, NOM and ACC must be related by Merge F (15a), because they are syncretic in the paradigm 'land'. However, since DAT is the same as ACC in the paradigm 'queen,' it is the DAT case which must be related to ACC via Merge F (15b). GEN must be placed outside of this constituent (15c).

Table 2.2: Syncretism in Icelandic (Harðarson 2016)

	'arm'	'land'	'queen'	'tongue'
NOM	arm-ur	land-Ø	drottning-Ø	tung-a
ACC	arm-Ø	land-Ø	drottning-u	tung-u
DAT	arm-i	land-i	drottning-u	tung-u
GEN	arm-s	land-s	drottning-ar	tung-u

(15) a.
```
        ACC
       /  \
      F   NOM
           ⌒
          ...
```
b.
```
        DAT
       /   \
     F2    ACC
          /   \
         F    NOM
               ⌒
              ...
```
c.
```
         GEN
        /   \
      F3    DAT
           /   \
         F2    ACC
              /   \
             F    NOM
                   ⌒
                  ...
```
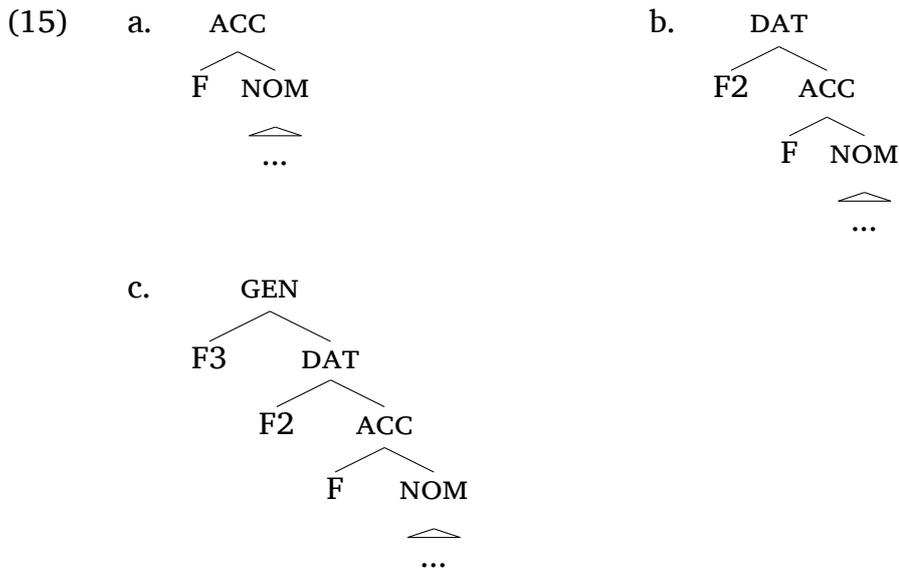
Harðarson (2016) mounts supporting evidence for the alternative hierarchy (15c) (with GEN larger than DAT) by noting that in Faroese, the possessor (i.e., the abstract genitive) requires a preposition coming on top of the dative. Within the confines of Caha's theory, such examples must be analysed as a scenario where GEN is derived from DAT by Merge F, reflected by stacking. This fact supports (15c), and seems impossible to reconcile with Caha's (14c).

(16)   Her eru húsini   hjá einum ríkum   manni.
       here are houses.the with a.DAT rich.DAT man.DAT
       'Here is a rich man's house.' (Faroese, Harðarson 2016)

In sum, the Icelandic and Faroese facts can still be derived under the assumption that stacking (16) and syncretism (table 2.2) reflect Merge F, but only if the order of cases is NOM—ACC—DAT—GEN, which is different from the order NOM—ACC—GEN—DAT required for Russian. A potential conclusion one could draw here is that the order of Fs is not fixed across languages, but varies from language to language.

2.5 Case in Skolt Saami

A similar set of problems is present in various Saami languages, here represented by Skolt Saami in table 2.3 (Feist 2010):[3]

Table 2.3: Skolt Saami (Feist 2010: 145)

	1ST.PL	hole, SG.	hole, PL
NOM	mij	kåå'pp	kåå'v
GEN	mij	kåå'v	koo'v-i
ACC	mi'jjid	kåå'v	koo'v-i-d
DAT	mi'jjid	kåpp-a	koo'v-i-d

In order to derive the patterns in table 2.3, while maintaining the premise that syncretism reflects Merge F, we must operate with a hierarchy where GEN is derived directly from NOM, because this is the only way in which we can allow for the NOM–GEN syncretism in the paradigm of the first person plural. (Other plural pronouns show the same syncretism.) In order to derive the remaining two syncretisms, GEN must be followed by ACC and DAT in that order. The fact that ACC must follow GEN is strengthened by the observation that the ACC is in fact derived from GEN by stacking in the plural paradigm of 'hole.'

Skolt Saami thus presents us with a similar puzzle as Icelandic: the facts do point to a particular linear order of cases, but this order seems to be highly variable across languages. Is there no end to variation?

2.6 Two datives

In a larger perspective, these puzzles are somewhat reminiscent of the ones explored in the seminal work by Cinque (1999). As is well know, Cinque proposes that adverbs are rigidly ordered and occupy fixed positions in the functional sequence. At one point, however, Cinque comes across what "seems to [be] a paradox, since we have postulated both

---

[3]I am labelling the case which is traditionally called *illative* as a *dative* in the table, since the case has a recipient function in a double object construction:

(i) jeä'nn uu'di leei'bid kooum päärnže.
     mother give.PAST bread.ACC three.ILL child.ILL
     'The mother gave the bread to three children.' (Skolt Saami, Feist 2010: 224)

The ability to express the recipient in a double object construction is a defining characteristic of a dative, so I follow this labelling for consistency with other languages. For a discussion of the DAT-ILL syncretism, see Caha (2017).

the structure [...] with *twice* higher than *intentionally* and the structure [...] with *twice* lower than *intentionally*. [...] The paradox, however, is not real, as there is evidence that *twice* belongs to a class of adverbs [...] that are systematically ambiguous between two interpretations, each associated with a different position."

It turns out that a solution along the same lines can be fruitfully pursued also in the current case, a proposal investigated for the Icelandic puzzle in Starke (2017). Specifically, Starke claims that the Russian recipient and the Icelandic recipient each correspond to a different type of dative. He calls the Russian dative for BIG DAT and the Icelandic dative for SMALL DAT. The complete functional sequence then contains both of these datives, one bigger than GEN and the other smaller than GEN, as in (17).

(17)

```
        BIG DAT
       /     \
     F4      GEN
            /    \
          F3   SMALL DAT
               /      \
             F2       ACC
                     /    \
                    F     NOM
                          /\
                         ...
```
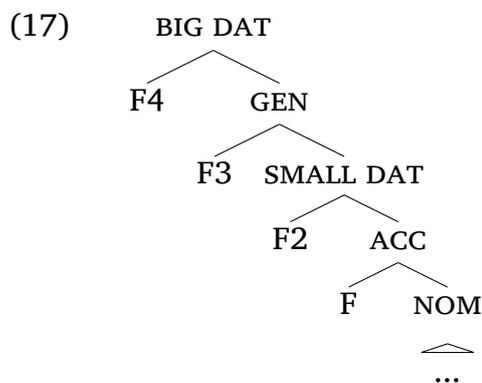
The fact that each of the two datives is found on the opposite side of the genitive leads to the surface observation that in Russian, datives are bigger than GEN (because they correspond to the BIG DAT), while in Icelandic they are smaller than GEN (because they correspond to the SMALL DAT).

The clearest evidence for postulating two datives comes from languages that actually show two different realisations of the recipient, each associated with a distinct syntactic position and distinct morphological realisation. This is the case in the so-called dative-shift alternation:

(18) a. Sally gave a toy to Justine.
      b. Sally gave Justine a toy.

Following Starke, *to Justine* corresponds to the BIG DAT, and the bare DP to the SMALL DAT. English further shows that the SMALL DAT must be smaller than GEN, because in English (similarly to Icelandic), the SMALL DAT (*Justine/him*) is syncretic with ACC and different from GEN.

That the BIG DAT of the dative-shift alternation should come on the other side of GEN, can be seen nicely in Arabic:

(19)   *Arabic* (Ryding 2011: 290-1)
        a.   a'ṭay-tu l-miftāḥ-a   li-l-bint-i
              I.gave   the-key-ACC to-the-girl-GEN
              'I gave the key to the girl.'
        b.   a'ṭay-tu l-bint-a   l-miftāḥ-a
              I.gave   the-girl-ACC the-key-ACC
              'I gave the girl the key.'

In (19b) we can see the SMALL DAT version of the recipient, which is identical to ACC (like in English). (19a) shows that the BIG DAT version is formed by the stacking of *li* on top of GEN. Since the BIG DAT is derived by stacking from GEN, it must contain GEN. In sum, the dative-shift alternation shows that we independently need two different datives, each with a distinct structure, and each in a different syntactic position.

Now consider the fact that neither Icelandic nor Russian has the dative alternation with the verb *give*. The languages thus allow only for a single frame, featuring a noun that is marked by a case that both descriptive traditions label as DAT. Which of the datives do the respective languages use? A logical possibility, suggested by Starke, is that Icelandic uses the SMALL DAT (because dative shift is obligatory), while Russian exhibits the BIG DAT (because dative shift is impossible). The consequence is that the Russian dative and the Icelandic dative are (structurally speaking) different cases, and that is why they are each at a different side of the genitive.

An independent piece of support for this conclusion comes from nominalisations. In Russian, when a verb takes a dative argument (20a), that argument must retain its dative in the nominalisation (20b).

(20)   *Russian* (Zimmermann 2002: 280)
        a.   izmenit'   žen-e
              betray.INF wife-DAT
              'to betray the wife'
        b.   izmena   žen-e
              betrayal wife-DAT
              'the betrayel of the wife'

In Icelandic (Maling 2001), when a verb takes a dative argument (21a), this argument turns genitive in nominalisation (21b).

(21)   *Icelandic* (Harðarson 2016)
        a.   Astrid bjargaði skinkunni.
              Astrid rescued ham.DEF.DAT
              'Astrid rescued the ham.'

b. björgun skinkunnar
rescue ham.DEF.GEN
'the rescue of the ham'

In traditional terms, this shows that the Icelandic dative counts as a structural case, while the Russian dative behaves like an oblique case. In the current theory, the facts indicate that all cases which are smaller than the genitive (including NOM, ACC and SMALL DAT) turn to genitive in nominalisations, while cases larger than GEN (including BIG DAT) don't.

To sum up: the solution to the puzzle pointed out by Harðarson (2016) is that there are two distinct datives, one larger than GEN (in Russian), and one smaller than GEN (in Icelandic). The solution is simple in the sense that we independently need two datives, and we independently need to acknowledge variation in the application of the dative shift. A simple combination of these two ingredients solves the *ABA puzzle, and correlates neatly with the patterning of the different datives in nominalisation. But now there is a larger issue to be addressed: how many cases lead such a double life, and what consequences does this bring?

2.7 Differential object marking

In discussing these issues, Starke (2017) suggests that also Differential Object Marking (DOM) should be treated in terms of two direct object cases, BIG ACC and SMALL ACC. An example of a DOM is shown in (22). It consists in the differential treatment of direct objects depending on notions such as specificity and animacy (Aissen 2003).

(22)   *Spanish* (Starke 2017)

a. María quiere **a**   un abogado [+anim, +specific]
Mary wants PREP a lawyer
'Mary wants a (specific) lawyer.'
b. María quiere un abogado [+anim, -specific]
Mary wants a lawyer
'Mary wants a lawyer (any lawyer).'

The reasoning is simple: since a single noun phrase ('a lawyer') can be marked in two different ways, we must acknowledge the existence of two different cases, BIG ACC and SMALL ACC. Although BIG ACC is frequently syncretic with DAT, as it is in Spanish (c.f. Manzini & Franco 2016), it cannot be fully reduced to DAT, since it can show a special marking of its own, as in Turkish. (23a,b) introduce the basics of DOM, (23c) shows that BIG ACC is different from DAT.

(23)   *Turkish* (Aissen 2003)

    a.   Ali bir kitab-ɨ   aldɨ.
        Ali one book-B.ACC bought
        'Ali bought the book.'

    b.   Ali bir kitap-Ø   aldɨ.
        Ali one book-S.ACC bought
        'Ali bought some book.'

    c.   Ben çocuğ-a şeker-ɨ   verdɨm.
   I   child-DAT candy-B.ACC gave
        'I gave the child the candy.'

BIG ACC must therefore be a self-standing case in the neighbourhood of BIG DAT. The question of whether it should be lower or higher than BIG DAT can be answered by looking at the third attested pattern of DOM, represented by Ossetic, where BIG ACC is syncretic with GEN (and different from DAT). (24a,b) give the DOM pair with SMALL ACC and BIG ACC respectively. (24c) shows that BIG ACC is the same as GEN and distinct from DAT:

(24)   *Ossetic* (Belyaev 2010)

    a.   miron fəš   argăfšt-a
        Miron ram.S.ACC slaughter-PAST
        'Miron has slaughtered a ram.'

    b.   lăpːu   čəžǯ-ə   fedt-a
        boy-NOM girl-B.ACC see-PAST
        'The boy saw the girl.'

    c.   mă-mad-ə   mad-ăn
        my-mother-GEN mother-DAT
        'to my mother's mother'

The data is summarised in the following table, where the differentially marked object (BIG ACC) is shaded, and any potential syncretism is highlighted by extending the shading to the relevant cell:

(25)   *The syncretisms of the BIG ACC*

	GEN	BIG ACC	BIG DAT
Spanish	A	B	B
Turkish	A	B	C
Ossetic	A	A	B

Considered together, the facts require that BIG ACC must be located between GEN and Caha's original DAT, which has now changed label to BIG DAT. The full structure is in (26).

(26)

```
                BIG DAT
               /      \
            F5       BIG ACC
                    /      \
                  F4       GEN
                          /    \
                        F3    SMALL DAT
                              /       \
                            F2      SMALL ACC
                                    /       \
                                   F       NOM
                                          /
                                         ...
```
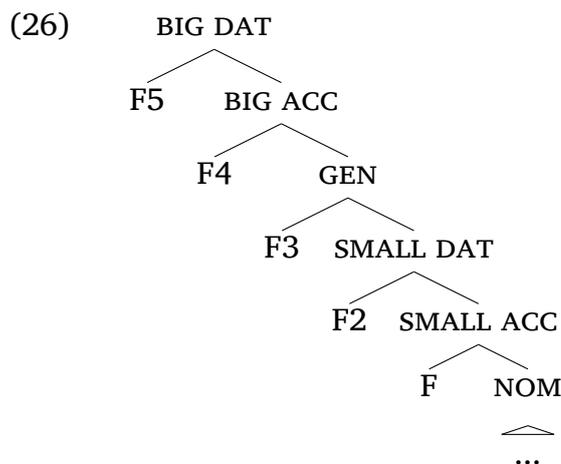
The basic analytical claim of such a hierarchy is that dative shift and differential object marking each require the existence of two related cases: the BIG DAT and the SMALL DAT for recipients, and the BIG ACC and SMALL ACC for patients. The BIG series is located above the genitive, while the SMALL series is below the genitive.

2.8   Back to Saami

At this point, it has become clear how languages with two datives and/or two accusatives fit into the Case sequence. The main point I want to make here is that this is in fact all we need in order to account for the problematic Skolt Saami facts.

In order to see this, let us turn our attention to languages which have neither the dative shift nor differential object marking. These languages must either have an obligatory dative shift (Icelandic), or lack it completely (Russian). Similarly, in languages without DOM, it must be the case that objects are either all marked by BIG ACC, or none are. Such a bi-furcation of languages with a single type of object marking has been proposed already in Aissen's (2003) seminal work.[4]

In abstract terms, the system thus allows for four logically possible types of single-marking languages, as depicted in the table (27). In the first two lines, I have placed Icelandic and Russian as representatives of languages with SMALL ACC and a single DAT.[5]

---

[4]Note that in these languages, both datives/accusatives exist in the underlying functional sequence, but the general working of syntax is such that one of them never gets to surface. From now on, I will sometimes be simplifying things by saying that Russian 'lacks' the small dative, but this is just a shortcut for saying that it has this constituent, even though it never gets to surface.

[5]It is possible that Russian has two different ACC cases, a big one for masculine animates, and a small one for the rest. However, the facts are also consistent with a

(27)   *Types of languages without alternations*

a.	SMALL ACC	SMALL DAT	Icelandic
b.	SMALL ACC	BIG DAT	Russian
c.	BIG ACC	SMALL DAT	?
d.	BIG ACC	BIG DAT	?

As the question marks in the table indicate, the combinatoric possibilities of the system open the space for the existence of two new language types. Should such languages be found, this would clearly strengthen the case for Starke's (2017) explanation of Harðarson's (2016) observation.

It turns out that the option depicted in (27d) (with both cases BIG) is the one we need in order to explain the formerly problematic facts in Skolt Saami. As we have seen, Skolt Saami requires the existence of the surface Case sequence in (28a), with the surprising placement of GEN next to NOM. Under the current proposal, this sequence corresponds to an expected incarnation of the universal hierarchy with both cases big, as in (28b).

(28)   a.   NOM—GEN—ACC—DAT—etc.
       b.   NOM—S-ACC—S-DAT—GEN—BIG-ACC—BIG-DAT—etc.

Interestingly, there is independent evidence for analysing the Skolt Saami paradigm along these lines. The evidence comes from comparing Skolt Saami to other Saami languages. I start with South Saami, which is the only Saami language that actually has DOM. When plural and indefinite, the marking of the object is identical to the nominative case, and referred to as such in the descriptive tradition, see (29a). When definite, it is marked by the ending *-ide*, referred to as the ACC, see (29b).

(29)   *South Saami* (Siegel 2017, attributed to Bergsland)

    a.   Laara treavk-ah   dorjeme
           Lars ski-PL.NOM make.PTCP.PERF
           'Lars has made skis.'

    b.   Dejtie   treavk-ide vööjnim
           dem.PL.ACC ski-PL.ACC see.PRT.1SG
           'I saw those skis.'

This shows that what the Saami grammatical tradition refers to as ACC in South Saami must in fact be BIG ACC. Not unexpectedly, the position of this case in the paradigm must be after the GEN, as the declension in table 2.4 shows.

---

single ACC approach, which I adopt here, as discussed in the main text. I will not dwell on this issue since the main point of this article lies elsewhere.

Table 2.4: South Saami declension (Ylikoski 2011)

	door, SG	door, PL
NOM	okse	oksh
GEN	oksen	oksi
ACC	oksem	oksi-de
DAT	oksese	oksi-de
INE	oksesne	oksi-ne
ELA	okseste	oksi-jste
COM	oksine	oksi-gujmie

The following two facts are relevant for the ordering of the cases. (i) In the plural, there is a syncretism between the ACC and the DAT, which is captured by placing the ACC after GEN in the paradigm. On its own, this fact could also be handled by placing the DAT before GEN (just like in Icelandic). However, what cannot be captured that way is that the ACC.PL is clearly derived from the GEN.PL by stacking. This confirms that what is called the accusative in South Saami is in fact BIG ACC, as one would expect given its DOM status.

However, DOM is rare among the Saami languages, in fact it is "a feature which is restricted to South Saami only" (Siegel 2017). For instance, the closely related Ume and Pite Saami show no DOM. In these languages, the single direct object case always carries the same marker (labelled ACC) even when inanimate, indefinite and/or non-specific. (30a,b) show this for Ume and Pite Saami respectively.

(30)　　*Indefinite/non-specific objects in Ume and Pite Saami*

　　a.　månne Arrviehaureje vuolgav　biebmo-ide　oasstiet
　　　　I　　　Arvidsjaur.ILL drive.1SG food-PL.ACC buy.INF
　　　　'I am driving to Arvidsjaur to buy food.' (Siegel 2017)

　　b.　färt　bäjjve mij bårojmä gulli-jd
　　　　every day　we ate　　fish-ACC.PL
　　　　'Every day, we ate fish.' (Wilbur 2014: 140)

Importantly, the marker we find on the objects is obviously a cognate of the South Saami BIG ACC marker. I will take this as an indication that the single ACC case in Pite and Ume Saami is the BIG ACC (an option allowed also in Aissen 2003). The BIG ACC status of the Ume/Pite ACC is confirmed by the ordering of the paradigms, given in table 2.5. For reasons of space, I only include here NOM, GEN, ACC and DAT.

Note that the GEN is placed after the NOM in the tables, with ACC after them, an ordering also used in the traditional descriptions of Saami languages. The important points are the following. (i) In both languages,

Table 2.5: Ume Saami (Siegel 2017) and Pite Saami (Wilbur 2014) declension

| | Ume Saami | | Pite Saami | |
	SG	PL	reindeer, SG.	reindeer, PL
NOM		-h	båts-oj	buhts-u
GEN	-n	-ij	buhts-u	buhts-u-j
ACC	-v/-b	-ij-de	buhts-u-v	buhts-u-j-d
DAT	-je	-ij-de	buhts-u-j	buhts-u-j-d-a

the ACC.PL contains the GEN.PL. (ii) In Pite Saami, this containment carries over to the singular. This requires an ordering with GEN before ACC, which is impossible to derive under the Caha hierarchy (13b), or any similar proposal to date, except as a reduced language-specific instantiation of the sequence proposed in Starke (2017).

In addition, note that in the Pite Saami plural, the ordering NOM—GEN—ACC—DAT is reflected by a steady increase in the complexity of marking, such that DAT contains ACC, ACC contains GEN and GEN NOM. If such a stacking is to be interpreted as a reflex of Merge F, then Merge F must derive DAT from ACC, ACC from GEN and GEN from NOM. This is precisely the ordering which is expected in languages with BIG-ACC and BIG-DAT under Starke's proposal.

The crucial point in the discussion up to now was to show that there is independent evidence for the Saami ACC being BIG, and since this ACC is syncretic with DAT, the DAT is big too. This general setup of the Saami case system, where both ACC and DAT are obviously more complex than GEN in terms of morphological complexity, paves the way for the existence of NOM-GEN syncretism, since they end up adjacent in a language where SMALL ACC and SMALL DAT are absent from the paradigm. And this is indeed the option which we have found in Skolt Saami (repeated here in table 2.6).

Table 2.6: Skolt Saami (Feist 2010: 145)

	1ST.PL	hole, SG.	hole, PL
NOM	mij	kåå'pp	kåå'v
GEN	mij	kåå'v	koo'v-i
ACC	mi'jjid	kåå'v	koo'v-i-d
DAT	mi'jjid	kåpp-a	koo'v-i-d

All summed up, there are three strong reasons to think that the Saami ACC is big. (i) It is cognate to the BIG ACC of South Saami. (ii) It contains

the GEN. (iii) It must not be in the way between NOM and GEN, because we find syncretism between NOM and GEN to the exclusion of ACC. I can see no way of accounting for these facts unless the Saami ACC is a BIG ACC, located after GEN in the Case sequence.

In addition, there is no way of accounting for the syncretism between NOM and GEN unless the DAT is a BIG DAT as well. This latter conclusion is in turn strengthened by the fact that the DAT in Saami also frequently contains the GEN, and that it tends to be syncretic with (what has to be) the BIG ACC.

Finally, BIG DAT must contain BIG ACC, because of the fact that DAT contains (what has to be) BIG ACC in Pite Saami. In addition, GEN is syncretic with BIG ACC in Skolt Saami, which also shows that BIG ACC must be next to GEN, and BIG DAT comes after these two.

2.9   Conclusion

Paradigms, as Plank (1991) puts it, are the oldest grammatical texts, with the first exemplars attested on Babylonian clay tablets going back to 1600 BC. In this article, I tried to explain what (for a Nanosyntactician in any case) is the main driving force that makes one keep looking at the same type of object three and a half thousand years later. The reason is that for the first time in the history of the field, we are in a position where paradigms stop being simple lists (whether arbitrary or ordered), and become highly structured objects, a product of the recursive application of the fundamental operation Merge F. The degree of unification between the traditional morphology and modern syntax achieved by this move keeps amazing me, and there is no better place for me to say this than exactly here.

The fact that the addition of the BIG/SMALL distinction on top of the Case hierarchy (13b) allows for a neat approach to Saami, is from this larger perspective a relatively insignificant result, but at the same time important, since without the BIG/SMALL distinction, the facts would be hard to unify with the world view just described.

References

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21(3). 435–483.

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

Belyaev, Oleg. 2010. Evolution of case in Ossetic. *Iran and the Caucasus* 14(2). 287–32. `http://ossetic-studies.org/biblio/05-Belyayev.pdf`.

Bobaljik, Jonathan. 2007. On comparative suppletion. Ms., University of Connecticut. `lingbuzz/000443`.

Caha, Pavel. 2009. *The nanosyntax of case*. CASTL, University of Tromsø dissertation. `lingbuzz/000956`.

Caha, Pavel. 2011. The parameters of case marking and spell out driven movement. *Linguistic variation yearbook 2010* 10. 33–78. `lingbuzz/001026`.

Caha, Pavel. 2017. How (not) to derive a *ABA: The case of Blansitt's generalisation. *Glossa: a journal of general linguistics* 2. 84. `lingbuzz/003782`.

Chomsky, Noam. 1995. Bare phrase structure. In Gert Webelhuth (ed.), *Government and binding theory and the minimalist program: Principles and parameters in syntactic theory*, 383–439. Oxford: Blackwell.

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. 315–332. `lingbuzz/000170`.

Cinque, Guglielmo & Luigi Rizzi. 2008. The cartography of syntactic structures. *STiL – Studies in Linguistics, CISCL Working Papers* 2. 42–58. `http://ciscl.media.unisi.it/doc/doc%7B%5C_%7Dpub/STiL-2008-vol2.pdf`.

Feist, Timothy. 2010. *A grammar of Skolt Saami*. University of Manchester dissertation.

Harðarson, Gísli Rúnar. 2016. A case for a weak case contiguity hypothesis-a reply to Caha. *Natural Language & Linguistic Theory* 34(4). 1329–1343.

Johnston, Jason Clift. 1996. *Systematic homonymy and the structure of morphological categories. Some lessons from paradigm geometry*. University of Sydney dissertation.

Maling, Joan. 2001. Dative: the heterogeneity of the mapping among morphological case, grammatical functions, and thematic roles. *Lingua* 111(4-7). 419–464.

Manzini, Rita & Ludovico Franco. 2016. Goal and DOM datives. *Natural Language and Linguistic Theory* 34(1). 197–240. `lingbuzz/001759`.

McCreight, Katherine & Catherine V. Chvany. 1991. Geometric representation of paradigms in a modular theory of grammar. In Frans Plank (ed.), *Paradigms: The economy of inflection*, 91–112. Berlin, New York: Mouton de Gruyter.

McFadden, Thomas. 2018. *ABA in stem-allomorphy and the emptiness of the nominative. *Glossa: a journal of general linguistics* 3. 8. `lingbuzz/003353`.

Plank, Frans. 1991. Rasmus Rask's dilemma. In Frans Plank (ed.), *Paradigms: The economy of inflection*, 161–196. Berlin: Mouton de Gruyter.

Ryding, Karin Christina. 2011. Arabic datives, ditransitives, and the preposition li. In Bilal Orfali (ed.), *In the shadow of Arabic. The centrality of language to Arabic culture*, 283–298. Leiden: Brill.

Siegel, Florian. 2017. Ume Saami – The forgotten language. *Études finno-ougriennes* 48. `http://journals.openedition.org/efo/7106`.

Smith, Peter W., Beata Moskal, Ting Xu, Jungmin Kang & Jonathan Bobaljik. 2018. Case and number suppletion in pronouns. Ms., Goethe University Frankfurt, Uconn. `lingbuzz/003110`.

Starke, Michal. 2009. Nanosyntax. A short primer to a new approach to language. In Peter Svenonius, Gillian Ramchand, Michal Starke & Tarald Taraldsen (eds.), *Nordlyd 36: special issue on Nanosyntax*, 1–6. Tromsø: University of Tromsø. `lingbuzz/001230`.

Starke, Michal. 2017. Resolving (DAT = ACC) ≠ GEN. *Glossa: a journal of general linguistics* 2(1). 104.1–8. `lingbuzz/003783`.

Van Baal, Yvonne & Jan Don. 2018. Universals in possessive morphology. *Glossa: a journal of general linguistics* 3. 11.

Wilbur, Joshua. 2014. *A grammar of Pite Saami*. Berlin: Language Science Press.

Ylikoski, Jussi. 2011. From compound nouns to case marking: Prolatives in South Saami and Lule Saami. In Cornelius Hasselblatt & Beáta Wagner-Nagy (eds.), *Finnisch-Ugrische Mitteilungen 39*, 283–298. Helmut Buske Verlag. `https://munin.uit.no/bitstream/handle/10037/10383/article.pdf?sequence=4`.

Zimmermann, Ilse. 2002. Structural cases in Russian. In *More than words: A festschrift for Dieter Wunderlich*, 275–298. Berlin: Akademie Verlag.

Zompì, Stanislao. 2017. *Case decomposition meets dependent-case theories*. Pisa, Italy: Università di Pisa MA thesis. `lingbuzz/003421`.

3 Gelukkige verjaardag!

Karen De Clercq & Guido Vanden Wyngaerd

FWO/Ghent University & KU Leuven

3.1 Introduction

There is in Dutch a minimal contrast between the evaluative adverb *gelukkig* 'fortunately', and its antonym *ongelukkig-erwijs* 'unfortunately'. Both are derived from the adjective *gelukkig* 'lucky, happy', but only the latter is marked by the overt adverbial suffix *-erwijs* '-ly'. We present an account of this contrast in terms of intervention by a Neg head, which forces the separate spellout of a higher portion of the functional hierarchy as an adverbial suffix.

3.2 Dutch evaluative adverbs

As a rule, adverbs are not formally marked in Dutch, and are therefore mostly syncretic with adjectives. An exception to this general rule are evaluative adverbs, which need to be formally marked, either morphologically or syntactically. Morphologically, they may be marked by the suffix *-erwijs*, which derives adverbs from adjectives. Some examples of evaluative adverbs are given in (1):

(1)   verrassend-erwijs   'surprisingly'
        interessant-erwijs   'interestingly'
        begrijpelijk-erwijs   'understandably'
        paradoxal-erwijs   'paradoxically'
        merkwaardig-erwijs   'curiously'
        opvallend-erwijs   'strikingly'

The meaning of these evaluative adverbs can be paraphrased as in (2), where A indicates the slot for the corresponding adjective, and *p* is the sentence in which the adverb occurs:[1]

---

[1]This is an informal approximation; for more detailed discussion of evaluative adverbs, see Bellert (1977); Bonami & Godard (2008); Mayol & Castroviejo (2013), and

(2)   The speaker finds it A that *p*

As an alternative to suffixation with *-erwijs*, the relevant adjectives may also be followed by *genoeg* 'enough' to derive the evaluative adverb interpretation:[2]

(3)   Merkwaardig genoeg was het rijtuig   een dwarsbalk verloren.
        curious   enough was the carriage a   crossbar lost
        'Curiously, the carriage had lost a crossbar.'

Barbiers (2001) calls this a case of *genoeg*-support, because this type of *genoeg* 'enough' does not appear to contribute anything to the meaning of the adjective, i.e. (4a) has a meaning equivalent to (4b) rather than (4c):

(4)   a.   Marie heeft vreemd genoeg niet gebeld.
             Marie has   strange enough not called
             'Strangely, Marie hasn't called.'
        b.   Het is vreemd dat Marie niet gebeld heeft.
             It   is strange that Marie not called has
             'It is strange that Marie hasn't called.'
        c.   Het is vreemd genoeg dat Marie niet gebeld heeft.
             It   is strange enough that Marie not called has
             'It is strange enough that Marie hasn't called.'

The meaning of (4a) is paraphrasable as in (4b), which is in essence that of (2). No semantic contribution is made by *genoeg* 'enough', which is semantically vacuous, but whose presence seems to be required for syntactic reasons only (much like *do*-support in the verbal domain in English).[3] To the paraphrase of (4b) we may now add *genoeg*, as in (4c), and in this case *genoeg* 'enough' does make a semantic contribution as an adjectival modifier. This meaning is roughly equivalent to 'sufficiently A (to VP)', as illustrated in (5):

(5)   Die   schoenen zijn groot genoeg voor mij om   te dragen.
        those shoes   are large enough for   me COMP to wear.

---

references cited there.

  [2]Marking of the adverb by *genoeg* 'enough' is more productive than marking by *-erwijs*, in that *genoeg* can be added to any adjective with an evaluative meaning, but *-erwijs* to only a subset of these (e.g. *vreemd genoeg* 'oddly' vs *vreemd-erwijs*). On the other hand, adverbs derived with *-erwijs* can have a wider range of meanings than those with *genoeg* 'enough', and be used as adverbs of alethic modality, expressions of durative aspect, or adverbs of manner (see De Belder & Vanden Wyngaerd 2018 for discussion).

  [3]*Genoeg*-support also recalls the phenomenon of *much*-support in English, as discussed in Corver (1997).

'Those shoes are large enough for me to wear them.'

As (5) shows, regular *genoeg* has argument structure of its own, like a PP with *voor* 'for', and an infinitival clause introduced by the complementizer *om*). As Barbiers observes, such argument structure is unavailable with *genoeg*-support, as shown in (6).

(6) *Marie heeft vreemd genoeg voor mij niet gebeld om   me
      Marie has   strange enough for   me not called COMP me
      zorgen   te maken.
      worries to make
      Intended: 'Marie hasn't called, and I find this strange enough to worry.'

As we already stated above, the evaluative adverb interpretation requires either the adverbial suffix *-erwijs*, or *genoeg*-support. Without either of these two present, the intended reading disappears, and the sentences become ungrammatical:

(7) *Merkwaardig/verrassend/begrijpelijk was hij gekwetst.
      curious/surprising/understandable   was he hurt
      intended: 'Curiously/Surprisingly/Understandably, he was hurt.'

There is, however, one exception to this general rule: the adjective *gelukkig* 'lucky, happy' may appear with the evaluative interpretation (to mean 'fortunately') in the absence of either the suffix *-erwijs* or *genoeg*-support:

(8) Gelukkig was er   een dokter in de zaal.
      fortunate was there a   doctor in the room
      'Fortunately, there was a doctor in the room.'

Interestingly, a minimal contrast arises with the polar opposite of *gelukkig* 'fortunate(ly)', the *on*-prefixed adjective *ongelukkig* 'unfortunate'. Although clearly morphologically derived from *gelukkig* 'fortunate(ly)' through *on*-prefixation, *ongelukkig* 'unfortunate' again falls under the general rule that either the suffix *-erwijs* or *genoeg*-support are needed for the evaluative adverb reading:[4]

(9) {Ongelukkigerwijs / ongelukkig genoeg / *ongelukkig} was
      unfortunately   / unfortunate enough / unfortunate   was
      er   geen dokter in de zaal.
      there no   doctor in the room

---

[4]The Historical Dictionary of the Dutch language (*Woordenboek der Nederlandse Taal,* WNT; `http://www.wnt.inl.nl`) also gives *(on)gelukkiglijk* as archaic forms of the adverb.

'Unfortunately, there was no doctor in the room.'

The purpose of this brief paper is to shed a light on this contrast. We shall argue, first, that *gelukkig* has a lexical entry that differs from that of other evaluative adverbs, and that allows it to spell out a larger structure, without needing to take recourse to a suffix. Second, we shall attribute the necessary appearance of a suffix with the negative *on-gelukkig* 'unfortunate' to an intervention effect caused by the negative prefix, which blocks the adjective from functioning as an evaluative adverb. We shall show that this blocking effect arises with evaluative adverbs, but not with manner adverbs. This leads us to hypothesize that there are (at least) two different adverbial heads in the functional sequence.

3.3 The internal structure of adverbs

The case of the Dutch evaluative adverbs shows that adverbs may be derived from adjectives through the addition of a suffix. This obviously fits into a broader pattern found in English and numerous other languages, where adverbs are visibly larger than adjectives (e.g. *quick, quickly*). Yet at the same time, the case of *gelukkig* instantiates a pattern that is also found elsewhere, namely one where the adverb is unmarked, i.e. syncretic with the adjective (e.g. *fast*). This state of affairs can be explained as a difference in size of the lexical items involved: an adverb is larger than an adjective, i.e. it spells out more structure, as witnessed by the fact that adverbs are often decomposable as adjectives plus an extra marker that turns them into adverbs. For concreteness, we assume that adverbs realise an extra feature ADV. Unmarked adverbs, i.e. adverbs that are syncretic with adjectives, spell out the entire structure, whereas marked adverbs arise on the basis of adjectives that can realise less features. We represent this rather informally as in the following table:

(10)

A	ADV
quick	ly
fast	

In the same way, the difference between morphologically marked evaluative adverbs in Dutch and the unmarked one *gelukkig* 'fortunately' is represented in (11):

(11)

A	ADV
merkwaardig	erwijs
gelukkig	

Taking the somewhat informal concept of the 'realisation of features'

42

to correspond to the nanosyntactic mechanism of phrasal spellout, we assume that a syncretic form like *fast* spells out ADVP. The syncretism between the adverb and the adjective arises as a familiar consequence of the Superset Principle. A nonsyncretic adjective like *quick* realises a smaller constituent, so that a separate exponent *-ly* is needed to spell out the adverbial head. Applied to the case of the evaluative adverbs in Dutch, the contrast between the marked and the unmarked adjectives reduces to the size of their lexical trees. The unmarked evaluative adverb *gelukkig* spells out an ADVP, i.e. a constituent that contains an ADV feature on top of the projection of the gradable adjective, as shown in (12). We furthermore take gradable adjectives to decompose into a gradability feature Q and a root feature (see De Clercq & Vanden Wyngaerd to appear), so that the full structure of *gelukkig* 'fortunate(ly)' looks as in (12):

(12)
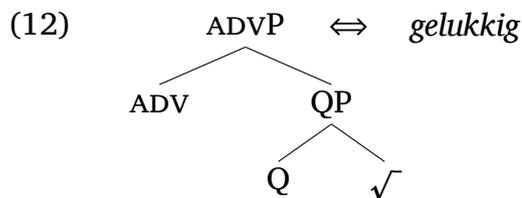
Standard Superset Principle logic implies that *gelukkig* may spell out both a gradable adjective (QP) and an adverb (ADVP).

Suffixally marked adverbs are lexically smaller: they involve an adjectival root of the size QP (i.e. that of a gradable adjective), and need an additional exponent spelling out the ADV feature. This requires spellout-driven movement of QP across ADV, allowing for the spellout of the suffix, as shown in (13) for the case of *merkwaardig-erwijs* 'curiously'.

(13)
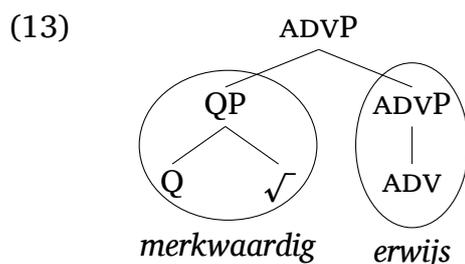
De Clercq & Vanden Wyngaerd (to appear) argue that the negative prefix *on-* 'un-' spells out a Neg feature as well as a gradability feature Q, and that this prefix is merged as a specifier in the adjectival spine (Starke 2018). Since the adjectival spine has an optional Neg position on top of QP, the negative specifier projects a NegP in the main spine (Starke 2004). This is shown in (14):

(14)

```
                    NegP
          ⟨ QP ⟩      ⟨ QP ⟩
         Q    Neg    Q    √
            on          gelukkig
```
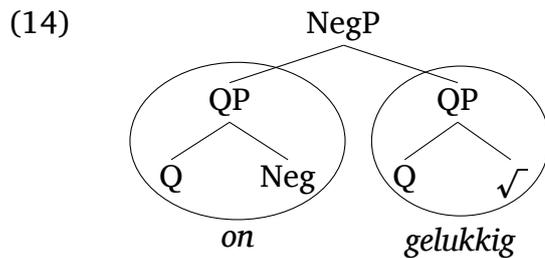
This is the structure for the negative adjective *ongelukkig* 'unhappy'. Now this can be turned into an adverb by merging the ADV feature on top of it:

(15)

```
              ADVP
        ADV          NegP
                ⟨ QP ⟩      ⟨ QP ⟩
               Q   Neg     Q    √
                  on          gelukkig
```
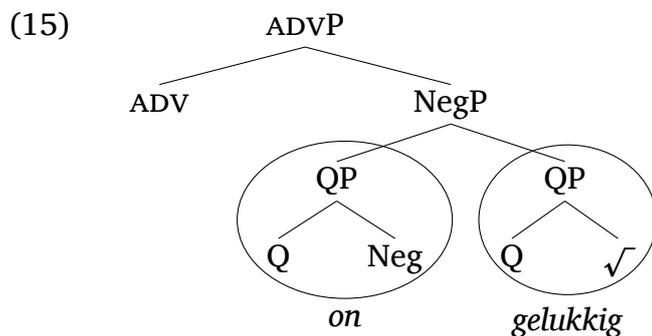
But there is no lexical item that can spell out this ADVP. The lexical tree for *gelukkig* 'fortunate(ly)' (which is as in (12) above) does not contain (15) as a subtree, due to the additional NegP in (15). This is what we described earlier as an intervention effect: the NegP projected by the prefix intervenes between the Q and ADV heads, and this prevents the spellout of ADVP by the lexical item *gelukkig*, even though the latter is big enough in principle to spell out ADVP. The reason that it cannot do so in this case is that the syntactic derivation contains an element which is optional in the functional sequence, namely NegP. Because of this intervening element, *gelukkig* is restricted to spelling out the lower projection QP, 'shrinking' as it were below NegP, and the higher features of the sequence (Neg and ADV) need to be spelled out by separate exponents (respectively *on-* and *-erwijs*). In particular, the ADV head needs to be spelled out by the available suffix because of the intervening NegP.

The way the ADV feature is spelled out in the negative adverb *ongelukkigerwijs* 'unfortunately' is that NegP raises to adjoin to ADVP, allowing the spellout of the latter as the suffix *erwijs*, as shown in (16):[5]

---

[5]In the formulation of the spellout algorithm of Starke (2018), this type of rollup movement would only be attempted after movement of the Spec of the complement of ADV, and subsequent unsuccessful spellout. In the formulation of Caha, De Clercq & Vanden Wyngaerd (2017), projecting Specs are treated differently from nonprojecting ones, in that with the projecting ones rollup movement is attempted first. Whichever option is chosen makes no difference in this particular case, however, since Spec-to-

(16)
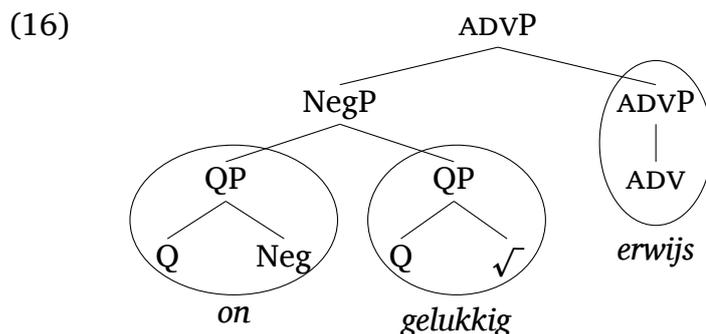
Finally, note that both in (15) and (16) the negative prefix is a structural, but not a linear, intervener between the adjectival root and the suffix. Since the Neg head sits on the opposite side of the root from the suffix, the adjectival root and the suffix are linearly adjacent, but not structurally adjacent. These data argue against possible alternative accounts of the intervention effect in purely linear terms, such as the account in terms of spanning proposed by Haugen & Siddiqi (2016).

An analogous case of lexical 'shrinking' as a consequence of Neg intervention is discussed in Caha, De Clercq & Vanden Wyngaerd (2017). They note that the Czech suppletive adjective pair *dobrý-lepší* 'good-better' reverts to the nonsuppletive root in the comparative if the adjective is prefixed with the negative prefix *ne-* 'un-':

(17)

POS	CMPR	
dobr-ý	lep-š-í	'good-better'
ne-dobr-ý	ne-dobř-ej-š-í	'bad-worse'
	*ne-lep-š-í	

This pattern is exactly the same as that observed with the Dutch evaluative adverb pair *gelukkig-ongelukkigerwijze* 'fortunately-unfortunately', under the assumption that the suppletive root is larger than the nonsuppletive one, and that the NegP is a structural intervener in the adjectival sequence between QP and the markers of the comparative, as shown in (18):

Spec movement of the prefix will not lead to successful spellout of ADVP in any event.

45

(18)

```
                              C2P
                         ┌─────┴──────┐
                        C1P          (C2P)
                    ┌────┴────┐         │
                  NegP      (C1P)      (C2)
              ┌────┴────┐     │         š
           (QP)      (QP)    (C1)
          ┌──┴──┐   ┌──┴──┐   ěj
          Q   Neg   Q    √P
              ne        dobr
```
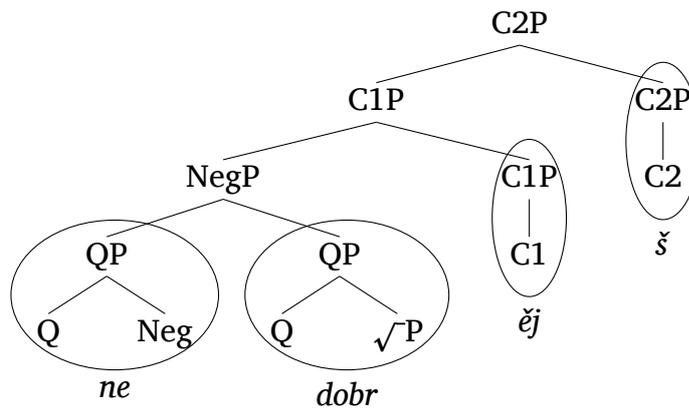
The suppletive root *lep* spells out the complex of QP and C1 (i.e. C1P) (see Caha 2017; De Clercq & Vanden Wyngaerd 2017; Caha, De Clercq & Vanden Wyngaerd 2017 for more discussion of the analysis of comparative suppletion along these lines). However, *lep* cannot spell out C1P in this case because C1P also contains NegP. As a result, the adjectival root has to shrink to spell out QP, and gets realised as *dobr*. In addition, C1 now needs the additional comparative marker *-ěj-* in the nonsuppletive comparative form *ne-dobř-ej-š-í* 'worse'. This is exactly parallel to the emergence of the adverbial marker *-erwijs* at the point where *gelukkig* shrinks to spelling out QP due to the intervening NegP.

3.4 More adverb positions

Let us return to the functional sequence that we have established so far:

(19)

```
          ADVP
       ┌───┴────┐
      ADV      NegP
           ┌────┴────┐
          Neg        QP
                  ┌───┴───┐
                  Q       √
```
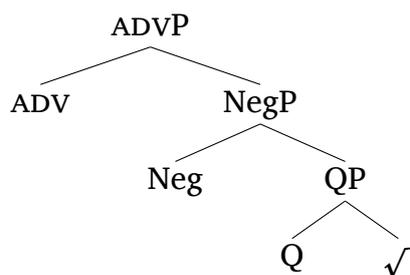
This hierarchy makes a concrete prediction, which is that no adjective with the negative prefix *on-* 'un-' should be syncretic with an adverb: as in the case discussed in the previous section, NegP will intervene and cause separate lexicalisation of the ADV feature. However, this prediction turns out to be wrong, in that there are numerous cases of manner adverbs which are syncretic with *un*-prefixed adjectives:

(20) a. Hij heeft haar on-menselijk/on-vriendelijk behandeld
          he has   her un-human/un-friendly   treated

46

'He has treated her cruelly/in an unfriendly manner.'
  b. Ze   eten on-gezond.
      they eat   un-healthy
      'They eat unhealthyly.'

These facts suggest that there are different ADV features, corresponding with different types of adverbs. Assuming that there is a feature for evaluative adverbs above NegP, and one for manner adverbs below NegP will give us what we need. This order respects the order for adverbs proposed by Cinque (1999).

(21)
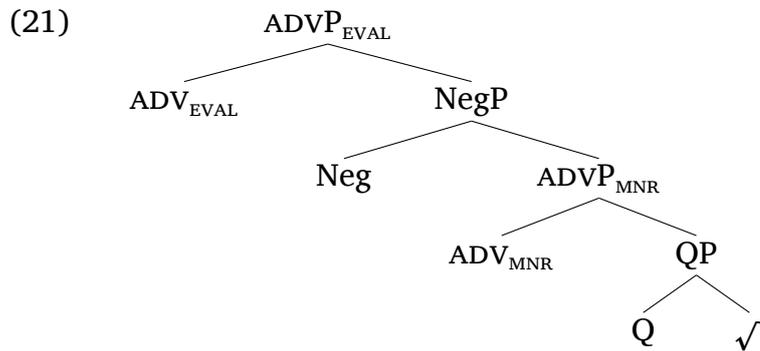
Now manner adverbs can be syncretic with adjectives and still be prefixed with *on* 'un-', since the negative prefix will not intervene between Q and ADV$_{MNR}$, thus allowing the adjective-adverb syncretism, while at the same time allowing the addition of the negative prefix on top of the adverb or adjective in subsequent stages of the derivation. The tree below gives the derivation for the manner adverb *ongezond* 'unhealthily':

(22)
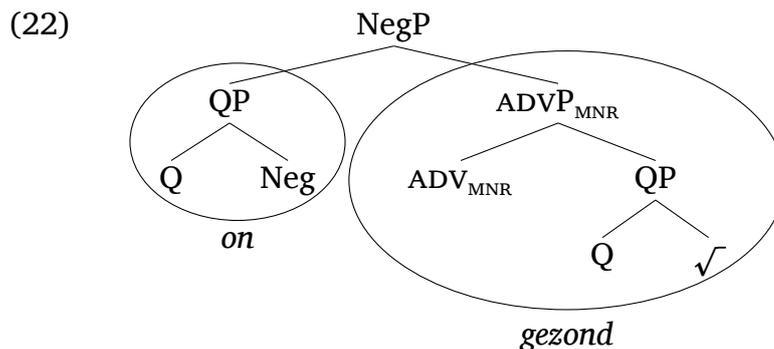
That there are two different adverbial positions in the sentential hierarchy is further confirmed by the fact that evaluative and manner adverbs can co-occur, even if they involve the same adjectival root:

(23)   a. Vreemd genoeg behandelde hij de dieren   vreemd.
            strange enough treated   he the animals strange
            'Strangely, he treated the animals strangely.'

b. Begrijpelijkerwijs probeerde hij zijn voorstel
   understandably   tried   he his proposal
   begrijpelijk   uit te leggen.
   understandable out to lay
   'Understandably, he tried to explain his proposal in an understandable way.'

3.5 Conclusion

In this paper we have investigated an intervention effect caused by a negative head in the sequence of heads entering into the internal makeup of evaluative adverbs in Dutch. The intervening head caused the adjectival root to shrink, and the adverbial head to be spelled out by a separate suffix. The fact that the intervention effect does not show up with manner adverbs led us to postulate two distinct adverbial heads, one for manner adverbs below NegP and one for evaluative adverbs above NegP. This proposal for the functional sequence inside words agrees with the one proposed for the clausal level in Cinque (1999).

References

Barbiers, Sjef. 2001. Is *vreemd genoeg* genoeg? In Bernardus Dongelmans, Josine Lalleman & Olf Praamstra (eds.), *Kerven in een rots. Opstellen over Nederlandse taalkunde, letterkunde en cultuur, aangeboden aan Jan W. de Vries bij zijn afscheid als hoogleraar Dutch Studies aan de Universiteit Leiden*, 15–28. Leiden: SNL. `https://pure.knaw.nl/ws/files/457476/genoeg.pdf`.

Bellert, Irena. 1977. On semantic and distributional properties of sentential adverbs. *Linguistic Inquiry* 8(2). 337–351.

Bonami, Olivier & Danièle Godard. 2008. Lexical semantics and pragmatics of evaluative adverbs. In Louise McNally & Chris Kennedy (eds.), *Adverbs and adjectives: syntax, semantics, and discourse*, 274–304. Oxford: Oxford University Press.

Caha, Pavel. 2017. Explaining Bobaljik's root suppletion generalization as an instance of the adjacency condition (and beyond). In Joseph Emonds & Markéta Janebová (eds.), *Language use and linguistic structure: proceedings of the Olomouc linguistics colloquium 2016*, 193–208. Olomouc: Palacký university.

Caha, Pavel, Karen De Clercq & Guido Vanden Wyngaerd. 2017. The fine structure of the comparative. Ms. Masarykova Univerzita, U Ghent, KU Leuven.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: a cross-linguistic perspective*. Oxford: Oxford University Press.

Corver, Norbert. 1997. Much-support as a last resort. *Linguistic Inquiry* 28. 119–164.

De Belder, Marijke & Guido Vanden Wyngaerd. 2018. Zingenteere verge-lijkenderwijs. In Jan Ceuppens, Hans Smessaert, Jeroen Van Craen-enbroeck & Guido Vanden Wyngaerd (eds.), *A coat of many colours. Vriendenboek voor Dany Jaspers naar aanleiding van zijn zestigste verjaardag*. Brussels: KU Leuven. `www.dj60.be/contents-3/`.

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

De Clercq, Karen & Guido Vanden Wyngaerd. to appear. Negation and the functional sequence. *Natural Language and Linguistic Theory*.

Haugen, Jason D. & Daniel Siddiqi. 2016. Towards a restricted realization theory: multimorphemic monolistemicity, portmanteaux, and post-linearization spanning. In Daniel Siddiqi & Heidi Harley (eds.), *Morphological metatheory*, 343–385. Amsterdam: John Benjamins.

Mayol, Laia & Elena Castroviejo. 2013. (Non)integrated evaluative adverbs in questions: a cross-Romance study. *Language* 89. 195–230.

Starke, Michal. 2004. On the inexistence of specifiers and the nature of heads. In Adriana Belletti (ed.), *Structures and beyond*, vol. 3, 252–268. Oxford: Oxford University Press.

Starke, Michal. 2018. Complex left branches, spellout, and prefixes. In Lena Baunaz, Karen De Clercq, Liliane Haegeman & Eric Lander (eds.), *Exploring nanosyntax*, 239–249. Oxford: Oxford University Press.

4 Theme vowels are verbs

Antonio Fábregas

University of Tromsø

4.1 Introduction and main claims

The goal of this squib is to provide several empirical arguments that support the view that the real verbal predicate in Spanish are the theme vowels that appear combined with what is traditional regarded as the lexical verb (1). In a nutshell, then, theme vowels are light verbs,[1] as suggested by Kayne (2016) following Solà (1994).

(1) a. cant-**a**
         sing-THV1
         'sing'
     b. beb-**e**
         drink-THV2
         'drink'
     c. viv-**i**
         live-THV3
         'live'

While in examples such as (1) this claim is essentially trivial if one assumes that the unit that the theme vowel combines with is an acategorial root (Marantz 1997; Arad 2003; Marvin 2002), the apparent problem of this approach is that theme vowels also seem to appear with morpholo-

---

[1]Kayne himself admits that the notion of 'light verb' is difficult to define in pretheoretical terms, and it is largely an empirical issue to determine which properties it should have. Here we adopt as a working definition the idea that a light verb should be viewed as a constituent that licenses the presence of verbal functional structure, such as tense, aspect, mood and subject or object agreement, but does not introduce enough conceptual information to be a predicate by itself; the light can also provide syntactic positions for arguments, but does not determine autonomously the interpretation of those arguments. Some verbs are always light in this sense, such as *ser* and *estar,* while some full verbs can double as light verbs, as *poner.* See Butt (1995; 2003); Butt & Geuder (2001); Grimshaw & Mester (1988); Lin (2001); Mohammad & Karimi (1992) for further properties.

gical verbalisers, as in the example (2). The standard analysis of such cases in Neo-Constructionist approaches is that the first constituent is a root, while the verbaliser should be decomposed in two parts: *ific* as the spell out of the verbal layer, and *a* as the theme vowel.

(2)   clas- ifica
        class- ify
        'classify'

We will also provide arguments that in (2) the verbaliser is in fact the theme vowel, and the *ific* constituent should be treated as a root modifying it (Lowenstamm 2014). Thus, we will argue for the following identity:

(3)   Theme vowel = Light verb = Verbaliser

The relevance of this empirical observation is that it makes it easier to dissolve a frequently cited *prima facie* counterargument against syntactic approaches to word formation. Blevins (2007), for instance, points out that the fact that some languages have theme vowels while others don't supports an idiosyncratic—and therefore lexicalist—treatment of word formation to the extent that whether a language has or does not have theme vowels is an arbitrary property of how verbs acquire their morphological shape with no consequences for syntax or semantics (unlike for instance PRO DROP, V2 or RICH AGREEMENT). If theme vowels are light verbs, then all languages have 'theme vowels' in the deep sense because all languages have verbs (cf. Kayne 2016 for the claim that English *e* in *grad-e-d* is a theme vowel). Second, theme vowels do not seem to correspond to independent syntactic objects in the available theories. In lexicalist proposals, the base is already a verb, and the theme vowel just marks the conjugation class (4). In Neo-constructionist theories (Halle & Marantz 1993), the verbaliser is the real verb and the theme vowel has to be introduced post-syntactically (as a dissociated morpheme, cf. M. Oltra-Massuet 1999) to mark the conjugation class (5).

(4)   [$_V$ cant-a]

(5)   [[[$_\sqrt{}$clas] ific]$_v$ + a]

This second problem is also dissolved, because if this proposal is right the theme vowel would be the spell out of the verbal layer(s), a relatively well-understood syntactic and semantic entity.

4.2 Theme vowels = light verbs

Kayne (2016) argues, following an initial suggestion by Solà (1994), that the right segmentation of an English regular past tense is (6).

(6)   rain$^{\text{Root}}$-e$^{\text{THV}}$-d$^{\text{Tense}}$

If theme vowels are light verbs, then a theme vowel and a light verb would in principle be introduced in the same position—assuming for the time being that there is a unique syntactic position for both. From here it follows that the best candidates to be light verbs in English should lack any presence of the /e/ that corresponds to the theme vowel, a prediction that seems borne out:

(7)   *beed, *haved, *doed, *goed, *comed, *taked, *bringed, *putted, *getted, *gived, *maked, *letted, *sayed, *canned, *willed, *shalled, *mayed, *musted

In what follows I will extend Kayne's observation to Spanish. Let us take the prototypical case of a light verb in Spanish, *ser*. As can be seen in (8), this verb is prototypically light in at least two senses: it lacks enough conceptual semantics to stand alone as a predicate, as witnessed by (8b), and it has plain uses as an auxiliary, in particular as a passive auxiliary (8c).

(8)   a.   Juan es alto.
              Juan is tall
              'Juan is tall'
        b.   *Juan es.
              Juan is
              'Juan is.'
        c.   El prisionero fue detenido.
              the prisoner   was detained
              'The prisoner was detained.'

In the present paradigm it is impossible to identify any segment that could plausibly correspond to the THV.

(9)   SG   1   so-y
             2   ere-s
             3   es
        PL   1   so-mos
             2   so-is
   3   so-n

Moreover there is no evidence that here we have a root that combines

with the theme vowel. The few cases that can be argued to relate to this verb through word formation are bizarre, infinitives or highly lexicalised expressions that are quite likely to come from Latin.

(10)   a.   esencia
            is.ance
            'essence' (not 'state of being')
       b.   ser
            to.be
            'entity'

My claim is that both facts—absence of a root + theme vowel make up and its light verb nature—are correlated and *be* lacks a theme vowel because both *be* and the theme vowel are the same object in the syntax. This explains the facts if we have an alternation along the shape of (11), with a 'standard' verb consisting of the light verb and the root, and *ser* being just the light verb without any root:

(11)   VP   (12)   VP
            /   \ |
          V   √   SER
          | |
          a   cant-
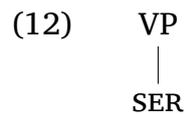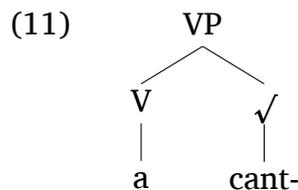
    Another prototypical case of light verbs are those used as aspectual auxiliaries: the progressive *estar* 'be', the perfect *haber* 'have' and the prospective *ir* 'go'.

(13)   a.   Juan está comiendo.
            Juan is   eating
            'Juan is eating.'
       b.   Juan ha comido.
            Juan has eaten
            'Juan has eaten.'
       c.   Juan va   a comer
            Juan goes to eat
            'Juan is going to eat.'

The first two do not have a use as semantically 'strong' verbs, but they are closely related in existential sentences, where they contrast depending on the definiteness of the only argument, with definites combining with *estar* (14b).

(14)   a.   Hay   un libro.
            there.is a   book
            'There is a book.'

54

b. Estaba   el libro.
       there.was the book
       'The book was there.'

The third verb is closely related to *ser* in its morphology: the perfective form of *ir* and *ser* is morphologically identical.

(15)   fui (3SG.PRF of *be* and *go*)

I want to argue that these three light verbs also lack a theme vowel in the present. Consider their paradigms.

(16)   *estar*   *haber*   *ir*
    SG 1   esto-y   he   vo-y
        2   está-s   ha-s   va-s
        3   está   ha/ha-y   va
    PL 1   esta-mos   he-mos   va-mos
        2   está-is   habé-is   va-is
        3   está-n   ha-n   va-n

Prima facie, the verb *estar* could be interpreted as consisting of a root *est-* and a first conjugation theme vowel *a,* but this causes problems for the algorythm that assigns stress in Spanish according to I. Oltra-Massuet & Arregi (2005), which should assign stress to the first vowel, as in *cantar.* The problem disappears if the verb is segmented as (19).

(17)   a. /es.tá/ (cf. /kán.ta/)
        b. /es.tán/ (cf. /kán.tan/)

(18)   a. est-á
        b. est-á-n

(19)   a. está
        b. está-n

In the case of *ir,* there is simply no evidence that *va* should be segmented: *v-* is not used as a root. In the case of *haber,* with the only possible exception of the 2pl *habéis,* which could be amenable to a segmentation *hab-é-is,* there is no segment that could correspond to the root—remember that in Spanish the letter 'h' has no sound. The same claim should be extended to the verb *dar,* another strong candidate to being a light verb:

(20)   a. doy
        b. das
        c. da
        d. damos
        e. dais

f.   dan

Here are some pieces of additional evidence that these verbs lack any theme vowel.

First of all, we have the fact that these verbs do not make an indefinite past as one would expect if the *a* were a Theme Vowel. The indefinite past of *cant-a* is *cant-a-ste,* but these verbs do not follow this pattern:

(21)   a.   *estaste
       b.   estuviste

(22)   a.   *daste
       b.   diste

(23)   a.   *vaste
       b.   fuiste

Second, nobody, to the best of my knowledge, has ever tried to regularise the past tense of these verbs, unlike what is the case with the verb *andar* (24), which can be segmented in a theme vowel and a root (25).[2]

(24)   a.   and-a-ste
       b.   anduv-i-ste

(25)   a.   /án.da/
       b.   /án.dan/

Contrast this with our light verbs that include the sequence */ub/* in their indefinite past:

(26)   a.   *estiste / estuviste
       b.   *histe / hubiste

Another interesting generalisation regarding these verbs without a theme vowel—albeit one I have no explanation for—is that they are precisely those that add a segment /I/ (ortographically, *-y*) in one form of the present indicative paradigm, normally the 1SG (with the verb *haber* adding it only in presentational contexts):

(27)   a.   so-y, esto-y, vo-y, do-y
       b.   ha-y

---

[2]As far as we can tell, the alternation between *andaste* and *anduviste* does not reflect two uses of the same verb, one as full verb and one as light verb. The form *anduviste* is getting lost in contemporary Spanish, and it is safe to say that speakers that use it learn it at schools, and some of them quickly forget them. For this reason, pending further research, it seems plausible to us that speakers treat *anduv* as an allomorph of the root and the structure of this verb is always the one associated to full verbs.

At the very least, that these verbs—and only these verbs—take this morpheme strongly suggests that they share some property with each other. My suggestion is that the property is that they are, in themselves, theme vowels.

4.3 Verbaliser = Theme Vowel = Light verb

There is a long tradition that views the verbalisers as light verbs (Harley 1995). Just as what we have descriptively labeled theme vowels, they combine with roots, and are associated to abstract meanings which in compositional cases function as semantic skeletons. Treating the verbalisers as light verbs is therefore not an unheard of idea, but the problem emerges when we combine this with the proposal that the theme vowel itself is a light verb, essentially because in principle the verbaliser and the theme vowel seem to co-occur. (28) lists the most productive verbalisers in contemporary Spanish; note that the last segment in each one of them is morphophonologically identical to a theme vowel, and also defines the resulting complex verb as belonging to the conjugation class associated to that theme vowel.

(28) a. **iza**   (autor-iza 'authorise')
      b. **ifica**   (clas-ifica 'classify')
      c. **ita**   (debil-ita 'debilitate')
      d. **ece**   (palid-ece 'to become pale')
      e. **ea**   (tont-ea 'to act silly')

Thus, a verb like *palidecer* inflects as a second conjugation verb, as expected if the final *e* in *ece* is the second conjugation theme vowel (29). A verb like *autorizar* inflects in the first conjugation (30), again as expected if the final segment is the theme vowel.

(29) a. comer > com-í-a (IPFV, 3SG)
      b. palidecer > palid-ecí-a (IPFV, 3SG)

(30) a. cantar > cant-a-ba (IPFV, 3SG)
      b. autorizar > autor-iza-ba (IPFV, 3SG)

I argue that in fact these verbalisers are allomorphs of the theme vowels.[3]

---

[3]An alternative would be to say that *ific, ec, iz* are roots modifying the light verb, as argued by Lowenstamm (2014) in the case of some apparent adjectivalisers / nominalisers. There are two reasons not to adopt this analysis: first, I am not aware of any use of the relevant morphemes as category-changing affixes producing nominalisations or adjectivalisations, which would make surprising that they are roots. Second, as Tarald Taraldsen (p.c.) made us notice, if they were modifiers of the VP layer, then we would expect them to be able to combine with light verbs.

I can provide three arguments supporting this analysis. First of all, the traditionally considered 'verbalisers' *ific, it, ec* and so on never ever appear without the segment we believe is the theme vowel. There are no plausible phonological reasons for this. Removing the theme vowel from *ifica* before a nominaliser like *ción* 'ation' would produce a sequence that is perfectly possible in other contexts, such as (31).

(31)   destruc-ción

Still there are no words in Spanish ending in a sequence *ificción* which can be plausibly analysed as the nominalisation of a verb. The same applies to **itción, *eción, *izción*. Note in contrast that there are many words ending in -*facción*:

(32)   a.   calefacción (heating)
        b.   rarefacción (rarefaction)
        c.   licuefacción (liquefaction)
        d.   putrefacción (putrefaction)

Importantly, in none of these cases does the noun derive from a verb:

(33)   a.   *calefacer
        b.   *rarefacer
        c.   *licuefacer
        d.   *putrefacer

The only case to my mind where there is a sequence *facción* and there is a related verb is *satisfacer* 'satisfy', where in case one wanted to segment it, the base would be *hacer* and the first member would be a prefix, not a verbalised root.

This property of verbalisers is explained automatically if *ifica* is just a version of the theme vowel. An account where *ific* is the verbaliser itself has no way of blocking that the theme vowel, itself a distinct morpheme, can be absent.

A second argument in support of the view that the verbalisers are the theme vowels is that they never appear in light verbs. None of these affixes can attach to something to make a light verb, something that we expect given that light verbs occupy the position of the theme vowel.

The third argument is that the connection between the morphemes *ific, e, it, ec* and the theme vowel is univocal. It is never the case that, for instance, *ific* appears with the theme vowel *e* or *i* instead of *a* (34).

(34)   a.   *ificer
        b.   *icer
        c.   *eir

d. *itir

This is not logically necessary, given that for instance with nominalisers and adjectivalisers, Spanish allows that the morpheme responsible for the category assignment combines with more than one distinct marking. In (35), the nominaliser *-ez* combines with *a* marking feminine in the first case, but with null marking in the second.

(35)   a.   pobr-ez-a
            poor-ness-A
            'poverty'
       b.   tont-ez-ø
            silly-ness-ø
            'stupidity'

The pattern is however explained if in the case of theme vowels they are the real verbalisers, and an exponent like *ifica* is another spell out they can get. In the case of nominalisers and adjectivalisers, we assume—following the traditional approach—that they are themselves responsible for the category change and the final vowel is a higher functional head. This also explains that in the case of nominalisers and adjectivalisers this final vowel is systematically cancelled:

(36)   a.   ceremoni- os- o
            ceremoni- ous- O
            'ceremonious'
       b.   ceremoni- os- idad
            ceremoni- ous- ity
            'ceremoniosity'
       c.   *ceremoni- os- o- idad
            ceremoni- ous- O- ity
            Intended: 'ceremoniosity'

Note, incidentally, that *ifica* can be historically related to the light verb *facere* 'to make / do' in Latin (cf. Spanish *hacer*).

4.4   Conclusions and further prospects

There are, therefore, several empirical arguments that there should be total structural identity between theme vowels, verbalisers and light verbs in the sense that they identify the same position within the verbal structure. If these arguments are right, then we would be one step closer to reducing what seems to be surface, idiosyncratic morphological variation to well-defined syntactic or semantic properties of universal struc-

tures.

Note that we have made the argument that light verbs are theme vowels—or vice versa—based on the present tense. One reason for this is that in the past tense additional segments that could be identified as theme vowels (M. Oltra-Massuet 1999) emerge:

(37)   a.   er-a
             was-THV
             'It was.'
       b.   esta-b-a
             be-IMPF-THV
             'It was.'

The way in which we interpret these facts follows Oltra-Massuet in the claim that the structure of the Romance verb includes more than one position for the theme vowel. Instead of treating theme vowels as dissociated morphemes, however, we will push our preliminary findings and propose that in (37) the visible theme vowel is a light predicate corresponding to [PST]. The question of how many theme vowels a particular verb form carries, in this view, is a function of the height at which the verb is initially introduced in the structure. A light verb like *ser* or *estar* is introduced below the position for Past Tense, and therefore it lexicalises a chunk of structure that does not include the projection that defines Past. Therefore, by the Exhaustive Lexicalisation Principle, when the light verb combines with the [PST] head, an additional exponent—itself a theme vowel—has to be combined with it. A natural extension of this proposal is to check whether, as expected, the number and position of the theme vowels that light predicates of different types carry follows the Functional Sequence, as one would expect from this type of treatment. We hope to explore this issue in the immediate future.

Acknowledgment

References

Arad, Maya. 2003. Locality constraints on the interpretation of roots: the case of Hebrew denominal verbs. *Natural Language and Linguistic Theory* 21. 737–778.

Blevins, James. 2007. Conjugation classes in Estonian. *Linguistica Uralica* XLIII(4). 250–267.

Butt, Miriam. 1995. *The structure of complex predicates in Urdu* (Dissertations in linguistics). Stanford, Ca.: CSLI Publications.

Butt, Miriam. 2003. The light verb jungle. In *Harvard working papers in linguistics*, vol. 9, 1–49. Cambridge, MA: Harvard.

Butt, Miriam & Wilhelm Geuder. 2001. On the (semi)lexical status of light verbs. In Norbert Corver & Henk van Riemsdijk (eds.), *Semilexical categories: the function of content words and the content of function words*, 000–000. Berlin: Mouton de Gruyter.

Grimshaw, Jane & Armin Mester. 1988. Light verbs and theta-marking. *Linguistic Inquiry* 19. 205–232.

Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building 20: essays in linguistics in honor of Sylvain Bromberger*, 111–176. Cambridge, MA: MIT Press.

Harley, Heidi. 1995. *Subjects, events, and licensing.* Cambridge, MA: MIT dissertation.

Kayne, Richard. 2016. What is suppletive allomorphy? NYU.

Lin, Tzong-Hong. 2001. *Light verb syntax and the theory of phrase structure.* UC Irvine dissertation.

Lowenstamm, Jean. 2014. Derivational affixes as roots. In Artemis Alexiadou, Hagit Borer & Florian Schäfer (eds.), *The syntax of roots and the roots of syntax*, 230–259. Oxford: Oxford University Press.

Marantz, Alec. 1997. No escape from syntax: don't try morphological analysis in the privacy of your own lexicon. In Alexis Dimitriadis & Laura Siegel (eds.), *Proceedings of the 21st Annual Penn Linguistics Colloquium* (University of Pennsylvania Working Papers in Linguistics), 201–225. Philadelphia: University of Pennsylvania.

Marvin, Tatjana. 2002. *Topics in the stress and syntax of words.* Cambridge, MA: MIT dissertation.

Mohammad, Jan & Simin Karimi. 1992. Light verbs are taking over: complex verbs in Persian. In J.A. Nevis & V. Samiian (eds.), *Proceedings of wecol 5*, 195–212. Fresno: Fresno State University.

Oltra-Massuet, Isabel & Karlos Arregi. 2005. Stress-by-structure in Spanish. *Linguistic Inquiry* 36(1). 43–84.

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.

Solà, Jaume. 1994. Morphology and word order in Germanic languages. In Werner Abraham (ed.), *Minimal ideas: syntactic studies in the minimalist framework*, 217–251. Amsterdam & Philadelphia: John Benjamins Publishing Co.

5 How to Grill a Chicken (in the Nanosyntactic Oven)

Michaela Faltýnková & Markéta Ziková

Masarykova Univerzita, Brno

5.1 Why the Chicken Has to Be Grilled

In Czech, nominal inflectional morphology consists of a stem followed by a case ending. Prototypically, the stem appears invariable throughout a paradigm, as is illustrated in Table 5.1. The table shows the paradigms of three nouns, 'machine', 'world' and 'tray', each involving a phonologically constant stem, to which a set of particular case endings is attached. (Note that case endings sometimes do not surface phonetically, which is marked as Ø below. We will come back to these zero markers later.)[1]

Table 5.1: Czech declension (fragment)

	'machine,' SG	'machine,' PL	'world,' SG	'world,' PL	'tray', SG	'tray,' PL
N	[stroj]-Ø	[stroj]-e	[svjet]-Ø	[svjet]-y	[plat]-o	[plat]-a
A	[stroj]-Ø	[stroj]-e	[svjet]-Ø	[svjet]-y	[plat]-o	[plat]-a
G	[stroj]-e	[stroj]-ů	[svjet]-a	[svjet]-ů	[plat]-a	[plat]-Ø
D	[stroj]-i	[stroj]-ům	[svjet]-u	[svjet]-ům	[plat]-u	[plat]-ům
L	[stroj]-i	[stroj]-ích	[svjet]-u	[svjet]-ech	[plat]-u	[plat]-ech
I	[stroj]-em	[stroj]-i	[svjet]-em	[svjet]-y	[plat]-em	[plat]-y

From the perspective of stem invariance, the noun 'chicken' (displayed in table 5.2) is exceptional: its case endings are concatenated with three different stem forms, namely [kuře], [kuřet] and [kuřat], differing in the threefold alternation *e∼et∼at*.[2]

---

[1]Throughout this paper, examples are presented as follows: orthographic forms are in italics, [square brackets] and /slash brackets/ denote [phonetic] and /phonological/ form, respectively; morpheme boundaries are marked by hyphen-s and glosses are enclosed by 'simple quotation marks'. Note further that an acute accent marks vowel length in Czech orthography; a grapheme <í> thus corresponds to [iː] and so on. Only a long [uː] is marked by a superscript circle, i.e. <ů>.

[2]In fact, the paradigm involves yet another alternation: the stem final [t] alternates

Table 5.2: Declension of 'chicken'

	'chicken,' SG	'chicken,' PL
NOM	[kuře]-Ø	[kuřat]-a
ACC	[kuře]-Ø	[kuřat]-a
GEN	[kuřet]-e	[kuřat]-Ø
DAT	[kuřec]-i	[kuřat]-ům
LOC	[kuřec]-i	[kuřat]-ech
INS	[kuřet]-em	[kuřat]-y

The aim of our paper is to capture both the phonological relatedness of the stem forms and their morphosyntactic properties, as revealed in their distribution. The main idea, summarised in (1), is that there are just two stems underlyingly, one for the plural and one for the singular. The variation of the singular stem will then be analysed in purely phonological terms as the deletion of the final coda in the nominative/accusative.

(1)   The two stem proposal:
    a.   PL = [kuřat]
    b.   SG = [kuřet] (+ coda deletion = [kuře])

Moreover, we argue that the two stems are not independent of each other, but stand in a containment relation. In particular, we postulate a single affix /At/, which is shared by both stems, and thus present in all inflectional forms. In the plural part of the paradigm, the affix is followed by the case endings directly, and it surfaces in its lexical form, i.e. with a low vowel [a]. We propose that in the singular part of the paradigm, there is an additional high-vowel affix /I/, which intervenes between the /At/ and the case markers, and which undergoes coalescence with the vowel in /At/. The coalescence of the high vowel /I/ and the low vowel in the affix /At/ then gives rise to a mid vowel [e], seen across all singular forms. In other words, we propose that the actual surface alternation [a]~[e], which tells apart the plural and singular stems, is in fact an alternation between /At/ and a bi-morphemic sequence /At/+/I/. The whole analysis is graphically depicted in (2).

(2)   SINGULAR   PLURAL
      kuř-   /At-I/   -ending | kuř-   /At/   -ending
                 ↙   ↘ | ↓
               [et]   [e] | [at]

_____
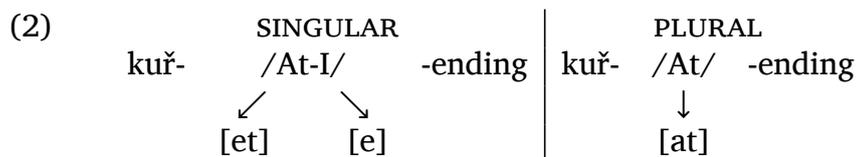
with a palatal [c] in the dative/locative singular form. However, the palatalisation is a regular phonological process triggered by the case-marking front vowel, we thus ignore it.

The proposed analysis makes two strong predictions concerning both the phonology and morphosyntax. First, if containment of forms indicates containment in terms of the syntactic structure (which is one of the core ideas of a cartographic approach developed by Cinque & Rizzi 2008 or Starke 2009), then the syntactic tree of the plural [at]-stem has to be a subpart of the singular [et]-stem. We therefore propose that both stems involve a shared piece of the syntactic structure which is spelled out by the affix /At/. The singular [et]-stem has in addition a singular-number head spelled out by the /I/-affix. The idea that the singular stem is morphologically more complex than the plural stem is elaborated on in section 5.2.[3]

On the phonological side, the two-stem analysis leads to the conclusion that there must be two phonologically distinct zero endings. The reason is that the consonant of the affix /At/ surfaces before the genitive plural zero (i.e. *kuř*[at]-Ø), but not before the nominative/accusative singular zero (i.e. *kuř*[e]-Ø). This indicates that despite appearances, the stems appear in two different phonological environments. In section 5.3, we show how such an analysis works and argue that postulating two phonologically distinct zeroes is not an ad hoc solution for this particular paradigm, but must be acknowledged independently of the facts at hand.

5.2 The Chicken Plucking Machine with a (Nano)syntactic Engine

The main idea we are going to develop in this section is that 'chicken' type nouns are compounds made up of two fully-fledged syntactic trees that bottom out in a root projection. The right-hand tree is then spelled out by the affix /At/, the left hand tree by the actual root. If a singular-number head is merged on top of the whole compound structure, the spell out adds the affix /I/. Let us now motivate the proposals in detail.

As is illustrated in table 5.3, 'chicken'-type nouns typically denote young animals. Moreover, we can see that these immature nouns (as we will call them) are derived regularly from nouns denoting the kind of the animal. The kind name is (most of the time) homophonous with that of mature individuals representing the kind (e.g., an individual ostrich).

---

[3]We note that this analysis does not automatically entail that singular in general is more complex than plural. It could easily be the case that there is also a plural head in the structure, which is expressed jointly with case by the portmanteau suffixes (which are plural specific, recall table 5.2). Our analysis only says that the singular stem [kuřet] is more complex than the plural stem [kuřat].

Table 5.3: Immature nouns

THE THREE STEMS			GLOSS	KIND	GLOSS
pštros-[e],	pštros-[et],	pštros-[at]	'young ostrich'	pštros	'ostrich'
sokol-[e],	sokol-[et],	sokol-[at]	'young falcon'	sokol	'falcon'
labuť-[e],	labuť-[et],	labuť-[at]	'young swan'	labuť	'swan'
tuleň-[e],	tuleň-[et],	tuleň-[at]	'young seal'	tuleň	'seal'

By the logic of the cartographic approach, we take the containment of the forms to indicate an actual containment in terms of the syntactic structure. We thus postulate a syntactic head IM(MATURE), which distinguishes age-marked nouns like *pštros-e* 'young ostrich' from their unmarked cousins, i.e. *pštros* 'ostrich'.[4] The easiest way to get the surface forms would be to say that the IM-head is spelled out by the affix /At/. However, this scenario, which appears the most reasonable at the first glance, cannot account for the data in table 5.4. The table shows that the immature nouns can contain suppletive roots, which differ in unpredictable ways from the kind-denoting root. (The kind root is also found in nouns referring to mature individuals, with the exception of *tur* 'Bos taurus', cattle,' where the individual-denoting roots are also suppletive.)

Table 5.4: Suppletive immature nouns

THE THREE STEMS			GLOSS	KIND	GLOSS
tel-[e],	tel-[et],	tel-[at]	'calf'	/tur/	'B. taurus'
sel-[e],	sel-[et],	sel-[at]	'piggy'	/pras/	'pig'
štěň-[e],	štěň-[et],	štěň-[at]	'puppy'	/pes/	'dog'
jehň-[e],	jehň-[et],	jehň-[at]	'lamb'	/ovc/	'sheep'

Nanosyntax accounts for such a root suppletion by means of phrasal spellout and lexical relatedness. Roughly speaking, the suppletive root (e.g. 'puppy') is a portmanteau morpheme which spells out (i) the lexically-related root ('dog') and (ii) an additional piece of the syntactic structure, where the obvious candidate in our case would be the IM-head.

To illustrate the portmanteau-based approach, consider the two lexical entries in (3) below. The entry for the suppletive root /bbb/ in (3a) spells out (i) all the projections that the root /aaa/ does and (ii) an additional syntactic feature Y. In technical terms, the entry in (3a) contains the so-called pointer ($\Rightarrow$) that points to the entry for the root /aaa/. The latter entry is given in (3b). As a consequence, the suppletive root /bbb/ spells out the whole syntactic tree [$_{YP}$ Y [$_{XP}$ X]], and applies only in case

---

[4]Such an IM head resembles the dedicated AGE projection proposed in Scott (2002).

XP had been spelled out by /aaa/.

(3)   a. $</bbb/, [_{YP} Y \Rightarrow </aaa/>]>$
       b. $</aaa/, [_{XP} X]>$

From this perspective, the suppletive immature roots such as those listed in table (2) above are an instantiation of such portmanteau morphemes: their lexical entry spells out the IM-head and contains a pointer to the corresponding age-unmarked root. To give an example, consider the pair *tur* 'Bos taurus' and *tel* 'calf.' The lexical entry in (4a) says that the suppletive root /tel/ 'calf' is an immature version of the root /tur/ 'B. taurus,' whose entry is given in (4b).

(4)   a. $</tel/, [_{ImP} IM \Rightarrow </tur/>]>$
       b. $</tur/, [_{NP} NP]>$

However, suppletive nouns like 'calf' form only a small subset of all immature terms: by default, the immature nouns share their root with their non-immature cousins, as has been illustrated in 5.3. In these regular nouns (like *pštros-e* 'young ostrich'), the affix /At/ is the only marker of the immature meaning. This is why we have already proposed that the /At/-affix also spells out the IM-head. However, since this affix attaches also to the suppletive roots like /tel/ 'calf' in (4a), it inevitably follows that the IM-head, encoding the immature meaning, is present twice in the structure of the suppletive nouns like *tel-e* 'calf' (once in the root, and once in the affix). To put it in traditional terms, the immature suppletive nouns are instances of the so-called *extended exponence*.

The existence of such an extended exponence identified in the nouns like *tel-e* challenges the simple approach alluded to above, according to which the affix /At/ (spelling out IM) is simply merged on top of the root. The reason is that there is a general principle that rules out recursive features in a single functional projection (Starke 2004, De Clercq & Vanden Wyngaerd to appear). Given this general principle, the affix /At/ and the suppletive root, which the /At/ is attached to, cannot belong to a single *fseq*.

We solve this recursiveness problem by analysing immature nouns as syntactic compounds. In particular, we propose that immature nouns are made up from two fully-fledged syntactic trees which bottom at root projections and where each of the functional sequences contains the IM-head.

The compound' proposal is depicted in (5): here, two trees are joined together and each has the structure $[_{ImP}$ IM $[_{NP}$ N]]. The only difference between the two nouns in (5) is their "morphological' status: while the left-hand noun is realised by an open-class root morpheme, the right-

hand one by a morpheme that would traditionally be classified as an affix.[5]

(5)

root ⇐ ImP   ImP ⇒ affix

Im   NP   Im   NP

...   ...
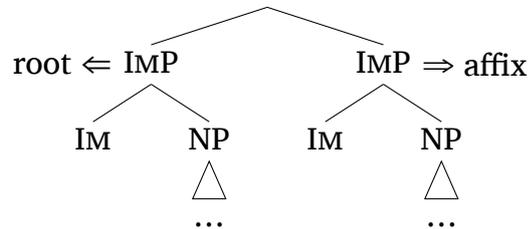
The lexical entry for this noun-like affix is given in (6): it involves a nominal-root phrase (NP) at the bottom that is dominated by the ImP. Paraphrased, the entry in (6) says that the affix /At/ denotes an immature object (without any specific conceptual meaning).

(6)   < /At/, [$_{\text{ImP}}$ Im [$_{\text{NP}}$ N]] >

Analyzing the affix /At/ as a noun with a general meaning 'immature object' not only solves the issue with suppletion, but enables us also to account for an additional class of immature nouns that are built on non-nominal bases. A couple of examples are given in table 5.5.

Table 5.5: Immature nouns from non-nominal bases

	The three stems			Gloss
a.	batol-[e],	batol-[et],	batol-[at]	'toddler'
b.	hol-[e],	hol-[et],	hol-[at]	'nestling'

On the first line, there is a manner-of-motion root *batol* 'toddle', which acquires the affix /At/ yielding the immature noun *batol-e* 'toddler'. In the b-example, the affix is attached to the adjectival root *hol* 'naked' producing the immature noun *hol-e* 'nestling' (literally 'naked youngster'). In effect, the affix /At/ serves as a nominaliser turning non-nominal bases to fully-fledged nouns (with the immature meaning), which is something that is expected from the syntactic structure associated with the affix. Technically speaking, the 'chicken'-type nouns like *batol-e* or *hol-e* are

---

[5]The proposal that affixes can realise fully-fledged syntactic structures (which include a root projection at the bottom) is not completely new – it has already appeared in the literature; see e.g. De Belder (2011) or Lowenstamm (2014), who have proposed something similar within a Distributed Morphology framework. In principle, both separate noun projections in (5) can be realised by (what a traditional morphologist would analyse as) a root morpheme, so that we obtain so-called coordinative compounds; these (two-rooted) immature compounds are typical, for example, for English, e.g. *tiger baby, lion cub* or Norwegian, e.g. *katt-unge,* lit. 'cat-youngster.'

syntactic structures where the nominal tree [$_{\text{ImP}}$ IM [$_{\text{NP}}$ N]], spelled by the /At/, is adjoined either to the verbal or the adjectival tree, respectively.

To this point, we have discussed 'chicken'-type nouns that denote immature animals or humans. However, given that the affix /At/ has the structure [$_{\text{ImP}}$ IM [$_{\text{NP}}$ N]], 'chicken'-type nouns without the immature meaning are also expected to exist. Their existence is predicted by the *Superset Principle* (Starke 2009, Caha 2009), according to which a given lexical item can spell out the syntactic structure equal to the size of that item – or smaller than it. In this particular case, the affix /At/ either spells out the immature tree [$_{\text{ImP}}$ IM [$_{\text{NP}}$ N]] or it spells out its subconstituent, i.e. the bare-noun projection. And in fact, there are nouns that contain the affix /At/ and that do not refer to immature objects. Table 5.6 provides a couple of examples, which include nouns denoting both animate objects (in 5.6.a) and inanimate ones (in 5.6.b).

Table 5.6: [-e/-et/-at] nouns without immature meaning

	THE THREE STEMS			GLOSS
a.	pras-[e],	pras-[et],	pras-[at]	'pig'
	morč-[e],	morč-[et],	morč-[at]	'guinea pig'
	zvíř-[e],	zvíř-[et],	zvíř-[at]	'animal'
b.	košť-[e],	košť-[et],	košť-[at]	'broom'
	rajč-[e],	rajč-[et],	rajč-[at]	'tomato'
	varl-[e],	varl-[et],	varl-[at]	'testicle'

Since the immature meaning is encoded in the syntactic IM-head (which is the central idea of our proposal), the 'chicken'-type nouns like *pras-e* 'pig' must differ syntactically from the immature nouns like *tel-e* 'calf'. We assume that both are syntactic compounds with the left-hand tree spelled out by the root and the right-hand tree spelled out by the affix /At/. Comparing the non-immature compound in (7) with its immature cousin in (8), the only difference is that the functional projections of the former lack the IM-head.

(7)

root ⇐ NP   NP ⇒ affix
         △   △
         ...   ...
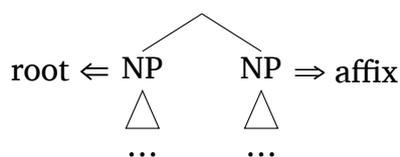
Having established the main syntactic ingredients of the 'chicken'-type nouns, let us return to looking at how these ingredients are spelled out by particular lexical items. The figure in (8) repeats the syntactic structure

69

of those 'chicken'-type nouns that have the immature meaning. Since the right-hand tree [$_{\text{ImP}}$ IM [$_{\text{NP}}$ N]] is invariably spelled out by the affix /At/ we will focus on the spell out of the left-hand tree below.

(8)

```
            root ⇐ ImP              ImP ⇒ affix
                  /    \            /    \
                IM      NP        IM      NP
                        △                 △
                       ...               ...
```
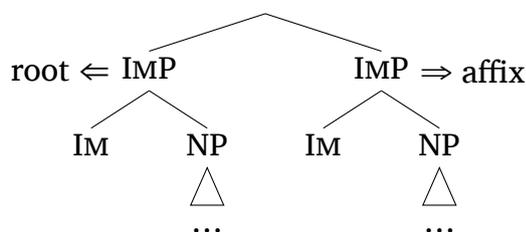
In Nanosyntax (see in particular Starke 2018), spell-out proceeds cyclically from the bottom up. Cyclicity means that each syntactic Merge triggers the scanning of the lexicon for an appropriate lexical item (in the sense of the *Superset Principle*). If such an item is found, the particular phrasal node is spelled out by it. In the next step, Merge creates a one-feature bigger syntactic constituent and the lexicon is scanned again. If an appropriate item is found (i.e. an item whose syntactic structure includes the structure of this newly created constituent), the form is inserted and this newly inserted form then overwrites the form inserted previously.

From this perspective, the suppletive immature nouns like *tel-e* 'calf' differ from the regular nouns like *pštros-e* 'young ostrich' only in how many lexical items are inserted during their spell out, as is illustrated in (9).

(9)   a.   ImP ⇒ /tel/   b.   ImP ⇒ /pʃtros/
              /   \   /   \
            IM   NP ⇒ ~~tur~~   IM   NP ⇒ ~~pʃtros~~
                       △   △
                      ...   ...
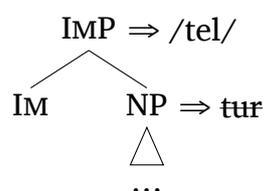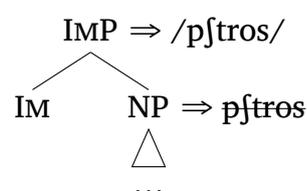
The figure in (9a) depicts the cyclic spell out of the left-hand tree of the immature noun 'calf' that involves two separate lexical items. During the first spell-out cycle, the root /tur/ is inserted into the NP. Its entry is in (10a). On the next cycle, i.e. after the merge of the immature feature, the root /tur/ is overwritten by the suppletive root /tel/, whose lexical structure is in (10b). The entry spells out the IM-head and contains a pointer to the root /tur/. By contrast, in the regular noun *pštros-e* 'young ostrich' in (9b), both spellout cycles are spelled out by the same lexical item, i.e. the root /pʃtros/. Since it is stored with the whole immature tree [$_{\text{ImP}}$ IM [$_{\text{NP}}$ N]] (see (10c)), it can spell out both the root projection NP as well as its merger with the IM-head. As a consequence, the root /pʃtros/ is shared by the non-immature noun *pštros* 'ostrich' (in which it

70

spells out the bare NP) and the immature noun *pštros-e* 'young ostrich' (in which the form /pʃtros/ spells out the [ImP IM [NP N]]).

(10)   a.   </tur/, [NP NP]>
       b.   </tel/, [ImP IM ⇒ </tur/>]>
       c.   </pʃtros/, [ImP IM [NP N]]>

Summing up, the suppletive and the regular immature nouns differ only in the cyclic spellout of the root tree: one root is replaced by the other in the suppletive nouns, while in the regular nouns, the single root is reinserted.

By comparing immature and non-immature nouns, we have identified the following two types of relation between their roots: (i) the root appears identically in both nouns (e.g. [pʃtros] 'ostrich' – [pʃtros]-e 'young ostrich'), (ii) the roots differ completely in their form (e.g. [tur] 'B. Taurus, cattle' – [tel]-e 'calf'). In addition, there exist also cases where the immature and the non-immature root are phonologically related (even though not identical). The table in 5.7 illustrates such pairs. There are two types of alternations that are involved in immature roots (these are on the left): (i) alternations in vowel length, (ii) palatalization alternations. In particular, there are three animal-denoting roots listed in table 5.7, whose immature variants display long vowels and palatalized consonants (i.e. either a palatal [ɲ] or a postalveolar [tʃ]). The non-immature forms in the right part of the table, on the other hand, show short vowels and non-palatalized consonants (i.e. an alveolar [n] and a velar [k]).

Table 5.7: Phonologically related immature nouns

THE THREE STEMS			KIND	GLOSS
sl[uːɲ]-[e]	sl[uːɲ]-[et]	sl[uːɲ]-[at]	sl[on]	'elephant'
klok[aːɲ]-[e]	klok[aːɲ]-[et]	klok[aːɲ]-[at]	klok[an]	'kangaroo'
r[aːtʃ]-[e]	r[aːtʃ]-[et]	r[aːtʃ]-[at]	r[ak]	'crawfish'

Since root pairs like /tel/–/tur/ are undoubtedly suppletive, the pairs involving segment alternations in table 5.7 can be, in principle, treated either suppletively or derivationally. In the latter case, only the non-immature root form will be stored and the immature one will be derived from it by phonology. To decide which of these two scenarios is more plausible, consider the examples in table 5.8. They show that palatalisation alternations are independent of vowel length alternations.

In sum, all roots ending in alveolars and velars appear palatalised in the context of the affix /At/, while only a subset of them shows lengthened vowels. To account for this asymmetry, we thus assume that

Table 5.8: Palatalisation alternations

THE THREE STEMS			KIND	GLOSS
biz[oɲ]-[e]	biz[oɲ]-[et]	biz[oɲ]-[at]	biz[on]	'buffalo'
lacht[aɲ]-[e]	lacht[aɲ]-[et]	lacht[aɲ]-[at]	lacht[an]	'sea lion'
hr[oʃ]-[e]	hr[oʃ]-[et]	hr[oʃ]-[at]	hr[ox]	'hippo'

the only active phonological process involved in the pairs like /sluːɲ/–/slon/ in table 5.7 is the palatalisation. As a consequence, there are two lexical entries for the root 'elephant' in (11), which (i) share the final consonant, i.e. the [n] in this particular case, and (ii) differ in vowel length. The long variant of the root in (11a) is then specified as spelling out the immature meaning.

(11)   a.   </sluːn/, [$_{ImP}$ IM ⇒ </slon/>] >
        b.   </slon/, [$_{NP}$ NP] >

To conclude, vowel length alternations found in immature forms are suppletive (because they are unproductive, i.e. they apply or fail to apply on a case-by-case basis), as opposed to palatalization alternations that are derived in the phonological component (because they are completely general). Since all 'chicken'-type nouns contain the affix /At/, we consider it as the palatalisation trigger.[6]

As a final point, let us address the morphosyntactic aspects of the stem allomorphy in the 'chicken'-paradigm. Recall the basic observation: the stem variation in a declension is number-dependent: the plural forms exhibit the [at]-stem and the singular forms the [et]-stem. In principle, two possible analyses can be proposed: one where the singular [et] is the default form and [at] is specified for plural, or the other way round. According to our analysis, the [at]-stem forms (which can be seen in the plural) are going to be number-unmarked: we propose that the affix /At/ itself is not associated with the plural feature. This proposal is motivated by the fact that [at]-stem forms regularly appear not only in the plural inflectional forms, but also in diminutives, as is shown in table 5.9 below. The diminutives of 'chicken'-type nouns (in the second column in the table) are all derived by the diminutive *k*-suffix attached to the [at]-stem; the final *-o* is a case marker. If the affix /At/ were

---

[6]In this paper, we remain agnostic as for the technical implementation of the idea that the non-front vowel affix serves as a palatalisation trigger. Note, however, that /At/ is not the only affix of this type. For example, the *a*-initial suffix *-an*, which derives inhabitant names such as *Kana*[ɟ-an] 'inhabitant of Canada' or *Newyor*[tʃ-an] 'inhabitant of New York', also triggers palatalisation of alveolars and velars, i.e. [d] and [k] turn to [ɟ] and [tʃ] in this particular case.

to spell out a number projection (encoding plurality), the distribution of the [at]-stem forms illustrated below could not be explained in a uniform way.[7]

Table 5.9: Diminutives

NOM.PL	DIMINUTIVE	GLOSS
ku[ř-at]-a	ku[ř-aːt]-k-o	'chicken'
slů[ɲ-at]-a	slů[ɲ-aːt]-k-o	'elephant'
pra[s-at]-a	pra[s-aːt]-k-o	'pig'

The [at]-stem forms thus contrast with the [e]-stem forms (i.e. *kuř*[e] and *kuř*[et]), which will be analyzed as arising from the addition of the number-marking affix /I/. In (12), we propose that the affix /I/ spells out the singular-number projection (SGP) which is merged only with the whole compound structure corresponding to the [at]-stem.

(12)   [et]-stem

    [at]-stem   SGP ⇒ /I/

      △

      ...
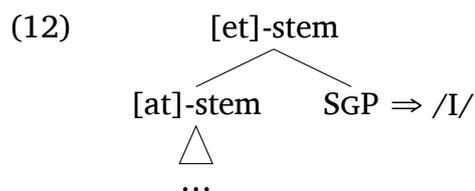
To sum up the morphosyntactic part of our analysis, the 'chicken'-type nouns are derived by two lexical items, namely the affixes /At/ and /I/. The affix /At/ is present in all nouns that belong to the 'chicken'-paradigm. The affix /I/ appears only in their singular forms and its merger with the affix /At/ leads to the *at~et* stem variation.

5.3 Gutting the Chicken with a Phonological Knife

In the previous section, we have elaborated on the syntactic structure of the two affixes that participate in the derivation of the 'chicken'-type nouns. This section is devoted to their phonology. Following the central ideas of Element Theory (Backley 2011) and Strict CV (Scheer 2004), we decompose melodic structure into privative elements and prosodic structure into strictly alternating C- and V-slots. With the background in place, we shall argue that both actual stem forms found in the singular part of the 'chicken'-paradigm, i.e. the [e]-form (in the nominat-

---

[7]The stem *a*-vowel is lengthened in the diminutive forms. The vowel lengthening is however a regular process triggered before the diminutive suffixes.

ive/accusative) and the [et]-form (elsewhere), arise from phonological computation applied to the (number-unmarked) [at]-stem.

The Element Theory postulates a containment relationship between the vowels [a i u]—the corners of the classical vocalic triangle—and other vocalic types. The idea is that the corner vowels are simplex segments, each involving exactly one of the three main melodic primes, namely the elements A, I and U. Their elemental makeup is summarised in (13). Furthermore, the figure in (13) demonstrates that mid vowels [e] and [o] are elemental compounds: both contain the element A, which is combined with either I (yielding a front mid vowel [e]) or U (in a back mid vowel [o]).

(13)　　　[i]　　/I/　　　　　　　　　/U/　　[u]
　　　　　　　　　　　↘　　　　　↙
　　　　　　　[e]　　/IA/　　　/IA/　　[o]
　　　　　　　　　　　↖　　　↗
　　　　　　　　　　　　/A/
　　　　　　　　　　　　[a]
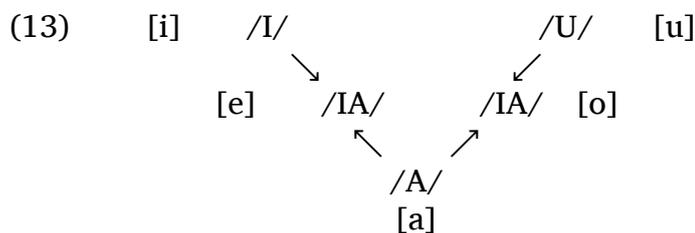
From the perspective above, the vocalic alternation a~e, which distinguishes the number-unmarked stem from the singular stem, can be viewed as involving an addition of the element I to the simplex low vowel [a], i.e. /A/ [a] + /I/ → /AI/ [e].

The phonological derivation of the e-alternant is depicted in figure (14). The figure shows lexical representations of the /At/-affix (14a) and the /I/-affix (14b). The figure in (14c) then depicts their phonological merger which is spelled out as [e]. (We will return to the status of [t] later.)

(14)　　a.　V　　　　　　　b.　　　　　　　c.　V
　　　　　　　|　　　　　　　　　　|　　　　　　　　|＼
　　　　　　　A　t　　　　　　　　I　　　　　　　A　t　I
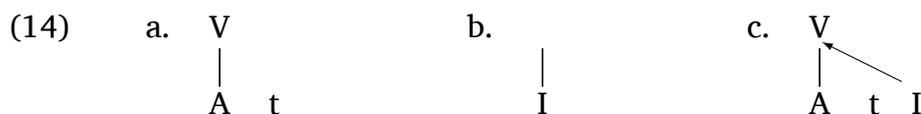
The representations above assume that phonological structure contains three ingredients, namely the melodic elements, the prosodic slots and the association lines. The /At/-affix contains all these three building blocks: its representation in (14a) involves the V-slot and the A-element, which are connected via an association line. The /I/-affix in (14b) contains the I-element provided with an association line, which is, however, not linked to any V-slot. The presence of the association line in the lexical representation ensures that the /I/-affix automatically associates with the V-slot provided by the /At/-affix. After the concatenation of the /I/-affix, the complex vowel /AI/ is thus derived in (14c) that is spelled out as [e].

The proposal that the mid vowel [e] in the singular stem arises from concatenation of the floating /I/-affix is further supported by the pattern in table 5.10. There are three nouns ('frog', 'bud' and 'peacock') with the following two properties: (i) they belong to the 'chicken'-paradigm, (ii) they involve labial-final roots. What we see in table 5.10 is a contrast between plain root-final labials [b p v], found in the plural, and their singular counterparts, which appear palatalised, with a palatal glide [j] following the labial (the actual forms are [bj pj vj]).

Table 5.10: Czech declension (fragment)

	'frog,' SG	'frog,' PL	'bud,' SG	'bud,' PL	'peacock,' SG	'peacock,' PL
N	žá[bj]-e-Ø	žá[b]-at-a	pou[pj]-e-Ø	pou[p]-at-a	pá[vj]-e-Ø	pá[v]-at-a
A	žá[bj]-e-Ø	žá[b]-at-a	pou[pj]-e-Ø	pou[p]-at-a	pá[vj]-e-Ø	pá[v]-at-a
G	žá[bj]-e-te	žá[b]-at-Ø	pou[pj]-et-e	pou[p]-at-Ø	pá[vj]-et-e	pá[v]-at-Ø
D	žá[bj]-et-i	žá[b]-at-ům	pou[pj]-et-i	pou[p]-at-ům	pá[vj]-et-i	pá[v]-at-ům
L	žá[bj]-et-i	žá[b]-at-ech	pou[pj]-et-i	pou[p]-at-ech	pá[vj]-et-i	pá[v]-at-ech
I	žá[bj]-et-em	žá[b]-at-y	pou[pj]-et-em	pou[p]-at-y	pá[vj]-et-em	pá[v]-at-y

The contrast between the singular and plural stems regarding labials can plausibly be explained as the result of I-affixation: the singular stems in table 5.10 involve the floating I-affix, which is what triggers the palatalisation of the root-final labials. In other words, the floating /I/-affix materialises in two ways in singular stems: (i) as a mid vowel [e], (ii) as a palatal glide [j] (after labial-final roots).

In fact, there are two palatalisation processes triggered in the 'chicken' type nouns, which must be treated separately. The contrast is between the labial-final roots on the one hand and the roots ending in alveolars and velars, discussed previously, on the other. The two palatalisation patterns are summarised in table 5.11 below. The table compares the genitive singular and the nominative plural forms of several 'chicken'-type nouns. The root-final alveolars and velars in table 5.11.a undergo palatalisation in both the singular [et]-stem (in the first column of the table) and the plural [at]-stem (in the second column). By contrast, the root-final labials in 5.11.b become palatalised only in the singular [et]-stem.

According to our analysis, the two palatalisation patterns above arise in two different contexts: the alveolar-velar pattern in table 5.11.a arises before the affix /At/, the labial pattern in table 5.11.b before the affix /I/. We propose that the uniform behaviour of alveolars and velars as opposed to labials can be captured by the following scale. Without going into phonological details, the scale in (15), is a one-way implicational hierarchy going from right to left. Its purpose is to derive three basic palatalisation patterns found in Czech: (i) all of labials, alveolars and velars undergo palatalisation, (ii) alveolars and velars undergo palatalisation,

Table 5.11: Two types of palatalisation

	GEN.SG	NOM.PL	ROOT	GLOSS
a.	krů[c-et]-e	krů[c-at]-a	krů[t]	'turkey'
	hro[ʃ-et]-e	hro[ʃ-at]-a	hro[x]	'hippo'
	vl[tʃ-et]-e	vl[tʃ-at]-a	vl[k]	'wolf'
b.	holou[bj-et]-e	holou[b-at]-a	holu[b]	'pigeon'
	čá[pj-et]-e	čá[p-at]-a	čá[p]	'stork'
	žel[vj-et]-e	žel[v-at]-a	žel[v]	'turtle'

or (iii) only velars are palatalised in the particular context.

(15)   velars   >   alveolars   >   labials
         [k g x]   [t d n]   [p b f v m]

Given the scale above, it is expected that there may be a suffix (like /At/) which only triggers the palatalization of alveolars and velars. However, in addition, we have to propose that the singular-marking affix /I/ affects labials, as was shown in table 5.11.b. Since labials constitute the lowest point of the palatalisation scale, we have to conclude that alveolars and velars undergo palatalisation process also before the /I/. However, the process is applied vacuously because roots ending in alveolars and velars already enter the singular stem with palatalised consonants.

Having explained the *a~e* alternation, which tells apart the stem unmarked for number and the singular stem, let us proceed to the second alternation found in the 'chicken'-paradigm, namely the *t~zero* alternation. Recall that the *t* of the affix /At/ is realised everywhere except in the nominative/accusative singular form. In approaching this process, we build on the insight of the Strict CV theory that segments alternating with phonetic zero are lexically floating pieces of melody. They get realised phonetically only when they receive association to the prosodic structure during the phonological computation. From this perspective, the contrast between the floating *t*-melody, which is not spelled out phonetically, and the associated *t*-melody surfacing as [t] will be as in (16).[8]

(16)   a.   ⇔   []   b.   C   ⇔   [t]
                                               |
              t   t

We propose that the floating *t*-segment is a lexical part of the affix /At/, which is involved in all inflectional forms in the 'chicken'-paradigm. A

---

[8]We do not depict melodic structure of the *t*-segment which in fact involves an ordered set of elements. Generally, we represent elemental makeup of segments only if the need arises.

question therefore arises why the C-slot, which makes the *t*-segment of the /At/-affix pronounceable, is inserted in all inflectional forms except in the nominative/accusative singular (in which the *t* is silenced).

Viewed from the surface perspective, *t*-realisation is triggered in disjunctive phonological contexts: the *t*-segment is pronounced before case markers beginning with a vowel (e.g. NOM.PL *kuřa*[t-a]), and also word-finally in the GEN.PL (*kuřa*[t#]). To make things (apparently) worse, there is an overlap between the latter context and the one where the *t*-realisation does not take place. In particular, while the *t*-segment is realised in the genitive plural form *kuřat*, it is not realised in the nominative/accusative singular form *kuřet*, despite the fact that the affix /At/ appears in the word-final position in both of these cases. To solve this puzzle, we propose that the contexts in which the *t*-segment is pronounced are unified at a more abstract level. We thus argue that the C-insertion (which makes the *t*-segment pronounceable) is triggered when the affix /At/ is followed by a V-slot on the prosodic level. This V-slot is either filled up with a melody, as in the nominative plural form *kuřata* in (17a), or it is empty, as in the genitive plural form *kuřat* in (17b). In other words, we assume that the genitive plural zero marker is in fact an abstract vocalic slot, which is the reason why the stem surfaces with [t] despite the fact that it appears in a word-final position in this morphological context. In both configurations in (17), the C-slot is inserted in order to repair two adjacent V-slots, which produce an ungrammatical prosodic string according to the Strict CV theory.

(17)   a.   C   V   C   V   C   -   V
              | | | | ↑ |
              k   u   ř   A   t   A

        b.   C   V   C   V   C   -   V
              | | | | ↑
              k   u   ř   A   t

By contrast, the *t*-segment does not surface in the nominative/accusative singular form, which also lacks an overt case marker. Translated into our phonological vocabulary, it means that the C-slot is not inserted in this particular morphological context. Given that the C-insertion is triggered only when a V-slot is attached to the affix /At/, we have to conclude that there is no empty V-slot in the nominative/accusative singular form. In other words, we propose that the two zero markers in the 'chicken'-paradigm differ: the genitive plural zero is an empty V-slot, while the nominative/accusative zero has no phonological structure at all.

The phonological computation of the *t*-less form *kuře* thus involves

two steps: (i) the concatenation of the /At/-affix in (18a), and (ii) the subsequent concatenation of the floating /I/-affix in (18b). Its association line docks automatically in the V-slot of the preceding affix (and this merger is spelled out as [e]).

(18)   a. C V C - V   b. C V C V   -

               k u ř A t   k u ř A t   I

The analysis thus assumes that case-marking zeroes differ phonologically. This is not an ad hoc solution for the t~zero alternation in the 'chicken'-paradigm, since it simultaneously opens the door for understanding other types of alternations in zero-derived inflectional forms. For example, word-final liquids alternate between postconsonantal and postvocalic depending on the morphological context. The alternating pattern, which is described in detail in Ziková (2013), is exemplified in (19).

(19)

	MASCULINE		FEMININE
NOM.SG	Pe[tr]-Ø	GEN.PL	Pe[ter]-Ø
GEN.SG	Pe[tr]-a	NOM.PL	Pe[tr]-a

In (19) we see the first name *Petr*, which ends in a *tr*-cluster. It has two gender paradigms (masculine and feminine), both involving zero case markers that appear in the same cases as in the 'chicken'-paradigm. Specifically, the nominative singular zero appears in the masculine paradigm and the genitive plural zero in the feminine paradigm. The pattern in (19) then demonstrates that the word-final *tr*-cluster is preserved in the context of the nominative singular (*Pe[tr]-Ø*), but it is broken by an epenthetic vowel [e] in the genitive plural (*Pe[ter]-Ø*). As a consequence, the two zero-derived inflectional forms are not syncretic as one would expect if they were fully identical.

    This effect is not due to homophony avoidance. To see this, consider the form *Pe[tr-a]*, where the cluster is followed by the vocalic ending -*a*. This form is syncretic between the genitive singular (in the masculine paradigm) and the nominative singular (in the feminine paradigm). This shows that there is nothing wrong about a masculine-feminine syncretism per se, and hence the reason for the differential behavior of the zero-derived form must be found elsewhere.

    Following the same logic as in the previous case, we take the fact that consonant-liquid (CL) clusters behave differently before zero endings to indicate that the nominative singular zero and its genitive plural counterpart have different underlying structures. Without going into the

technical details, we simply assert that the alternation in CL-stems can also be derived from the underlying contrast between the nominative zero (which we take to be phonologically empty) and the genitive plural zero (which corresponds to an abstract vocalic slot).

To sum up our phonological proposal, the two alternations that distinguish particular stem allomorphs in the 'chicken'-paradigm, i.e. *a*∼*e* and *t*∼zero, arise from regular phonological computation applied to two stem-building affixes /At/ and /I/. The crucial assumption is that lexical entries for these affixes involve incomplete phonological structure. Specifically, the following two phonological operations are applied: (i) the merger of the floating I-element which produces a complex vowel /AI/ (spelled out as [e]), (ii) the linking of the floating *t*-segment which makes it pronounceable as [t].

5.4   The Chicken on the Grill

In Czech, 'chicken'-type nouns are exceptional in displaying different stems in a single inflectional paradigm. We have approached the stem allomorphy in terms of morphological containment. Put simply, we propose that the singular stem contains the plural stem: it is derived from it by the number-marking affix. By contrast, the affix in the plural stem does not mark the plural number, but rather expresses the immature meaning. Analyzing the plural stem as a number-unmarked form enables us to capture its distribution in diminutives, where it appears to the left of the diminutive suffix.

In our account, the two affixes differ syntactically in a significant way. The immature affix corresponds to a fully-fledged syntactic tree that bottoms out in a nominal root projection. The singular-number affix, on the other hand, corresponds to a piece of the nominal functional projection. This approach opens the door for rethinking the traditional terms such as affix vs. root or derivational vs. inflectional morphology (De Belder 2011, Lowenstamm 2014) using the tools that came along with the Nanosyntactic framework (Starke 2009, Starke 2018).

References

Backley, Phillip. 2011. *Introduction to element theory*. Edinburgh: Edinburgh University Press.

Caha, Pavel. 2009. *The nanosyntax of case*. CASTL, University of Tromsø dissertation. `lingbuzz/000956`.

Cinque, Guglielmo & Luigi Rizzi. 2008. The cartography of syntactic structures. *STiL – Studies in Linguistics, CISCL Working Papers* 2. 42–58. `http://ciscl.media.unisi.it/doc/doc%7B%5C_%7Dpub/STiL-2008-vol2.pdf`.

De Belder, Marijke. 2011. *Roots and affixes: eliminating lexical categories from syntax*. Utrecht: Utrecht University dissertation.

De Clercq, Karen & Guido Vanden Wyngaerd. to appear. Negation and the functional sequence. *Natural Language and Linguistic Theory*. `lingbuzz/003794`.

Lowenstamm, Jean. 2014. Derivational affixes as roots. In Artemis Alexiadou, Hagit Borer & Florian Schäfer (eds.), *The syntax of roots and the roots of syntax*, 230–259. Oxford: Oxford University Press.

Scheer, Tobias. 2004. *What is CVCV and why should it be?* Walter de Gruyter.

Scott, Gary-John. 2002. Stacked adjectival modification and the structure of nominal phrases. In Guglielmo Cinque (ed.), *Functional structure in DP and IP*, vol. 1 (The Cartography of Syntactic Structures), 91–122. Oxford University Press.

Starke, Michal. 2004. On the inexistence of specifiers and the nature of heads. In Adriana Belletti (ed.), *Structures and beyond: the cartography of syntactic structures, vol. 3*, 252–268. New York: Oxford University Press. `lingbuzz/000004`.

Starke, Michal. 2009. Nanosyntax. A short primer to a new approach to language. In Peter Svenonius, Gillian Ramchand, Michal Starke & Tarald Taraldsen (eds.), *Nordlyd 36: special issue on Nanosyntax*, 1–6. Tromsø: University of Tromsø. `lingbuzz/001230`.

Starke, Michal. 2018. Complex left branches, spellout, and prefixes. In Lena Baunaz, Karen De Clercq, Liliane Haegeman & Eric Lander (eds.), *Exploring nanosyntax*, 239–249. Oxford: Oxford University Press.

Ziková, Markéta. 2013. Not all zeros are the same: Phonology of zero case markers in Czech. In Alexander Podobryaev (ed.), *Formal approaches to Slavic linguistics. The second MIT meeting 2011*, 305–319. Ann Arbor: Michigan Slavic Publications.

6 Balancing between roots and thematic vowels

Richard Holaj

Masarykova Univerzita, Brno

6.1 Introduction

Most of the Czech present verb forms exhibit an analytical inflectional pattern. This pattern corresponds to the root, followed by the thematic vowel (*-e, -í* or *-á*) and by the φ-features (1a). For example, for the second person singular of the verb *holit* 'to shave', we get the form *hol-í-š*, where *-í* (*í-verb*) is the thematic vowel and *-š* is the second person suffix, see (1b). For the verb *dělat* 'to do' in (1c), we will get *děl-á-š*, with the same second person suffix, but a different thematic vowel *-á* (*á-verb*). The same goes for the verbs with the thematic vowel *-e* (*e-verbs*) such as *lézt* 'to climb', where we get form *lez-e-š*, again with the same φ-features but a different thematic vowel.

(1)   a. $\sqrt{}$-THEME-φ
      b.   hol-í-š   'you shave'
      c.   děl-á-š   'you do'
      d.   lez-e-š   'you climb'

The forms of these verbs belonging to three different verb classes thus share the same form of the φ-features, but differ in their thematic vowel. This analytical pattern of inflection is kept for most person-number combinations. It is most prolific in the traditional fourth conjugation (*í-verbs*) and in the fifth conjugation (*á-verbs*). In these classes, the pattern is kept in all forms except for the third person plural, where it is difficult to maintain the neat segmentation into a theme vowel and φ, where φ would be constant across the three columns (see (2)).

(2)

Conjugation class	I-III	IV	V
1.SG	-u	-í-m	-á-m
2.SG	-e-š	-í-š	-á-š
3.SG	-e	-í	-á
1.PL	-e-me	-í-me	-á-me
2.PL	-e-te	-í-te	-á-te
3.PL	-ou	-í	-a(j)-í

The form of the third person plural is determined by the thematic vowel. We always get the third person plural *-ou* for the thematic vowel *-e, -í* for the thematic vowel *-í* and *-a(j)í* for the thematic vowel *-á.*

The thematic vowel itself seems to be selected by the root. However, it is not the phonology of the root that determines the form of the thematic vowel, since we can get different thematic vowels in the same phonological context (after the same root). For example, the root *sed* 'sit' can be followed both by *-í,* as in *seď-í,* and by *-á,* as in *sed-á* (3a). Similarly, for *leh* 'lie' we get *lež-í–leh-á* (3b), and for *běh* 'run' we have *běž-í–běh-á* (3c).[1]

(3)   a.   seď-í   sed-á
            'he is sitting'   'he is sitting down'
      b.   lež-í   leh-á
            'he is lying'   'he is lying down'
      c.   běž-í   běh-á
            'he is running'   'he runs'

Additionally, the thematic vowels seem to be influenced by the form of the following morpheme not only in the third person plural, but also in other forms such as in the imperative or in verbal nouns (4).

(4)   Present thematic vowel correlation

root	3.SG/TV	3.PL	IMP.2.SG	deverbatives
hol 'shave'	-í	-í	ø	-ení
děl 'do'	-á	-a(j)í	-ej	-ání

In this paper I want to focus mostly on the third person plural of the *í-verbs* and *á-verbs.* In Section 6.2, I will elaborate on two possible analyses of these forms. The first analysis considers *-í* in the third person plural

---

[1]In these examples, *d* and *h* are palatalized by *í* to *ď* and *ž* respectively. But we still consider the roots to be phonologically the same underlyingly. The forms given in the text differ in their meaning, and so ultimately the differences between the thematic vowels will have to be linked to differences in the functional sequence, but for now the point is simply that the relationship between the vowel and the root is not phonological selection.

of the *á-verbs* to be different from that in the *í-verbs*. The second analysis considers them to be the same. In Sections 6.3 and 6.4, I will explore the imperative forms, where we can observe a pattern similar to the one in the third person plural. Based on these observations, Section 6.5 argues for the correctness of the single *-í* hypothesis, and explores the consequences of this analysis for the structure of the thematic vowels.

6.2 Third person plural in Czech

In the third person plural of the *á-verbs* such as *děl-a(j)-í* 'they do', it seems that the analytical pattern (root, thematic vowel, and φ-features) is kept. The thematic vowel and the plural are thus spelled out separately as *-a(j)* and *-í*: such a structure is shown in (5a). I am assuming that the *a(j)/á* alternation is governed by phonology and that *j* has a function of the hiatus breaker.[2]

We can clearly see the opposite (synthetic) pattern for present verbs with the thematic vowel *-e*, where we get a synthetic form for the third person plural. The thematic vowel and the plural are spelled out together as the portmanteau morpheme *-ou,* resulting in the structure (5b).

(5)   a.   Y   b.   X
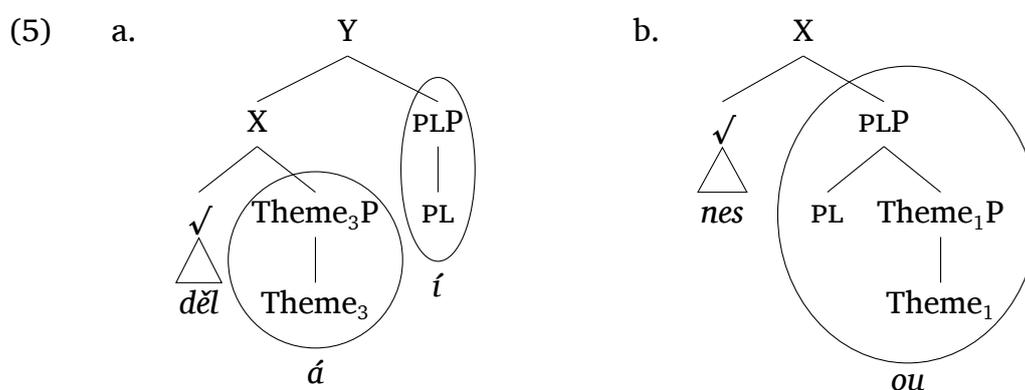
At present, we thus have two morphemes spelled out as *-í*: the 3PL marker and the thematic vowel of the *í-verbs*. At this point, it doesn't look like they could be unified, since they each realise different features.

However, there is yet another (i.e., third) context where we get the morpheme *-í*, namely the third person plural of the *í-verbs*. There are two possible analyses of this. First, the structure of the *-í* verb could correspond closely to structure of an *e-verb*. This is shown in the structure (6a), where *-í* is a portmanteau for the thematic vowel and the plural.

---

[2]Alternation appears in the context before *í*, so for *děl-á-í* 'they do' we get *del-a(j)-í.* This should not be confused with the words like *páj-í* 'they are soldering', where *j* is part of the root.

83

However, the structure of the *-í* verb could also correspond to structure of an *á-verb*. A structure along these lines is given in (6b), where we would have to assume that only one *-í* is pronounced as a result of phonology.

(6)　　a.

b.
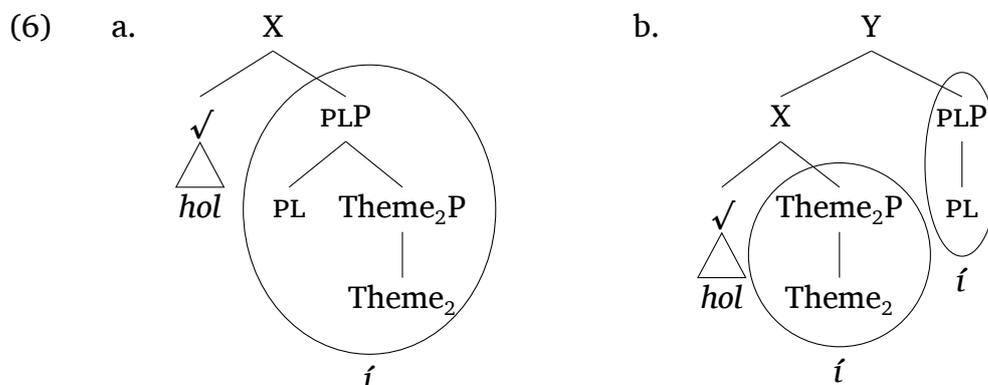
Both analyses need two lexical items for *-í*. The portmanteau analysis (6a) needs (7b) and (7c) (the latter for the 3PL marker in the *á-verbs* and the *e-verbs*). The analysis (6b) needs (7a) and (7b).

(7)　　a.　Theme₂P
　　　　　　|
　　　　　Theme₂

　　b.　PLP
　　　　　|
　　　　　PL

　　c.　PLP
　　　　　⟨PL　Theme₂P⟩
　　　　　　　　　|
　　　　　　　　Theme₂
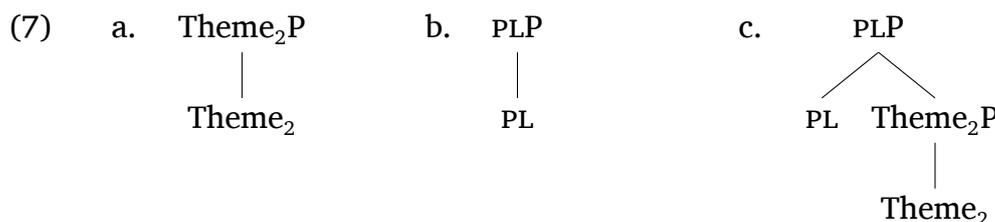
In the first analysis the features of the lexical items are in the subset/superset relation, but their structures do not fit the requirements of the Superset principle, since they do not share the same foot (Starke 2009). However, there is a way to unify them as (7c) , despite the fact that this may seem counterintuitive at first sight. To achieve this unification, we will first look at how the *í-verbs* and the *á-verbs* behave in the imperative, where we encounter an analogous situation.

6.3　The Czech imperative

There are several ways in which imperatives are realized in Czech verbs. These strategies are depicted in table 6.1. We can see there that the imperative may be just the root without any suffix, or the root with one of the following suffixes: *-i, -j, -ej*. In the first two cases (no affix or *i* suffix), the final consonant of the root can be palatalized. Palatalization happens when the final consonant is one from the set {t, d, n, k, g, ch,

h}. Palatalization is marked by ' in the table.[3]

Table 6.1: Variations in the Czech imperative

root	context	TV	IMP.2.SG	IMP. suffix
nes 'carry'	CVC	e	nes	ø
vrž 'creak'	CCL̦C	e	vrž	ø
ved 'lead'	CV[t|d|n|k|g|ch|h]	e	veď	'
kašl 'cough'	CVCC	e	kašli	i
vadn 'wither'	CVC[t|d|n|k|g|ch|h]	e	vadňi	'i
bi 'beat'	CV	(j)e	bij	j
sa 'suck'	CV	(j)e	saj	j
hol 'shave'	CVC	í	hol	ø
trp 'suffer'	CL̦C	í	trp	ø
chod 'walk'	CV[t|d|n|k|g|ch|h]	í	choď	'
mysl 'think'	CVCC	í	mysli	i
mast 'grease'	CVC[t|d|n|k|g|ch|h]	í	masťi	'i
děl 'do'	CVC	á	dělej	ej
vrt 'drill'	CL̦C	á	vrtej	ej
chyt 'catch'	CCV[t|d|n|k|g|ch|h]	á	chytej	ej
sedl 'saddle'	CVCC	á	sedlej	ej
lept 'cauterize'	CVC[t|d|n|k|g|ch|h]	á	leptej	ej
d 'give'	C	á	dej	ej

Based on the data, we can observe that *á-verbs* have the suffix *ej* independent of the phonological context, while in other classes, phonology plays a role.

Leaving palatalisation aside, we can see three possible variants of the imperative in the two latter classes: *ø, -i* and *-j*. Ziková (2016) proposes that they can all be represented as one underlying form /I/ (an abstract melodic segment) with three context-dependent realisations as shown in (8).

(8)   Realisation of Czech imperative of *í-verbs*

$$/I/ \rightarrow \begin{cases} \text{[j] / CV\_\_} \\ \text{ø / VC\_\_} \\ \text{[i] / VCC\_\_} \end{cases}$$

Based on (8), we can represent the imperative of the *í-verbs* (as well as the imperative of the *e-verbs*) as the root plus the imperative marker /I/, as seen in (9).

---

[3]L̦ in the environment CL̦C is syllabic consonant, which behaves as a vowel; see Zikova (2007).

(9)   a.   hol   - /I/ → hol!
             shave - IMP → 'shave!'
      b.   mysl - /I/ → mysli!
             think - IMP → 'think!'
      c.   sa   - /I/ → saj!
             suck - IMP → 'suck!'

If we look at the *á-verbs*, we can see that their imperative form always ends with the suffix *-ej* independently of the root's phonology. How can we fit this into our analysis? First we have to ask whether this suffix is a single morphemes or if it is bimorphemic.

If we accept the single morpheme hypothesis, we would either have to add an additional rule to (8), that would realize /I/ as *ej*, or assume a completely different allomorph. The first solution does not work, because we cannot specify the phonological context for this rule: recall that *-ej* does not depend on it. So this would lead to the system where the imperative suffix form is determined by the phonology and/or by the thematic vowel.

What would be the situation with the bimorphemic interpretation of *ej*? We would end up with the imperative morpheme realized as *j*. This morpheme would be preceded by the morpheme *e*, a vowel. This gives us the phonological context CV_ for all of the *á-verbs*. The rules in (8) then automatically give us [j] as the surface realization of the underlying imperative morpheme /I/ in context CV_ and that is exactly what we need.

From the perspective of the explanatory adequacy, the bimorphemic hypothesis of *ej* seems to be more plausible, since it doesn't require additional rules and it is able to derive the variation in the imperative only by the phonology. Adopting the bimorphemic nature of the suffixes gives us the segmentation (10) for the *á-verbs*:

(10)   děl-e-/I/ → dělej!
          do-α-IMP → 'do!'

Now we get an extra morpheme α in the *á-verbs*. In this paper we won't elaborate on the nature of this morpheme; rather we will show what are the consequences of its presence, more precisely we will explore what it tells us about the structure of the different roots.

## 6.4 The size of the roots in the Czech imperative

In the previous section, we have seen that the imperative is always spelled out by the morpheme /I/. Adopting Nanosyntactic constituent spellout (Starke 2009, Starke 2018), we know that one and the same morpheme always has the same foot, let's call it K. In addition, granted a universal *fseq* of an imperative, the imperative morpheme /I/ corresponds to the topmost feature or features, with K a foot, thus the first one of them (see (11)).

(11)   imperative fseq
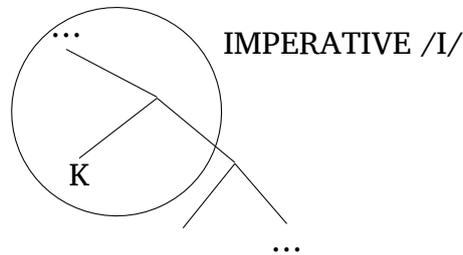
Features lower than K have to be realised by the remaining morphemes, i.e. the root in case of the *í-verbs* and the root plus the α-morpheme in case of the *á-verbs*. Since every morpheme has to spell out at least one feature, the structure below K is partially spelled out by the α-morpheme and partially by the root in case of the *á-verbs*, while in the *í-verbs*, the root spells out the whole structure (the same applies for the *e-verbs*).

This has one important consequence that can be easily seen when we compare (12) and (13). Roots of the *á-verbs* are smaller than the roots of the *í-verbs* (and the *e-verbs*), because features that are spelled out in *í-verbs* by the root only are spelled out partially by the *root* and partially by the α-morpheme in *á-verbs* (cf. Caha, De Clercq & Vanden Wyngaerd 2017).

87

(12)   Structure of the *í-verb hol-/I/*

IMPERATIVE /I/

…

K

ROOT *hol*

…
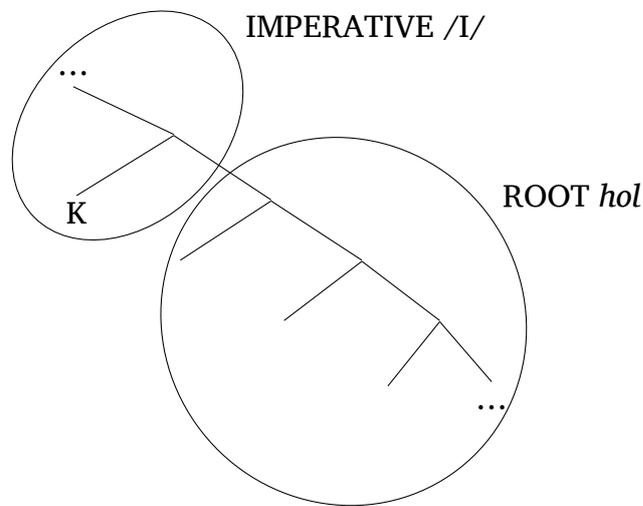
(13)   Structure of the *á-verb děl-e-/I/*

IMPERATIVE /I/

…

K

α-MORPHEME *e*

ROOT *děl*

…
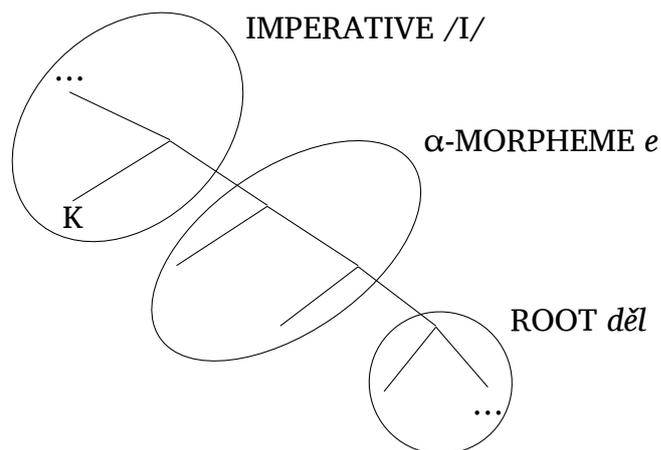
At this point we have support for the roots of *á-verbs* being smaller then the roots of *í-verbs*. For further analysis we will assume this to apply also for the other forms of those verbs, including the third person plural.

6.5   The third person plural and the thematic vowel

In the end of Section 6.2 we ended up with two morphemes *-í*, one for the thematic vowel of the *í-verbs* (respectively the thematic vowel plus the plural portmanteau) and one for the plural (seen after *a(j)* in *á-verbs*). For convenience, I repeat the relevant forms of the *í-verbs* (14) and the *á-verbs* (15) below, along with corresponding forms of the imperative.

88

(14) a. hol-í (shave-3SG) (15) a. děl-á (do-3SG)
      b. hol-í (shave-3PL)   b. děl-a(j)-í (do-3PL)
      c. hol-/I/ (shave-IMP)   c. děl-e-/I/ (do-α-IMP)

Assuming that *í* in the third person plural of the *í-verbs* and the *á-verbs* is the same (without any further assumption about its structure, except that the plural is the topmost feature or features), we will analogically to the analysis of the imperative in the Section 6.4 end up with the structures (16) and (17).

(16)

(17)
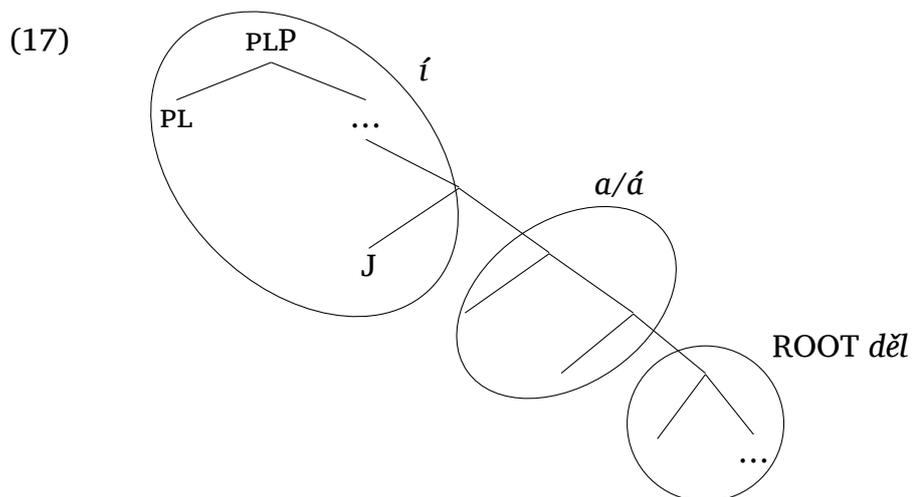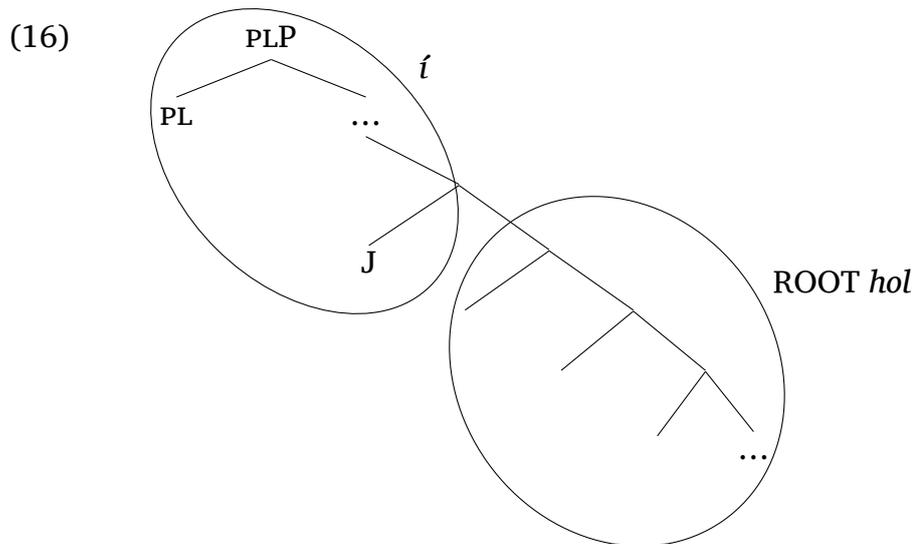
Notice that we have not said as yet which features correspond to the thematic vowel. To explore this, we need to consider how we can unify the third person singular forms with this analysis. Considering that the third person plural structure is built on top of the third person singular

by adding one or more plural features, we end up with the third person singular structure of the *í-verbs* such as (18), which is just like (16), but without PL on top.

(18)   Structure of hol-í (3SG)
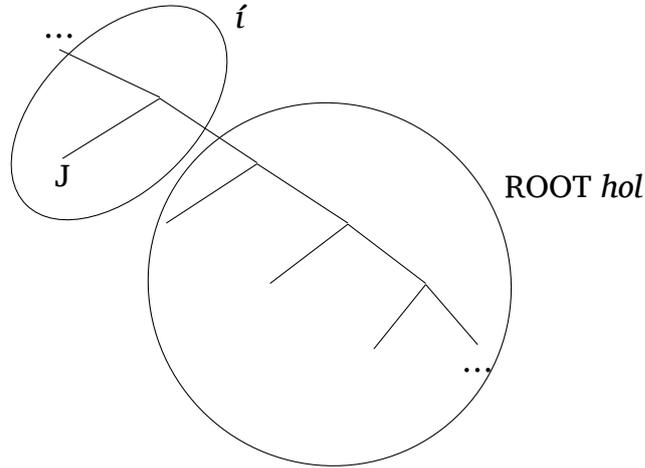
This structure can be spelled out by the same morpheme *-í* which we have seen in the plural structures. Its lexical entry would have a structure like (19).

(19)   Lexical item for *-í*
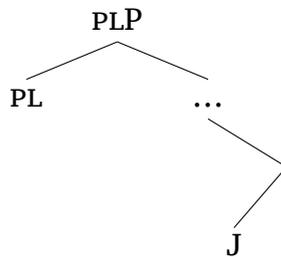
The plural in the *á-verbs* and the *í-verbs* has the same semantics, thus it corresponds to the same projections in the syntax and thus to the same features. Considering this and the structure of the third person plural of the *á-verbs* (17) we necesarily end up with the structure of the third person singular like (20).

90

(20)   Structure of děl-á (3sg)
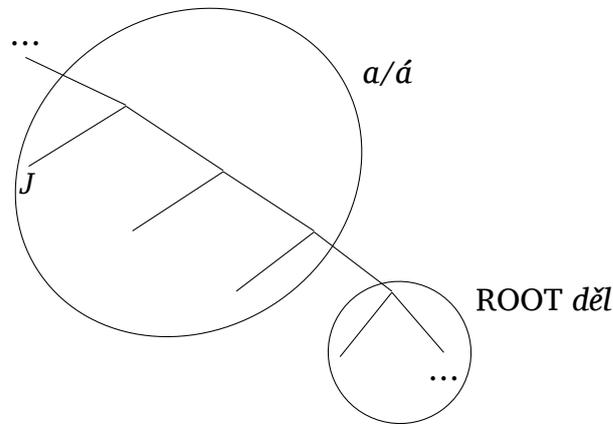
(21)   Lexical item for -á
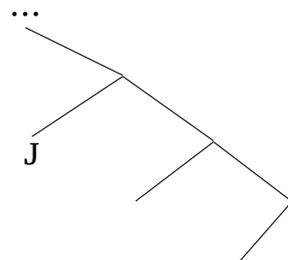
Notice at this point that the thematic vowel *á* has its foot necessarily lower than *J*, i.e. lower than where the thematic vowel *í* has its foot. However, the thematic vowel *á* stops lower than *í*: specifically, it does not spell out PL.

What consequences does this have for the structure of the thematic vowels in the third person singular? The conclusion is that independent of the full structure of the thematic vowels *-í* and *-á*, the thematic vowel *-í* seems to be contained in the thematic vowel *-á*. It also does not seem to be coincidential that the thematic vowel *á* spells out the same features as the thematic vowel *í* plus some of the lower features. Since the *á-verbs*, as we have shown, have smaller roots than the *í-verbs*, their thematic vowel *-á* has to spell out the features that are in the *í-verbs* already spelled out by the root (considering the same semantics of the third person plural of both verbs).

Having the lexical items (19) and (21), where only the former can spell out the plural, has several consequences. The first one is that *-í* can be represented by one lexical item: the thematic vowel plus the plural, as shown in (19) and as we suggested in Section 6.2. Also, we can derive the difference between the third person plural of the *í-verbs* and the *á-verbs* as a consequence of the thematic vowel *-á* not being able to spell out the plural.

I am assuming that in the third person plural of the *á-verbs*, when the full lexicalisation potential of *-á* is exhausted, and the plural feature merges on top of the tree, there is no lexical item to spell out just the plural. Therefore, backtracking applies (see Pantcheva 2011, Starke 2018), so that the derivation goes back all the way to the feature J; this feature, recall, still comes on top of the root and the foot of *-á* (21). From J on, *-í* applies, and spells out all the features going from J all the way to the top of the third plural tree. All considered together, this results into the suffix *-a(j)-í*, as seen in (17).

6.6   Conclusion

In this paper we have investigated the differences between the forms of the *í-verbs* and the *á-verbs* with a particular focus on the third person plural and the imperative. We have shown that the different forms of the imperative are caused by the phonology, and that suffix *-ej* that appears in the *á-verbs* imperative is bimorphemic, with *-j* being the realization of the imperative morpheme. The additional morpheme *-e* that appears in the *á-verbs* imperative is a consequence of the *á-verb* roots being smaller than the *í-verb* roots.

This difference between the root sizes causes the thematic vowel *-í* to be contained in the thematic vowel *-á*. This can be interpreted as the smaller *á-verbs* roots selecting the bigger thematic vowel *-á* and the bigger *í-verbs* roots selecting smaller thematic vowel *-í*.

This containment is visible in the third person plural of the *á-verbs*, where the higher part of the fseq that corresponds to the thematic vowel *-í* is spelled out separately along with the plural feature. The lower part remains accessible for insertion, and since it is still a subset of the thematic vowel *-á*, it is spelled out as *-a* resulting into the illusion of two thematic vowels being spelled out together. In the third person singular of the *á-verbs*, the structure corresponding to the thematic vowel *-í* is spelled out together with the lower features as *-á*. That is allowed by he absence of the plural. Furthermore this analysis allows us to unify *-í* as a single morpheme.

6.7   Acknowledgment

92

References

Caha, Pavel, Karen De Clercq & Guido Vanden Wyngaerd. 2017. The fine structure of the comparative. Ms. Masarykova Univerzita, U Ghent, KU Leuven. `lingbuzz/003790`.

Pantcheva, Marina. 2011. *Decomposing path: The nanosyntax of directional expressions*. Tromsø: University of Tromsø dissertation.

Starke, Michal. 2009. Nanosyntax: a short primer to a new approach to language. *Nordlyd* 36. 1–6.

Starke, Michal. 2018. Complex left branches, spellout, and prefixes. In Lena Baunaz, Karen De Clercq, Liliane Haegeman & Eric Lander (eds.), *Exploring nanosyntax*, 239–249. Oxford: Oxford University Press.

Zikova, Marketa. 2007. Czech syllabic consonants revisited. Paper presented at FDSL 7, Leipzig. `http://www.phil.muni.cz/cest/lide/zikova/FDSL%5C%207_hdt_final.pdf`.

Ziková, Markéta. 2016. When prosody follows syntax: verbal stems in Czech. *Linguistica Brunensia* 64. 163–185. `http://hdl.handle.net/11222.digilib/135456`.

7 A Recalcitrant Syncretism

Tarald Taraldsen

University of Tromsø

7.1 The Issue

In Romanian, a class of (neuter) nouns inflect like masculine nouns in the singular and like feminine nouns in the plural:[1]

(1)

		M	N	F
SG	NOM/ACC	N-u	N-u	N-ă
	GEN/DAT	N-u	N-u	N-e
PL	NOM/ACC	N-i	N-e	N-e
	GEN/DAT	N-i	N-e	N-e

Likewise, articles and modifiers agreeing with a noun in this class have the same form as articles and modifiers agreeing with a masculine noun in the singular, but the same form as articles and nouns agreeing with a feminine noun in the plural. I will use the term 'neuter nouns' to refer to the members of the relevant class.[2]

In addition, the F.SG.DAT/GEN is identical to the F.PL (in all cases).[3] This syncretism, however, plays only a marginal role in what follows. I will show in section 7.4 that a paradigm like (2), where all F.SG forms are identical to each other and distinct from the F.PL forms, presents exactly the same type of analytical challenges as the paradigm (1).

---

[1] I proceed on the assumption that the M/N.SG ending is *-u*, which is dropped in word-final position except after certain consonant clusters, e.g. *codru* 'wood'. Some neuter nouns have the plural ending *-uri*, which likewise triggers feminine plural agreement, but which is not found on feminine nouns, except for *vreme–vremuri* 'time'. I tentatively assume that this calls for a distinction between declension classes similar to the Latin contrast between *templum*.SG / *templa*.PL 'temple' and *tempus* / *tempora* 'time' (Romanian *timp / timpuri*). Adjectival inflection is identical to nominal inflection. The inflection of the definite article raises issues that will not be addressed here.

[2] Many of these neuter nouns descend from Latin neuter nouns.

[3] Both the F.SG.DAT/GEN *e* and the F.PL *e* are replaced by *i* in certain nouns, e.g. *bunică–bunici–bunici* 'grandmother' and *ramură–ramuri–ramuri* 'branch'. This alternation may be phonologically controlled.

(2)

		M	N	F
SG	NOM/ACC	N-u	N-u	N-ă
	GEN/DAT	N-u	N-u	N-ă
PL	NOM/ACC	N-i	N-e	N-e
	GEN/DAT	N-i	N-e	N-e

In other words, whether or not F.SG.DAT/GEN = F.PL neither improves or complicates the situation. The paradigm in (1) exhibits no ABA patterns and should be obtainable by positing lexical entries that will make neuter nouns syncretic with masculine nouns in the singular and syncretic with feminine nouns in the plural. But we will see that the paradigm cannot actually be derived on standard nanosyntactic assumptions as I understand them.[4]

7.2 Bridging fseq gaps

The distribution of forms in (1) is controlled by three different sets of features: gender features, number and case features. Correspondingly, the lexical entries we need will consist of three different layers. To make it possible for the neuter nouns to syncretise both with masculine nouns and feminine nouns, one must decompose the traditional gender features into sets of more primitive features as in (3a) or (3b):

(3)   a.   feminine   =   [F [N [M]]]   >
          neuter   =   [N [M]]   >
          masculine =   [M]
      b.   masculine =   [M [N [F]]]   >
          neuter   =   [N [F]]   >
          feminine   =   [F]

To create the syncretism between the F.SG.DAT/GEN and the F.PL, number must also be broken up into two features, e.g. plural = [PL [SG]] > SG.[5] Likewise, case will be represented as originally proposed by Caha (2009).

Suppose now we go by (3a) and try to capture the syncretism between masculine and neuter nouns in the singular by positing the entry in (4):[6]

---

[4]It is also not possible to adapt Caha's (2016) treatment of the Czech F/N.SG.GEN = F/N.PL.ACC/NOM syncretism since N.SG.GEN ending is not identical to N.PL in Romanian.

[5]Alternatively, the singular is just the absence of a number feature. As far as I can tell, this would not affect the conclusions in the text.

[6]It has been pointed out in the literature that using pointers leads to the prediction that certain ABA patterns should be possible (Taraldsen 2012; Vanden Wyngaerd

(4) $u \leftrightarrow$ [SG [N [M]]]

If matching is defined as in (5) ('bidirectional mapping'), (4) will not apply to a masculine noun, since the node N in (4) finds no match in (6):

(5)   The tree X in an entry M $\leftrightarrow$ X matches a tree Y generated by the syntax if and only if the root node A of Y has the same label as a node B in X and
    (i) every daughter of A matches a daughter of B and
    (ii) every daughter of B matches a daughter of A.

(6)   [SG [M]]

Similarly, if we go by (2b), we will have (7), which will not be applicable to neuter nouns (in (8)), because of clause (ii) of (5):

(7) $u \leftrightarrow$ [SG [M [N [F]]]]

(8)   [SG [N [F]]]

In the next subsection, we will see that there is no solution based on 'pointers' following the proposal by Caha & Pantcheva (2012), because neuter nouns syncretize both with masculine nouns and feminine nouns.

Then, I will look at the prospects of deriving the paradigm in (1) by removing clause (ii) from (5), as suggested by Pavel Caha (p.c.):

(9)   The tree X in an entry M $\leftrightarrow$ X matches a tree Y generated by the syntax if and only if the root node A of Y has the same label as a node B in X and every daughter of A matches a daughter of B.

It will turn out that (1) is still intractable.

7.3   No account using pointers

As already stated, I begin by abstracting away from F.PL = F.SG.DAT/GEN and case features will therefore not appear in any structural representations. I also assume that N.SG = M.SG, N.PL = F.PL require F > N > M or M > N > F because of *ABA.

For either arrangement $x > $ N $ > y$, the syncretism of N with x and the syncretism of N with y call for pointers to two distinct morphemes M and F that must determine the choice of the correct surface ending when the number layer is reached. But the smaller one of M and F will block the bigger one for [N [x]] by the Elsewhere Principle.

Assuming F > N > M, N.SG = M.SG leads to postulating the following entries (with (10) the item pointed to by (11)):

(10) $M \leftrightarrow$ [N [M]]

(11) $u \leftrightarrow$ [SG [M]]

Accounting for N.PL = F.PL would require (12)-(13) (where likewise (13) contains a pointer to (12)):

(12) $F \leftrightarrow$ [F [N [M]]]

(13) $e \leftrightarrow$ [PL [SG [F]]]

but by the Elsewhere Principle, (10) blocks (12) for [N [M]], with the consequence that *-e* is not assigned to neuter nouns in the plural.

With M > N > F, we would have (14)-(15) to account for N.SG = M.SG.

(14) $M \leftrightarrow$ [M [N [F]]]

(15) $u \leftrightarrow$ [SG [M]]

and N.PL = F.PL must be obtained by positing (16)-(17):

(16) $F \leftrightarrow$ [N [F]]

(17) $e \leftrightarrow$ [PL [SG [F]]]

By the Elsewhere Principle, (16) blocks (14) for [N [F]] so that the N.SG should now have the ending *-e*. Thus, adopting (5) and using pointers to bridge fseq gaps seems to fail to provide an account of the Romanian facts. But I will return to this issue in section 7.6.

7.4   No account using unidirectional matching

Abandoning pointers in favor of 'unidirectional matching' as defined in (9), we will still find that the paradigm in (1) cannot be derived even when abstracting away from F.PL = F.SG.DAT/GEN, i.e. if the paradigm were to have been as in (2). Essentially, the problem is now that differentiating forms based on gender distinctions is rendered impossible by the fact that the difference between the number of heads in the gender layer in two entries can be neutralized by the number of heads in higher layers and vice versa so that the Elsewhere Principle fails to choose between two competing entries.

Starting with F > N > M (and still leaving out case features), N.SG = M.SG should be attributed to the entry in (18):

(18) $u \leftrightarrow$ [SG [N [M]]]

Since the plural -*i* is unique to masculine nouns, we will also have:

(19) $i \leftrightarrow$ [PL [SG [M]]]

but (18) and (19) are the same size so that the Elsewhere Principle will not prevent -*i* from occurring also on masculine singular nouns in free variation with -*u*. (One will notice that the extra node that makes (18) as big as (19) is the N that would make (18) inapplicable to masculine nouns under bidirectional matching.)

Shifting to M > N > F doesn't help. The entry in (20) will now account for N.SG = M.SG while (21) accounts for N.PL = F.PL:

(20) $u \leftrightarrow$ [SG [M [N [F]]]]

(21) $e \leftrightarrow$ [PL [SG [N [F]]]]

but (20) and (21) are the same size and so, N.SG = *e* is not excluded.

7.5 Adding F.PL = F.SG.DAT/GEN doesn't help

We have seen that a paradigm like (2), which differs from (1) in that there is no case-sensitive SG = PL syncretism, cannot be derived on standard assumptions.

Putting F.PL = F.SG.DAT/GEN back in so that we return to (1), does not alleviate our problems. Assuming F > N > M, we would have (22)-(23), which are identical to (18)-(19) (repeated below as (24)-(25)) with the same number of case features added to each:

(22) $u \leftrightarrow$ [DAT [GEN [ACC [NOM [SG [N [M]]]]]]]

(23) $i \leftrightarrow$ [DAT [GEN [ACC [NOM [PL [SG [M]]]]]]]

(24) $u \leftrightarrow$ [SG [N [M]]]

(25) $i \leftrightarrow$ [PL [SG [M]]]

So, the Elsewhere Principle still does not prevent -*i* from occurring in masculine singular nouns. In addition, adding case features leads to a new problem. Capturing F.PL = F.SG.DAT/GEN = *e* while F.SG.ACC/NOM = *a*, requires (26)-(27):

(26) $e \leftrightarrow$ [DAT [GEN [ACC [NOM [PL [SG [F [N [M]]]]]]]]]

(27) $\breve{a} \leftrightarrow$ [ACC [NOM [SG [F [N [M]]]]]]

But (27) is smaller than (22) so that N/M.SG.ACC/NOM should have *a* rather than *u*. Adopting M > N > F removes this problem, but another problem remains. We would now have:

(28)   *e* ↔ [DAT [GEN [ACC [NOM [PL [SG [N [F]]]]]]]]

(29)   *ă* ↔ [ACC [NOM [SG [F]]]]

(30)   *i* ↔ [DAT [GEN [ACC [NOM [PL [SG [M [N [F]]]]]]]]]

(31)   *u* ↔ [DAT [GEN [ACC [NOM [SG [m [N [F]]]]]]]]

(28) and (31) are the same size so that N.SG should alternate freely between *u* and *e*.

7.6   Discussion

The singular form of neuter nouns in Romanian seems to be inherited from Latin where singular neuter and masculine nouns also had the same inflection except in the nominative. But the N.PL = F.PL syncretism is unexpected in this perspective. In Latin, neuter nouns had the ending *-a* in the nominative and the accusative case in the plural contrasting both with the masculine ending and the feminine *-ae* (nominative) and *-as* (accusative) in the declension classes were feminine nouns inflect differently from masculine nouns. While the Romanian plural *-e* is a predictable reflex of *-ae* in phonological terms, it seems unlikely that it could also have derived from the N.PL *-a*. That is, it seems unlikely that N.PL = F.PL = *-e* could be set aside as an instance of accidental homophony due to sound change.

If so, it seems significant that the paradigm in (1) cannot be derived by the usual tools made available by Nanosyntax, and we should look for ways of adding or refining our assumptions. In particular, one might want to take a closer look at the Elsewhere Principle.

In the preceding discussion of the analytical options, it was taken for granted that this principle simply compares the number of nodes in competing entries. In particular, all the problems that have arisen along the way, are due to the assumption that nodes in different structural layers (or 'sub-fseqs') are weighted equally at any point of the lexicalization process. But I have not been able to find a way of changing this that would be consistent with the fact examined here.

Also, the problems identified in sections 7.4 and 7.5 would not have arisen if we didn't adopt the definition of matching in (9) (unidirectional matching) instead of the one in (5) (bidirectional matching). But given what we saw in section 7.3, sticking with (5) seems impossible to the extent that (5) calls for pointers.

However, there is a way of overcoming the problem discussed in section 7.3, if the standard nanosyntactic toolbox is enriched by a last resort mechanism allowing the lexicalization procedure to backtrack and overwrite an earlier choice enforced by the Elsewhere Principle if it turns out that this earlier choice makes full lexicalization impossible. Suppose, for example, we posit the entries (32)-(35), where pointers appear only as needed in view of (5):

(32) $u \leftrightarrow$ [DAT [GEN [ACC [NOM [SG [M]]]]]]

(33) $i \leftrightarrow$ [DAT [GEN [ACC [NOM [PL [SG [M]]]]]]]

(34) $e \leftrightarrow$ [DAT [GEN [ACC [NOM [PL [SG [F]]]]]]]

(35) $\breve{a} \leftrightarrow$ [ACC [NOM [SG [F [N [M]]]]]]

By (5), (33) is inapplicable to singular nouns, and is correctly confined to masculine plural nouns, and (35) is only applicable to feminine singular nouns.

The pointers to M and F are introduced by (36)-(37) (identical to (10) and (12) in section 7.3):

(36) $M \leftrightarrow$ [N [M]]

(37) $F \leftrightarrow$ [F [N [M]]]

In section 7.3, we observed that the Elsewhere Principle would select M over F for the input [N [M]], raising the question how (34) could ever get to assign -*e* to neuter plural nouns. But notice that neither (33) nor (34) is applicable to (38) with M previously chosen as the spellout of [N [M]]:

(38)   [PL [SG [M]]]

In other words, the number features cannot be lexicalized. At this stage of the derivation, I suggest, the system backtracks to the point where [N [M]] was lexicalized by M and removes M as a candidate so that F is chosen instead. If so, the *e* will subsequently be introduced into neuter plural nouns by (34) applying to [PL [SG [F]]].

If this suggestion is on the right track, bidirectional matching as defined in (5) must be maintained together with the use of pointers. But as noted in footnote 6, the use of pointers allows certain ABA patterns to be created, and the question should be examined whether these patterns are robustly documented.

References

Caha, Pavel. 2009. *The nanosyntax of case*. Available at http://ling.auf.net/lingBuzz/000956. CASTL Tromsø dissertation.

Caha, Pavel. 2016. GEN.SG = NOM.PL: a mystery solved? In Markéta Ziková & Pavel Caha (eds.), *Linguistica brunensia. a special issue dedicated to petr karlík*, 25–40. Brno: Masaryk University.

Caha, Pavel & Marina Pantcheva. 2012. Tools in Nanosyntax. Handout for a talk given at Decennium: The First Ten Years of CASTL.

Taraldsen, Tarald. 2012. *ABA and the representation of features in syntax. Talk presented at BCGL 7, Brussels.

Vanden Wyngaerd, Guido. 2018. The feature structure of pronouns: a probe into multidimensional paradigms. In Lena Baunaz, Karen De Clercq, Liliane Haegeman & Eric Lander (eds.), *Exploring nanosyntax*, 277–304. Oxford: Oxford University Press.

8 Czech '-able'

Lucie Taraldsen Medová
TROMSØ

8.1 Outline

Consider some examples of Czech adjectives with 'able' meaning in (1).

(1)   'able' forms = (PREF)-root-THEME-TEL-N-ADJ

    a.   u-   děl- a- tel- n- ý
            PREF- do- AJ- TEL- N- ADJ.MASC.SG
            'doable'

    b.   forward- ova- tel- n- ý
            forward- OVA- TEL- N- ADJ.MASC.SG
            'forwardable'

    c.   léč- i- tel- n- ý
            cure- I- TEL- N- ADJ.MASC.SG
            'curable'

The morpheme composition of the forms in (1) seems straightforward enough: an optional prefix, root, theme, the suffix *-tel*, the suffix *-n* and the adjectival agreement.[1]

In what follows, I argue for a particular nanosyntactic view of the sequence THEME-*tel*-N, such that the morphemes are a particular spellout of the same features that can be seen in the paraphrase in (2). The paraphrase contains a form of the verb 'give' with an impersonal SE and the infinitive of the verb.

(2)   *paraphrase* = GIVE SE V$_{INF}$

    a.   dá   se u-děl-a-t
            give.3.SG SE do.INF

---

[1]I am not concerned here with the perfectivity of the verb: both perfective and imperfective verbs can get an 'able' derived adjective, though it seems that in general, perfective verbs are more prone to creating them (and I do not discuss this any further). It is also true that many of the 'able' adjectives are better negated, but I leave this aside as well.

         'it can be done'

  b.   dá   se forward-ova-t
        give.3.SG SE forward.INF
        'it can be forwarded'

  c.   dá   se léč-i-t
        give.3.SG SE cure.INF
        'it can be cured'

A note on the paraphrase: I have chosen a paraphrase where modality is expressed by the modal verb 'give,' which is homophonous with the regular verb 'give.' There are multiple such paraphrases possible, also without 'give,' but with another modal verb. For instance, a paraphrase to (2b) might be *může být forwardován* 'it can be forwarded'. However, I argue that there are independent reasons to think that the morpheme *-n* in (1) behaves like 'give' also in other respects, and that is why I concentrate precisely on this paraphrase. I now address the individual morphemes of the 'able' form in turn.

8.2 Theme

Looking at the morphological makeup of Czech 'able', the first component after the root (and the optional prefix, which I leave outside the discussion here) is the thematic suffix. The themes in Slavic introduce the argument structure of the verb. I will adopt an analysis according to which there are 7 of them: Ø, OVA, E, EJ, NU, AJ and I.[2] I list the classes below in a table, indicating for each theme roughly what kind of verb type it is associated with, and also whether it has an external argument, which is something that is going to be relevant shortly.

---

[2]There are various approaches to the classification of verbs depending on their theme. One tradition uses the so-called present stem for classification, see, e.g., Holaj (2018). I follow here the one-stem verb system introduced by Jakobson (1948) and brought into the current shape by Townsend & Janda (1996). It uses the past-tense stem that is 'more telling' for the (Russian) conjugation system; its advantage is that it helps to distinguish between classes that—in other classifications—would remain lumped together, cf. EJ and I in the traditional Czech classification (Holaj's *í*-verbs). It however means that the theme in (1a) is marked AJ, despite the fact that in its past stem the glide J does not appear for phonological reasons. For the sake of consistency, the theme is still marked as the present stem theme AJ, cf. *děl-AJ-í* 'they do'.

(3)   The verbal types

Theme	Verb type	External arg.
Ø	*unproductive*	–
E	unacc/stative	–
EJ	unacc/psych	–
NU	degree achievement	–
NU	semelfactive	+
I	causative	+
AJ	all types	+
OVA	all types	+

The theme NU, for instance, creates either transitive semelfactive verbs, as in (4), or unaccusative degree achievements, as in (5) (see Taraldsen Medová & Wiland 2018 for some details).[3]

(4)   *křik-nou-t* 'shout-NU-INF'

       Karel na mě   drze   křik-nu-l.
       Karel to me.ACC cheekily shout-NU-ACT.PART
       'Karel shouted at me cheekily once.'

(5)   *hloup-nou-t* 'stupid-NU-INF'

       Karel stár-nu-l   a   hloup-nu-l.
       Karel old-NU-ACT.PART and stupid-NU-ACT.PART
       'Karel was getting old and stupid.'

Importantly, not all of the themes can appear in the 'able' form, and the decisive factor is whether the thematic suffix does or does not introduce an external argument. Thus, the 'able' suffix (more specifically, its TEL part) can attach to the verbal theme AJ, OVA and I, as shown in (1). There are also 'able' adjectives derived from the NU themes (for semelfactives only), as in (6).

(6)   NU 'able'

    a. ne-   o-   po-   mi-   nu- tel- n- ý
        NEG- PREF- PREF- ROOT- NU- TEL- N- MASC.SG
        'unforgettable'
    b. ne-   dotk- nu- tel- n- ý
        NEG- ROOT- NU- TEL- N- MASC.SG
        'untouchable'
    c. ne-   zvlád- nu- tel- n- ý
        NEG- ROOT- NU- TEL- N- MASC.SG
        'impossible'

---

[3]I use the name 'degree achievement' to be consistent with the literature, but as everybody else, I don't see these verbs as achievements.

d. po- stih- nu- tel- n- ý
   PREF- ROOT- NU- TEL- N- MASC.SG
   'affectable'

However, there are no 'able' adjectives derived from the themes E and the (unproductive) theme ø, as shown in (7).

(7) *E, *ø 'able'
   a. *vid -ě- tel- n- ý
      see- Ě- TEL- N- MASC.SG
      Intended: 'visible'
   b. *slyš- ě- tel- n- ý
      hear- Ě- TEL- N- MASC.SG
      Intended: 'audible'
   c. *pro- vés- Ø- tel- n- ý
      PREF- lead- Ø- TEL- N- MASC.SG
      Intended: 'doable'
   d. *s- nés- Ø- tel- n- ý
      PREF- carry- Ø- TEL- N- MASC.SG
      Intended: 'tolerable'

And yet, it is possible to express the intended meanings—cf. 'visible' in (8a)—if the theme changes from E/ø to the theme I, as shown in (8).[4]

(8) *E, *ø 'able' → I 'able'
   a. vid- i- tel- n- ý
      see- I- TEL- N- MASC.SG
      Intended: 'visible'
   b. slyš- i- tel- n- ý
      hear- I- TEL- N- MASC.SG
      Intended: 'audible'
   c. pro- ved- i- tel- n- ý
      PREF- lead- I- TEL- N- MASC.SG
      Intended: 'doable'
   d. s- nes- i- tel- n- ý
      PREF- carry- I- TEL- N- MASC.SG
      Intended: 'tolerable'

I propose that the theme E introduces an external argument that is not

---

[4]It is not the case that none of the ø-themes can create an 'able' adjective. If we take the verb *krýt* 'cover' (assumed to belong in the ø theme class), the derived 'able' adjective is perfectly fine: *za-kry-tel-n-ý* 'coverable'. It seems that only vowel-final roots are prone to such derivation, cf. *nés-ø-t* 'carry' doesn't allow an 'able' in (8d) vs. *krý-ø-t* 'cover' above does. Thanks to Pavel Caha (p.c.) for the observation; I leave the issue open.

a 'proper' agent and that the 'able' adjective needs a proper ([+HUM])
agentive argument. Among the Czech themes, AJ, OVA and I are such
that they introduce such an argument, and that is why they allow for the
'able' forms, recall (1).[5] In other verb classes, the theme I must replace
the original theme (recall (8)), because I is the canonical introducer of
an external volitional argument.[6]

This analysis is supported by the fact that the thematic E-I alternation
is visible in other contexts as well, always in such a way that E appears in
a context where the external argument is 'suppressed', cf. in the passive
-N/-T participle *slaz-E-n-ý* 'sweetened'—derived from the I theme verb
slad-I-t 'sweeten,' which shows an I whenever the [+HUM] external ar-
gument is at play, cf. the finite forms: *slad-í-m* 'I sweeten' and *slad-I-l*
'he sweetened'. If correct, the very same alternation is observed in the
'able' constructions for certain verb classes, i.e. for the transitive E and
Ø verbs.

8.3   TEL

The claim that the Czech 'able' structure needs a [+HUM] external argu-
ment is corroborated by the fact that a part of the 'able' form—the suffix
-tel—is a suffix otherwise found (exclusively) in agent-denoting nouns
derived from verbs, see (9).

(9)   léč- i-   tel
       heal- THEME- TEL
       'a healer'

I take it that it is precisely this suffix TEL that 'eats up' the [+HUM]
portion of the external argument featural makeup.

This is confirmed by the minimal contrast between agentive nouns de-
rived by *-tel* and *-č*. Let me illustrate this using the verb *nos-I-t* 'carry$_{\text{HABIT}}$'.
On the one hand, there is the noun *nos-i-č* 'carri-er,' derived by *-č*, which
can be inanimate: 'a thing that is used for carrying something.' It can
also be animate, but then, it has a low degree of prestige: *nos-i-č* 'a per-
son who carries luggage'. On the other hand, the TEL-derived noun is
only animate, and bears a high degree of prestige: *nos-i-tel* 'a person who

---

[5]I leave the theme EJ that introduces mainly secondary imperfectives out of the
discussion.

[6]Needless to say, we are dealing with strong tendencies, which are muddled by
the phonological changes in the historical development of the Slavic languages. Thus,
there are also unaccusative verbs with the I theme, such *pěnit* 'foam'. However, the ma-
jority of the I theme verbs are transitive and among them, the majority has a volitional
external argument.

bears (carries) a distinction.'[7]

Apparently, then, it is the suffix TEL what is responsible for the restriction on the argument structure of the underlying verb, which has to have a human agent. Since such argument-taking properties are linked in Czech to theme vowels, this leads to a restriction on the kinds of themes that *tel* combines with. To illustrate this, in (10) I show that the thematic vowels which allow for the 'able' forms combine with TEL. In (11), I show that TEL does not combine with the thematic vowels that refuse 'able'.

(10) *TEL-assuming verbs*
    a. škol-i-tel
        supervisor
    b. řed-i-tel
        director
    c. kon-a-tel
        doer
    d. zřiz-ova-tel
        establisher

(11) *TEL-refusing verbs*
    a. *vid-ě-tel
        Intended: 'seer'
    b. *slyš-ě-tel
        Intended: 'hearer'
    c. *pro-vés-Ø-tel
        Intended: 'doer'
    d. *snés-Ø-tel
        Intended: 'tolerant person'

To sum up: the suffix TEL is inherently linked to the [+HUM] feature.

8.4 N

Recall the paraphrase in (2) with the modal verb 'give.' There are reasons to think that in the *tel-n-ý* adjective, this modal is reflected by the suffix -*n*. In particular, there is a class of verbs in Czech (semelfactives) which can also be paraphrased with 'give,' and where 'give' is also reflected by *n*. Thus, (4) above can be paraphrased as 'Karel gave me a cheeky shout,' and other examples are shown in (12).[8]

---

[7]There are a few TEL-derived nouns that do not refer to a human being: *dělitel* 'divisor' and *jmenovatel* 'denominator'. They were introduced in the 19th century as a part of new (German-replacing) Czech math terminology. These nouns 'feel' more agentive than verb-derived nouns introduced by another suffix, -EC, for instance *dělen-ec* 'dividend'.

[8]For detailed proposal on the 'give' in semelfactives, see Taraldsen Medová & Wiland (2018).

(12)   GIVE parahrase of NU-based semelfactives[9]

      a.   kop-n-ou-t
             kick-**N**-U-INF
             '**give** a kick'
      b.   štěk-n-ou-t
             bark-**N**-U-INF
             '**give** a bark'
      c.   hvízd-n-ou-t
             whistle-**N**-U-INF
             '**give** a whistle'

My proposal is that the analysis of the two constructions can be unified if indeed the semelfactive *n* is the same *n* that appears in 'able', and they both mean 'give.' The 'able' forms have the modal reading for the same reason that the Czech 'give' may have a modal meaning; I don't however discuss the precise mechanics of this analysis, and leave the issue open.

8.5   SE

In the previous sections, I presented the morphological pieces of the 'able' form; now we look at the paraphrase (2). The paraphrase contains 'give,' and I claim that it corresponds to the morpheme N in the 'able' form. Furthermore, the 'able' form contains a theme vowel—and the theme vowel finds its parallel in the infinitive. But there is one morpheme in the paraphrase that does not seem to correspond to anything in the 'able' form: SE. At the same time, there is one morpheme in the 'able' form that does not correspond to anything in the paraphrase—and that is the TEL morpheme. So, if the parallel is worth pursuing, then SE must correspond to TEL. We already know that TEL is linked to [+HUM], so, if the parallel is correct, SE must be linked to [+HUM] as well. Incidentally, this is precisely what I argue for in Medová (2009).

Rather uncontroversially, I assumed that there are external arguments of various types. Actors are external arguments, but it is rather difficult to see them as real, volitional agents. An example would be the noun *sun* in *The sun dried all the puddles up*. Agents on the other hand have [+HUM] external arguments, and an extra feature: volitionality. Following Taraldsen (2010) (and many others), I proposed that the volitionality feature is syntactically manifested and I identified it as a GEN-case layer in the external argument part of the structure. (The

---

[9]The NU theme is visible in its lexical shape (*nu*) in participles. However, in infinitives, *u* is lengthened to the diphtong *ou*. On templatic effects in Czech infinitives, see Caha & Scheer (2008).

non-volitional, *sun* type of arguments are in this story born as ACC and they lack the GEN layer altogether.)

So, how can we technically link SE to [+HUM]? And in general, how can we link a verb to having a human agent (cf. the [+HUM] feature)? Given that we no longer have lexical selection, we do selection by peeling (Medová 2009, Caha 2009, Starke 2014). There are reasons to think that SE can spell out the peels of the human agent (associated to SE in Slavic). And we can use the same technology to link TEL to the human agent. Hence: all is unified!

8.6   Outcome

The 'able' adjectives are derivable only from verbs that have the true agent, which I identified with the GEN case layer in the external argument VP spine. In the 'able' adjective, the theme must be big enough to include the GEN layer—that's presumably why the E-themes are not *'ablizable'*: they lack the agentive part provided by the I theme when it steps in for the adjectives in (8).

Thus, there are two ways to spell out the same structure: either by a roll-up derivation in (13), in which case the 'able' adjectives are derived, or by the paraphrase, with the spans spelled out by different morphemes in (14).

(13)   Roll-up derivation: *léč-i-tel-n* 'curable'
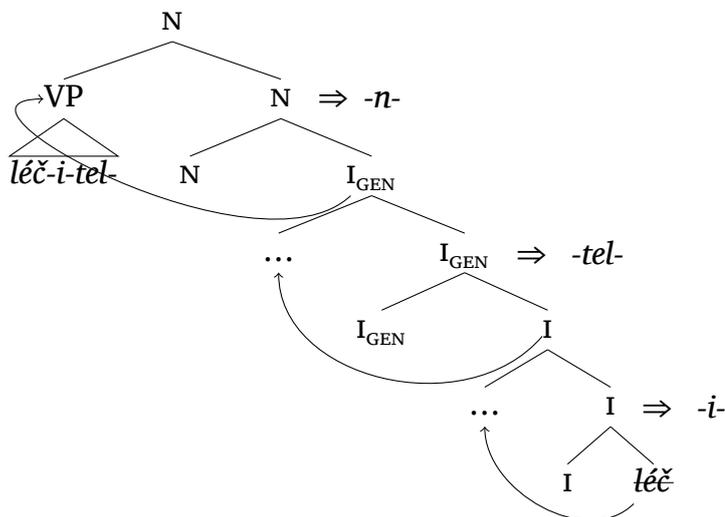
(14)   In-situ spell-out: *dá se léčit* 'it can be cured'
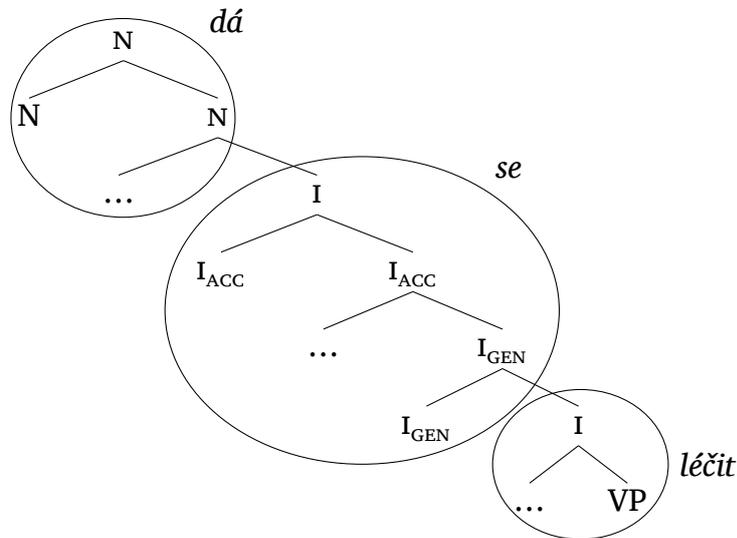
The I theme corresponds to two layers: the I_{ACC} layer and the I_{GEN} layer, the latter could be spelled out as part of the I theme—or by another morpheme—TEL or SE.

In the paraphrase, the I_{GEN} and N layers start a new zone: any kind of infinitive can be embedded under them, so, they must be independent of the actual theme of the embedded verb. On the other hand, not every verb can be turned into an 'able' adjective, indicating that in the roll-up derivation we deal with a single zone.

8.7   Acknowledgment

References

Caha, Pavel. 2009. *The nanosyntax of case*. Available at http://ling.auf.net/lingBuzz/000956. CASTL Tromsø dissertation.

Caha, Pavel & Tobias Scheer. 2008. The syntax and phonology of Czech templatic morphology. In Christina Bethin, Andrei Antoneko & John Bailyn (eds.), *Annual Workshop on Formal Approaches to Slavic Linguistics. The Stony Brook Meeting 2007*, 68–83. Ann Arbor: Michigan Slavic Publications.

Holaj, Richard. 2018. Balancing between roots and thematic vowels. In Pavel Caha, Karen De Clercq & Guido Vanden Wyngaerd (eds.), *The unpublished manuscript*, 81–93. Lingbuzz.

Jakobson, Roman. 1948. Russian conjugation. *Word* 4. 155–167.

Medová, Lucie. 2009. *Reflexive Clitics in the Slavic and Romance Languages. A Comparative View from an Antipassive Perspective*. Available at http://ling.auf.net/lingBuzz/001028. Princeton University dissertation.

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

Taraldsen Medová, Lucie & Bartosz Wiland. 2018. Semelfactives are bigger than degree achievements. Ms., submitted.

Taraldsen, Knut Tarald. 2010. Unintentionally out of control. In Maia Duguine, Susana Huidobro & Nerea Madariaga (eds.), *Argument Structure and Syntactic Relations from a Cross-Linguistic Perspective*, 283–302. Amsterdam-Philadelphia: John Benjamins.

Townsend, Charles E. & Laura A. Janda. 1996. *Common and Comparative Slavic: Phonology and Inflection*. Slavica Publishers.

9 Ordering paradoxes in a cross-categorial paradigm: on syncretisms with the declarative complementizer

Bartosz Wiland

Adam Mickiewicz University in Poznań

9.1 Introduction

This paper shows that incorporating non-definite demonstratives into the same fseq which covers syncretisms with the declarative complementizer, discussed in Baunaz and Lander's (2018a) work, is necessary to explain syncretic alignment and morphological containment in such paradigms in a systematic way. The paper also resolves an apparent *ABA violation in such a paradigm in Basaá, a language which shows syncretism between the demonstrative and the relativizer to the exclusion of the declarative complementizer.

9.2 Background: syncretisms with the declarative complementizer

Cross-linguistically attested syncretisms between the nominal declarative complementizer (COMP for short), the demonstrative pronoun (DEM), the interrogative pronoun 'what' (WH), and a relative pronoun (REL) suggest that these four categories form a paradigm, understood as an ordered set of related grammatical forms.[1] The syncretic alignment

---

[1] I will narrow down the discussion to syncretisms with the nominal complementizer, an equivalent of the English declarative *that*, as opposed to verbal complementizers. The latter are derived from assertive verbs like 'say', as for instance the Yoruba *pé*, which is a syncretic form for the verb 'say', as in (i), and a declarative complementizer which introduces clauses subordinated under verbs like 'forget', as in (ii):

between these categories have led Baunaz and Lander (2017; 2018a; 2018b) to propose that they are syntactically contained as in (1) (where ">" indicates dominance):

(1)   DEM > COMP > REL > WH

This ordering is motivated by the premise that syncretism anchors structural containment since it only targets contiguous layers of a syntactic structure, a generalization coined in Bobaljik (2007) as the *ABA. The following sample of languages illustrates syncretic alignments consistent upon the order in (1):

(2)

	DEM	COMP	REL	WH
English:	that	that	that	what
Modern Greek:	ekíno	pu	pu	tí
Italian:	quello	che	che	che
Romanian:	acel	că	ce	ce

In nanosyntax, the unattested ABA patterns are a consequence of the spell-out mechanism, which operates in concert with the Superset Principle, defined in Starke (2009) as follows:

(3)   An exponent of a lexical item is inserted into a syntactic node if its lexical entry has a subconstituent which matches that node.

The Superset Principle is amended with the Elsewhere clause, which makes sure that whenever there are lexical items that are in competition for insertion into a syntactic node, it is the exponent of the more specific lexical item (the one with fewer features unspecified in the node) that is inserted.

To briefly illustrate how the Superset Principle rules out the ABA patterns, let us work with the English declarative complementizer, demonstrative, wh- and relative pronoun, the categories which in Baunaz and Lander's analysis have a complex structure as in the tree in (4), whose left branch spells out as the prefix on the nominal stem (marked here as the triangle N), and the right branch spells out the ϕ-agreement features as the suffix.[2]

(i)   Olú pé awon ti   dé
        Olu say they have arrived
        'Olu says they have arrived.'

(ii)   Olú gbàgbé pé   Bólá ti jáde
        Olu forget   COMP Bola PF go.out
        'Olu forgot that Bola has gone out.' (data from Lawal (1991: 75))

[2]For the sake of concreteness, let us note that the nominal element marked as *n*

(4)

```
                                            φ ⇒ t
                                            △

        DEM ⇒ th                   N ⇒ a
                                   △
     DEM      COMP

          COMP      REL

             Rel      WH ⇒ wh

                   WH    n
```
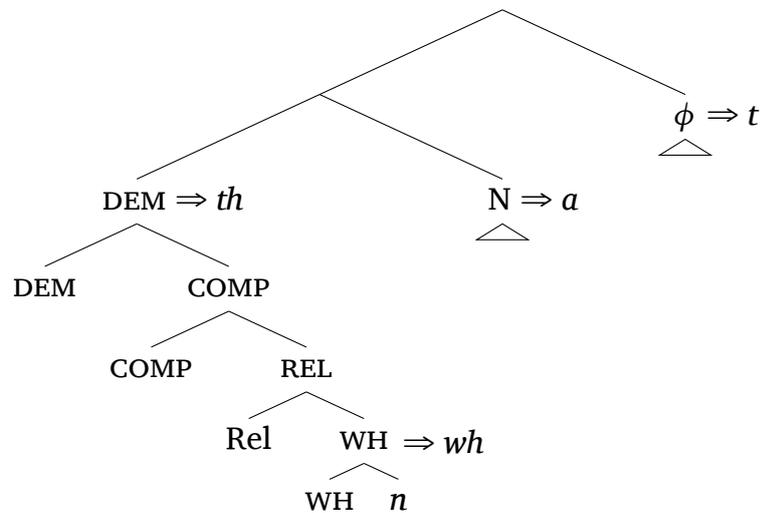
Given the lexical entries as in (5a) and (5b), *th-* and *wh-* come out as prefixes on *-at* (or rather, strictly following Baunaz and Lander's description, a bi-morphemic *-a-t*), which spells out an underlying middle and the right branch.

(5)   a.   [DEM [COMP [REL [WH *n*]]]] ⇔ *th-*
   b.   [WH *n*] ⇔ *wh-*

The phrasal nodes labelled as DEM, COMP, and REL all spell-out as *th-* as they constitute, respectively, the superset and the subsets of the lexical entry in (5a). The WH node, also a subset of the entry in (5a), is spelled out as *wh-* on the strength of the Elsewhere clause, since (5b) is a more specific match for the WH node than (5a). The *ABA is, thus, derived in a straightforward way: since the lexical entry for *wh-* is contained within the lexical entry for *th-*, it is impossible for *wh-* to lexicalize a structure bigger than *th-*.

Two remarks are in place before we proceed. First, it is important to note that the labelling used in (4) is a simplified way to illustrate Baunaz and Lander's analysis, in the sense that a 'demonstrative pronoun', a 'complementizer', a 'relativizer', and a 'wh-pronoun' lexicalize both the left branch, the middle, and the right branch of the tree (4) in their analysis, irrespective of morphological complexity of these categories. This is a natural consequence of phrasal spell-out. For instance, in Baunaz and Lander's system, the *ch-* prefix in the Italian *che* spells out both the left and the middle branch of the representation in (4) as a portmanteau,

---

at the bottom of the left branch of the tree in (4), the stem for the merger of the Wh feature, is described as a classifier-like lexical noun in Baunaz & Lander (2018a) and as non-lexical indeterminate noun in Baunaz & Lander (2018b). This issue, however, is largely orthogonal to what follows.

forming a 'complementizer', a 'relativizer', and a 'wh-pronoun', while *-e* spells out the right 'ϕ-agreement' branch as the suffix, as in:

(6)   a.

```
              COMP ⇒ ch      φ ⇒ e
           COMP           N
      COMP      REL
           Rel      WH
                WH    n
```

     b.

```
              REL ⇒ ch      φ ⇒ e
          REL           N
     Rel      WH
          WH    n
```

     c.

```
           WH ⇒ ch    φ ⇒ e
        WH      N
     WH    n
```
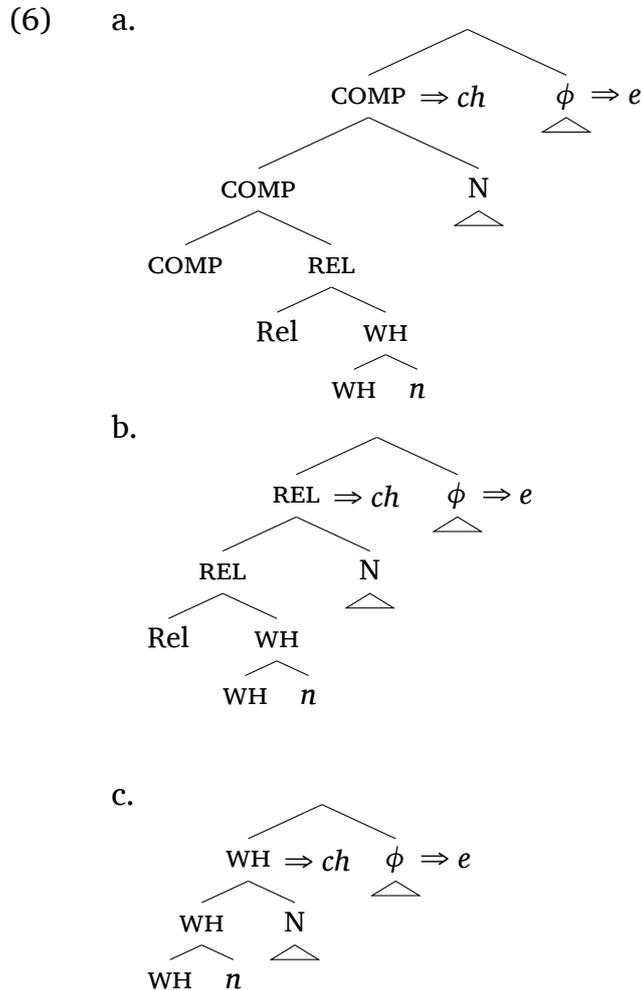
Thus, the features indicated here as 'DEM', 'COMP', 'REL', and 'WH' should be understood as subcomponents of the demonstrative, the complementizer, etc., rather than features that solitarily encode the properties of the categories they head. For example, the spatial deictic contrast in English demonstratives *th-{is, at}* is morphologically realized by the stem, not by the definite *th-* prefix. For this reason, Baunaz & Lander (2018a) describe the 'Dem' head in (4) as an instantiation of the definite article, a subcomponent of the demonstrative rather than the source of spatial deixis, an issue that will be taken up in a greater detail shortly. In order to keep the demonstrative pronoun in a language like Basaá, which does not have definiteness marking but shows syncretism with the relativizer in the picture, unless indicated otherwise, I will use the 'DEM' label more broadly so that it describes both kinds of demonstrative pronouns. Whenever it will be needed to differentiate between demonstratives with

and without definiteness marking, I will refer to them specifically as DEM$_{\text{DEF}}$ and DEM$_{\text{INDF}}$, respectively.

Second, both definite or indefinite DEMs, COMP, REL, and WH should not be necessarily treated as inherently simplex categories beyond their containment. The degree of their internal complexity is irrelevant to the discussion of patterns of syncretism between the categories they form, though. For example, we can follow Lander & Haegeman (2016) and split the demonstrative into a hierarchy of three distinct deictic features which cumulatively form the proximal ('close to speaker'), the medial ('close to hearer'), and the distal ('remote from speaker and hearer'), as in:

(7)

```
              DIST
             /    \
         Deix₃    MED
                 /   \
             Deix₂   PROX
                      |
                    Deix₁
```
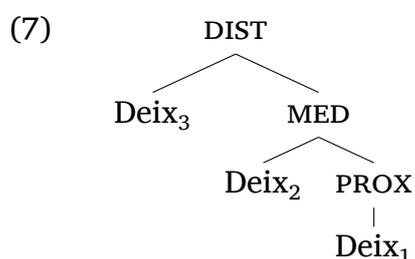
Splitting the deictic component of a demonstrative pronoun in this way is presupposed but it bares no consequences for its ordering with respect to COMP, REL, and WH, so I will continue to refer to it generically as 'DEM' for the ease of exposition.

9.3 Ordering paradoxes with the demonstrative

The DEM = COMP syncretism found in a number of languages, e.g. certain Germanic languages including English, Dutch, and German as in (8), points to the hierarchy 'DEM > COMP > REL > WH'.

(8)

	DEM	COMP	REL	WH
Dutch:	dat	dat	dat	wat
German:	das	dass	das	was

Some other languages, however, indicate that the order between these categories is different.

One class of challenges to 'DEM > COMP > REL > WH' comes from morphological containment of DEM in the structure of the other three categories, which we find in Slavic languages like e.g. Russian or Serbo-Croatian, as in:

(9)

	DEM	COMP	REL	WH
Russian:	to	čto	čto	čto
Serbo-Croatian:	to	što	što	što

The morphological containment of DEM within COMP, REL, and WH is opposite to what we expect if DEM is the biggest of all these categories. This problem is recognized in Baunaz & Lander (2018a), who propose to solve it by eliminating demonstratives without definiteness marking from the sequence so that it applies only to languages with morphologically marked definiteness on demonstratives, that is:

(10) DEM_{DEF} > COMP > REL > WH

More precisely, Baunaz & Lander (2018a) argue that only DEM_{DEF} projects as the top layer of the left branch of the tree in (4) and in languages like Russian and Serbo-Croatian DEM_{INDF} is restricted to the nominal stem, i.e. the middle branch of the tree in (4) (plus the right branch with the ϕ-agreement).

However, such a solution creates a paradox: on the one hand the hierarchy in (10) applies to the categories that always spell out both the left and the middle branch of the tree in (4) (either synthetically as in English or as a portmanteau morpheme in Italian), on the other hand it is defined only on the basis of the left branch of that tree, excluding the middle branch (putting aside the right branch with the ϕ-agreement, for a moment).

The other challenge comes from (what appears to be) a *ABA violation in Basaá (Bantu, A.43), which show syncretism between the relative pronoun and the demonstrative to the exclusion of the complementizer, as in:

(11)

	DEM	COMP	REL	WH
Basaá:	nú	Ø,	nú,	kíí
		lέ	lέ	

The status of Basaá as a language with non-definite demonstratives must be qualified. Demonstratives in Basaá can appear in pre- and post-nominal positions. A noun pre-modified by a demonstrative is focused, a noun post-modified by a demonstrative obligatorily takes the augment *í*-prefix, which marks definiteness/specificity (Jenks, Makasso & Hyman 2017), as shown on the example of class 1 demonstrative *nú* (other noun classes are subject to the same constraint):[3]

---

[3]A syncretic *í*-prefix is reported to function also as a locative marker as in e.g. *í-ꜜndáp* 'at home' (Makasso 2010: 148, as cited in Jenks, Makasso & Hyman 2017).

(12)   a.   **í-mut**$_i$ nú
                AUG.1.person 1.that.DEM
                'that person'

        b.   nú   mut
                1.that.DEM 1.person
                'THAT person' (Makasso 2010)

We can, thus, conclude that while demonstratives in Basaá contextually license an augment prefix which marks definiteness on the head noun, they are themselves not morphologically marked for definiteness.

There are two immediate ways to try to resolve this *ABA problem. One is to go along with the idea that Basaá does not belong to the sequence in (10) since it does not have definiteness morphology (save for the augment marker). This will eliminate the *ABA problem but we will be left with the DEM=REL syncretism which does not belong anywhere in the sequence. On the one hand, placing non-definite demonstratives in an alternative sequence 'COMP > DEM$_{INDF}$ > REL > WH' or 'COMP > REL > DEM$_{INDF}$ > WH' fixes the *ABA violation in Basaá but at the same time it makes the wrong prediction about the morphological containment of DEM$_{INDF}$ in WH in Russian and Serbo-Croatian. On the other hand, placing non-definite demonstratives at the bottom of 'COMP > REL > WH > DEM$_{INDF}$' explains the morphological containment in Russian and Serbo-Croatian but it reintroduces the *ABA violation in Basaá in a different place of the sequence by sandwiching the non-syncretic WH between syncretic REL=DEM$_{INDF}$.

A different way to go is to assume that since this complementizer is phonologically null, the COMP layer is not projected in Basaá at all. Such an explanation is a little hard to argue since a dialect of Basaá does have a different overt declarative complementizer *lέ*, as in (13), which is syncretic with the relativizer, as in (14):

(13)   mɛ ŋ̀-kâl   lέ   Tonye a ŋ̀-kŋ́ yàání
          I   PRS-say COMP Tonye SM PRS-go tomorrow
          'I say that Tonye will go tomorrow.' (Bassong 2010: ch.3, 30-a)

(14)   ɓaúdú ɓá gwĕ malĕt lέ a ŋ̀-kâl ɓɔ́ mam
          students SM have teacher REL SM PRS-tell them things
          'The students have a teacher that tells them stories.' (Bassong 2010: ch.4, 22-b)

As observed in Bassong (2010), the relativizer *lέ* is indeclinable and its distribution in relative clauses is more restricted than in the case of *nú*, nevertheless, its status as a declarative complementizer must be considered when the nonexistence of the declarative (factive) COMP layer

is postulated in order to salvage the hierarchy in (10) for Basaá. An intuitive option would be, thus, to assume that COMP is a skippable layer of structure (an 'fseq gap') only on top of the *nú* REL but not on top of the *lɛ́* REL.

In order to incorporate the facts from Slavic and avoid postulating an fseq gap in the *nú*-based paradigm in Basaá, I consider a solution which rests on two independent ingredients: (i) updating the sequence in (10) with DEM$_{INDF}$ and (ii) showing that the *ABA problem in Basaá is apparent, a conclusion we can draw from inspecting the syntax behind the REL-cell in the Basaá paradigm.

9.4 Low DEM$_{INDF}$

It appears that what constitutes an obstacle in resolving the ordering paradoxes for the sequence in (10) is that it describes the categories realized by both the left and the middle branche of the tree in (4), while the sequence applies only to the properties of the left branch, not the right middle 'base' branch (the N triangle). Let us, thus, consider what happens if we relax Baunaz and Lander's constraint that a demonstrative, a complementizer, a relativizer and a wh-pronoun are always realizations of more than one branch of the tree in (4).

I have argued elsewhere (Wiland (2018)) that the base for the formation of the wh-pronoun 'what' in Slavic is the indefinite demonstrative, which constitutes the bottom of a monotonically growing singleton branch, as in:[4]

(15)
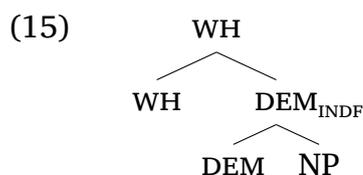
The nominal ingredient of a demonstrative pronoun—the NP stem that

---

[4]More precisely, I've argued in Wiland (2018) that the base for the formation of the Polish *co* and Russian *čto* 'what' is the medial demonstrative *to*, not the distal *tamto* (Pol) or *eto* (Ru). This makes an interesting prediction about the English *what* as potentially also based on a medial rather than distal *that* (and certainly not on proximal *this* given the non-existent *whis*). An extension of this idea for the English *th-at* and *wh-at*, however, would have to be based on a proviso that the *-at* suffix is either structurally identical in *th-at* and *wh-at* and its realization as in /ð-æt/, /w-ɒt/, /w-ɒt/, or /w-ɔt/ but not as in the expected */ð-æt/ is a matter of phonology (a possible but unlikely scenario given that /ðæ/ is an attested word-initial cluster in Modern English) or that /æt/ in *th-at* and /ɒt/ (and other variants) in *wh-at* are realizations of different subsets of the same fseq. We will return to this issue at the end of section 4.

spatial deixis merges—is responsible for the projection of a separate case fseq on its top (marked below as K_1, a stand-in for neuter nominative singular). All these layers are merged in the one and only branch, i.e. they constitiute the one and only fseq, as in the structure with a bare DEM_INDF and WH:

(16) a.

```
          K₁P
         /   \
       K₁    DEM_INDF
             /    \
           DEM    NP
```

b.

```
          K₁P
         /   \
       K₁    WH
             /  \
           WH   DEM_INDF
                /    \
              DEM    NP
```

Let us consider how these structures are lexicalized in Polish, a language with bimorphemic forms for all four categories, as shown in:

(17)

	DEM	COMP	REL	WH
Polish:	t-o	ż-e	c-o	c-o

In Polish, the spell-out of the 'WH > DEM_INDF' sequence involves a simple overriding. Given the lexical entries for the exponents of the forms in (17) as in:[5]

(18) Lexical entries in Polish

    a. [DEM NP] ⇔ *t-*
    b. [REL [WH [DEM NP]]] ⇔ *c-*
    c. [COMP [REL [WH [DEM NP]]]] ⇔ *ż-*
    d. [K_1] ⇔ *-o*

the spell-out of the WH overrides the earlier spell-out of DEM_INDF as in (19a), and the spell-out of K_1 requires the evacuation movement of its complement, as in (19b):

(19) a.

```
          K₁P
         /   \
       K₁    WH  ⇒ c
             /  \
           WH   DEM_INDF  ⇒ t
                /    \
              DEM    NP
```

---
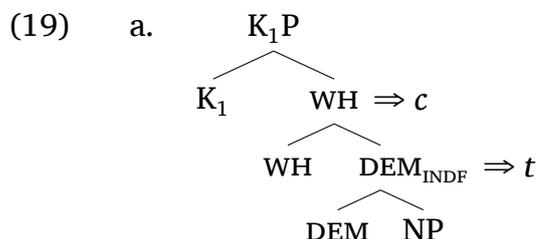
[5]The case suffix on the complementizer *ż-e* does not require a separate lexical entry than *-o*. As Baunaz & Lander (2018a) point out, the suffix *-o* /o/ shifts into *-e* /e/ after a soft consonant *ż-* /ʒ/.

b.

```
              /\
             /  \
      WH ⇒ c      K₁P ⇒ o
       /\          /\
      /  \        /  \
    WH   DEM_INDF ⇒ t  K₁   W̶H̶
         /\
        /  \
      DEM   NP
```
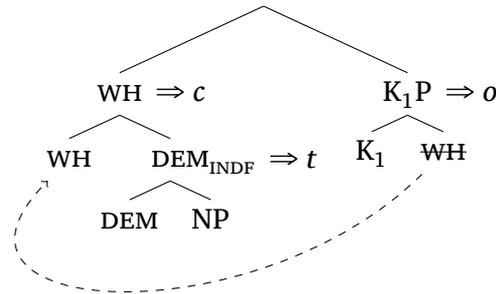
There is no need to postulate a second branch (e.g. the N triangle in (4)) if DEM$_{\text{INDF}}$ is already part of WH > DEM$_{\text{INDF}}$. With the lexical entries in (18), the lexicalization of REL and COMP layers takes place, again, by spelling out the one and only sequence:

(20)   Lexicalization of the sequence in Polish

```
        K₁P
        /\
       /  \
     K₁    COMP ⇒ ż
           /\
          /  \
       COMP   REL ⇒ c
              /\
             /  \
           REL   WH
                 /\
                /  \
              WH   DEM_INDF ⇒ t
                    /\
                   /  \
                 DEM   NP
```
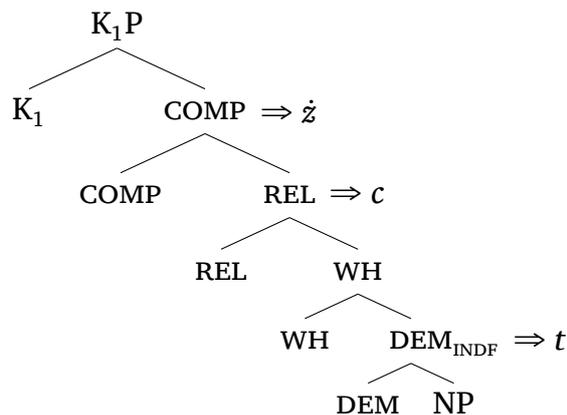
Note that the hypothesis that there is a singleton underlying branch for the sequence 'COMP > REL > WH > DEM$_{\text{INDF}}$' does not exclude the possibility that it may have to be reshaped in order to facilitate spell-out. This is a natural consequence of the spell-out algorithm in nanosyntax, as summarized in (21), but it does not equal the idea that a reshaped tree is underlyingly multi-branching.

(21)   stay ⇒ move ⇒ subderive

In fact, Starke (2018) advances that the subderivation of the left branch takes place as a last resort operation which facilitates spell-out after 'stay' and 'move' do not lead to lexical insertion. This is precisely the source of the difference between the pattern we see in Polish and Russian (and Serbo-Croatian), as argued in Wiland (2018). That is, while the shapes of the lexical entries in Polish allow the fseq in (20) to be spelled-out by 'stay' (ignoring case), the shape of the lexical entry for the Russian *č* as

in (22) requires the formation of the left branch.

(22)   Lexical entry in Russian

    [COMP [REL [WH DEM]]] ⇔ *č*-

If the lexical entries for the demonstrative *t*- and case *-o* are identical in Polish and Russian, then the lexicalization of WH, REL, and COMP will require the formation of the left branch in Russian, given the entry for *č*-. In contrast to Polish, only the bottom DEM$_{\text{INDF}}$ of the fseq in (20) can be spelled out by 'stay' (as *t*-) and movement operations are unable to reshape the tree in (20) in such a way that it matches (the subset or superset of) the entry for *č*-, either. The final available option is to launch a subderivation by providing the last spelled out feature from the mainline, i.e. the 'Dem' feature of DEM$_{\text{INDF}}$, as the basis for the merger of the 'Wh' feature, resulting in:

(23)

```
        WH
       /  \
     WH   DEM
```

Upon the merger of this subderivation with DEM$_{\text{INDF}}$, the resulting structure comes out as a bimorphemic *č-t-* (ignoring, again, the case suffix):

(24)

```
              /    \
       WH ⇒ č      DEM_INDF ⇒ t
       /  \        /   \
     WH   DEM    DEM   NP
```

Subsequent mergers forming REL and COMP will extend (what comes out as) the left branch, as discussed in Starke (2018), yielding:

(25)   Lexicalization of the sequence in Russian

```
              /         \
      COMP ⇒ č           DEM_INDF ⇒ t
       /    \              /   \
    COMP    REL          DEM    NP
            /   \
          REL   WH
               /   \
             WH    DEM
```

    If this is on the right track, then the contrast in the shapes of the lexical items in Polish and Russian informs us that the Polish pattern is more

basic, in the sense that the lexicalization of the same fseq is achieved by 'stay', while its lexicalization in Russian requires 'subderive', the last resort.[6] We can, thus, conclude that the underlying fseq comprises the non-definite demonstrative at its bottom, as in:

(26)   COMP > REL > WH > DEM$_{\text{INDF}}$

Such a sequence is enough to cover the facts form languages like Polish or Russian, but it needs to be updated with definite demonstratives in order to describe languages like English. This issue essentially reduces to the question: "where does definiteness belong to among the other categories?"

9.5 High DEM$_{\text{DEF}}$

There are at least two scenarios to consider. The first one is a variant of (26) in which definiteness is projected as a separate category at the bottom of the sequence, as in:

(27)   COMP > REL > WH > DEM$_{\text{DEF}}$ > DEF

Initially, this looks like an attractive option since not only does it suggest that definiteness applies directly to the nominal root, as in (28), but it also reflects the fact that definite markers can be contained in the structure of a demonstrative pronoun (e.g. English *th-at,* or Italian *quel-lo*).

(28)

$$
\begin{array}{c}
\text{DEM}_{\text{DEF}} \\
\diagup \quad \diagdown \\
\text{DEM} \quad \text{DEF} \\
\qquad \diagup \diagdown \\
\qquad \text{Def} \quad \text{NP}
\end{array}
$$

The idea that definiteness applies to the nominal root parallels with the

---

[6]The geometry of the tree in (25) resembles the structure for the Russian *č-t-* as in *č-t-o* in Baunaz & Lander (2018a), where it is based on a multi-branching tree in (4). Note, however, that there are two essential differences between these two representations. The first one is that in Baunaz and Lander's analysis, the Russian *t-* is an invariant nominal core, a kind of 'base' component, while the *t-* in (25) is the medial demonstrative pronoun (modulo the case suffix). The second difference concerns the nature of both representations. In Baunaz and Lander's analysis, the bi-morphemic *č-t-* realizes the nominal base and the prefix branch of an underlying multi-branching representation in (4). In the alternative in (25), the bi-moprhemic *č-t-* is created solely as a result of the spell-out algorithm, hence, there is technically no 'base' component or a pre-defined prefix branch; instead, the underlying representation is mono-branching just like it is in Polish (or any other language, for that matter).

situation observed with lexical nouns, as e.g. *the car*, where the definite article can appear without demonstrative morphology.

However, extending such a structure into WH, REL, and COMP leads to the *ABA violation: if the English DEF marker *th-* and the medial/distal demonstrative marker *-at* spell out such a structure, the demonstrative *-at* will come out as the suffix, following DEF-movement, as in:

(29)

DEF $\Rightarrow$ *th*   DEM$_{\text{DEF}}$ $\Rightarrow$ *at*

DEF   NP   DEM   ~~DEF~~
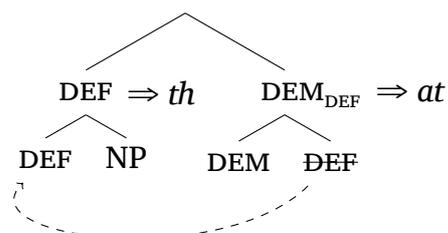
This appears to be a desired result. However, if the remainder of the sequence is 'COMP > REL > WH', then the addition of these layers will result in the *ABA pattern by sandwiching the *wh-* for WH between a lower *th-* for DEF and a higher *th-* for REL and COMP.

In the alternative scenario, definiteness applies to the entire fseq with the nominal root at its bottom, as in:

(30)   The updated singleton fseq

DEM$_{\text{DEF}}$

DEF   COMP

COMP   REL

REL   WH

WH   DEM$_{\text{INDF}}$

DEM   NP
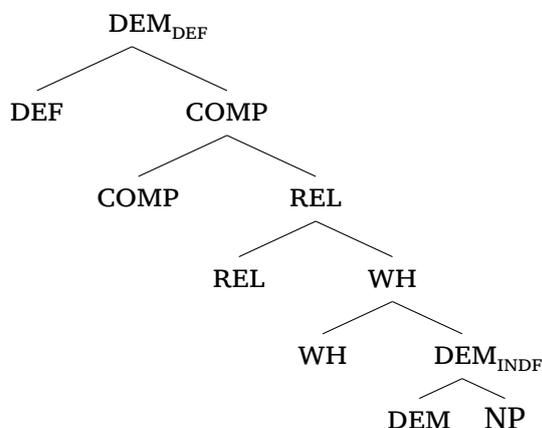
This sequence differs from the one that applies to both Polish and Russian (cf. (20)) only by the top layer and captures the fact that the deictic demonstrative is a stem for the formation of all higher categories.[7] Its lexicalization in English requires the formation of the complex left branch,

---

[7]This option, shown in (i) below without the intermediate WH, REL, COMP layers, is in essence compliant with Leu (2015: ch.2). Leu's work makes a case for the architecture of the Germanic definite demonstrative which contains the definite article and a 'proper' deictic element—an abstract HERE/THERE in Leu's (p. 15) analysis of German *der Tisch* 'the table', as in (ii):

as in (31), and with the addition of DEF, it mimics what we see in the lexicalization of this sequence in Russian in (25), modulo the DEF on top.

(31)   Lexicalization of the English *wh-at* and *th-at*
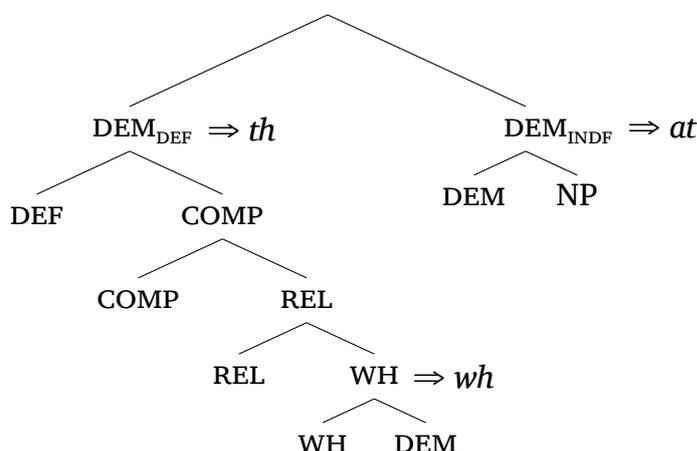
To sum up, defining the sequence as in (30) leads to the reordering in the paradigms of languages without definiteness marking, which should be represented as in (32).

(32)

	DEM$_{DEF}$	COMP	REL	WH	DEM$_{INDF}$
English:	th-at	th-at	th-at	wh-at	-at
Russian:		č-to	č-to	č-to	to

As already noted in footnote 4, the *-at* morpheme in *th-at* /ð-æt/ and in *wh-at* /w-ɒt/ has different exponents, even across the varieties of English. This contrasts with what we observe in Russian, where *to* is syncretic in all four forms (note that since the Russian *to* is a free form we can easily determine its 'own' phonology, which is impossible for the English bound morpheme *-at*). As pointed out by a reviewer, this fact does not appear to result in an ABA pattern, nevertheless, it may suggest that the stem for the formation of the WH, REL, COMP, and DEM$_{INDF}$ is not constant throughout the English paradigm the way it is in Russian. That is, the English /ɒt/ and /æt/ may reflect a subset-superset relation that is realized by different exponents, a scenario that is in principle possible

---
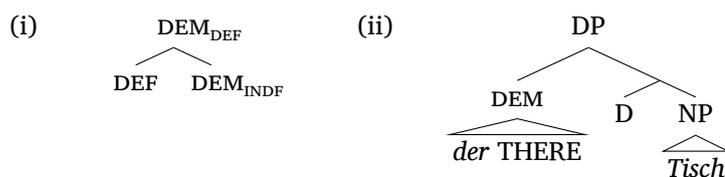
since the DEM_{INDF} stem is internally complex.[8] I leave this issue at this point.

With the sequence that covers both DEM_{DEF} and DEM_{INDF}, let us return to the second problem, the (apparent) *ABA violation that we find in Basaá, a language with non-definite demonstratives.

9.6   Resolving the *ABA violation in Basaá

While (30) resolves the morphological containment problem for Slavic, it only pushes the *ABA problem to a different place of the paradigm in Basaá:

(33)

	DEM_{DEF}	COMP	REL	WH	DEM_{INDF}
Basaá:		Ø,	nú,	kíí	nú
		lέ	lέ		

The violation of the *ABA generalization in Basaá is arguably unsolvable if we do not go beyond the cells of this paradigm.[9]

In order to resolve this problem, let us first recall from (12) (repeated below for convenience) that demonstratives in Basaá can be both post- and pre-nominal: a post-modified noun takes the augment í-prefix, a pre-modified noun is focused and does not take the augment prefix:

(12)　　a.　**í**-mut　　　　nú
　　　　　　AUG.1.person 1.that.DEM
　　　　　　'that person'
　　　　b.　nú　　　　　mut
　　　　　　1.that.DEM 1.person
　　　　　　'THAT person'

We observe a similar behavior of noun phrases modified by relative clauses, as in:

(34)　　a.　**í**-mut_i　　　nú　　[__i a　　bí ꜜjέ bíjέk　]
　　　　　　AUG-1.person 1.REL　　1.SBJ P2 eat 8.food

---

[8]The complexity of DEM_{INDF} concerns both the spatial deictic contrast as in Lander and Haegeman's (2016) decomposition in (7) but also its (pro)nominal component, marked in (30) and elsewhere as the bottom-most NP constituent. In Wiland (2018) I explore a possibility where the Russian and Polish NP t- (of t-o, as described in (18a) and (19a)/(25)) spells out subsets of a nominal sequence specified for THING and PERSON (in the sense of Cysouw 2004), a scenario more transparently visible in the English forms wh-at and wh-o rather than in the Russian č-to 'what' and k-to 'who' with a syncretic stem to.

[9]Except perhaps assuming that the WH cell altogether belongs to a different paradigm in Basaá, a possibility I will not explore here.

'that person that ate the food'

b. nú   mut$_i$ [$_{-i}$ a   bí ꜜjɛ́ bíjɛ́k]
   1.that 1.person   1.SBJ P2 eat 8.food
   'THAT person that ate the food'

In a two-step analysis of relativization in Basaá, Jenks, Makasso & Hyman (2017) first derive the pre-nominal order from the post-nominal order of the demonstrative, following a strand of research on morphologically complex demonstratives which submits that they derive from a lower, adjectival position (e.g. Leu 2015, a.o.), as in:

(35)   [$_{DP}$ nú$_{DEM}$ (*í-) [$_{NP}$ mut] n̶ú̶$_{DEM}$]

This descriptively explains the complementarity between the pre-nominal *nú* and *í-* in such a way that the presence of one blocks the presence of the other.[10] Second, they argue that both types of relative clauses are derived via head raising, in a similar way to what has been proposed for other Bantu languages (e.g. Ngonyani (2001); Carstens (2005), among others). Following Kayne (1994), the head nouns move to the specifier of the relative clause (here labelled as 'Rel-CP'), which can be selected by the D-head, as in their derivation of the relative clause (34a) with the pre-nominal demonstrative:

---

[10]To be more precise, Jenks, Makasso & Hyman (2017) assume that this blocking effect is an instance of a generalized version of the "Doubly-filled COMP Filter", which states that a head and its specifier cannot be both lexicalized. That is, the D-head cannot be lexicalized when the demonstrative occupies spec-DP. Suffice it say, nothing in what follows depends on that particular explanation, instead, what is important is the very observation that the fronting of the DEM *nú* from a post-nominal position blocks the merger of the augment *í-*.

(36)

```
                          DP₁
                  ┌────────┴────────┐
                 D₁              Rel-CP
                 │          ┌───────┴────────┐
                 í         DP₂             ┌──┴──┐
                      ┌─────┴─────┐       REL    TP
                     NP          Op        │   ┌──┴───┐
                     △      ┌────┴────┐     Ø  DP₂     T̄
                    mut    △    D₂   NP        ╶╴  ╶────────╴
                          nú    │              a bí ꜛjɛ́ bíjɛ́k
                                Ø
```
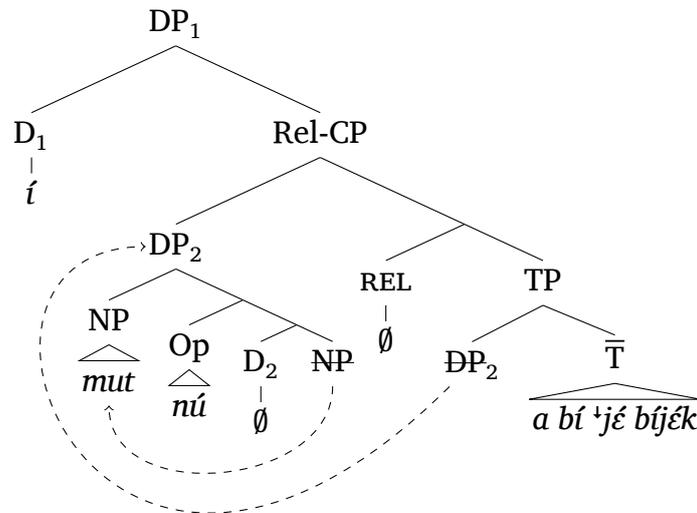
In such a derivation, the head noun *mut* 'person' is first fronted DP₂ internally to a position before the demonstrative "operator" *nú* and in the next step the entire DP₂ is raised to the specifier of the Rel-CP. The selecting head D₁ is lexicalized as the augment *í*-prefix.

Jenks, Makasso & Hyman (2017) refer to *nú* in the post-nominal position (36) as the "operator", which does not receive a focused interpretation in the specifier of Rel-CP. This contrasts with the derivation of the relative clause (34b) with a pre-nominal *nú* where, as Jenks, Makasso & Hyman (2017) argue, *nú* is a genuine demonstrative. In this case, it is the entire relative DP₂ that moves out of Rel-CP, which results in the demonstrative *nú* receiving a focused interpretation:

(37)

```
                           DP₁
                  ┌─────────┴─────────┐
                 DP₂            ┌──────┴──────┐
          ┌───────┴──────┐     D₁          Rel-CP
         DEM    ┌────┴────┐    │       ┌─────┴─────┐
         △     D₂    NP    Ø  DP₂     ┌──┴──┐
        nú     │     △         ╶╴    REL    TP
               Ø    mut               │   ┌──┴───┐
                                      Ø  DP₂     T̄
                                       ╶╴  ╶───────────╴
                                          a bí ꜛjɛ́ bíjɛ́k
```
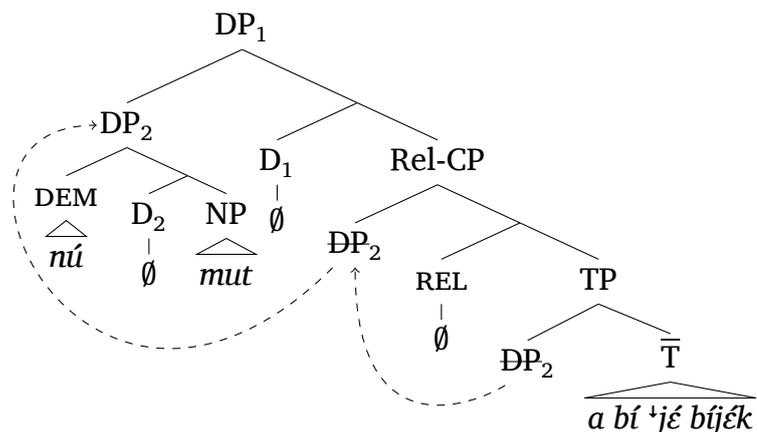
Such an account explains the complementary distribution between demonstratives and (what otherwise would appear to be a distinct category of) relativizers in Basaá, which holds in relative clauses involving a gap—both in subject and object relative clauses and both with pre- and post-nominal demonstratives, with other nominal classes, too, as for

instance in:

(38) **í-maaŋgɛ́**$_i$ nú (*nú) [mɛ ń yí ___$_i$]
AUG-1.child 1.DEM 1.REL 1sg PR know
'this/that child that I know'

(39) lí lí-wándá$_i$ (*lí˧) [___$_i$ lí bí ꜜjɛ́ bíjɛ́k]
5.DEM 5-friend 5.REL 5.SBJ P2 eat food
'THAT friend that ate the food' (Jenks, Makasso & Hyman 2017)

If Jenks et al.'s derivation of non-wh-relatives in Basaá is on the right track, then it resolves the *ABA problem in (33) straightforwardly. When we juxtapose the syntax of non-wh-relatives in Basaá to the syntax of non-wh-relatives in languages like English, we immediately realize that we are not comparing apples to apples and oranges to oranges when we are looking at the REL-cells of both kinds of paradigms. This is so since in relative clauses in languages like English, *that* does not form a constituent with the head noun and, instead, it serves as a genuine relativizer, the head of the relative clause, as in:

(40)

```
                    DP₁
                   /    \
                 D₁      Rel-CP
                 |      /      \
                the   NP        \
                     /  \        \
                  person REL      TP
                         |       /  \
                        that    NP    T̄
                                    /    \
                              ate my pizza
```
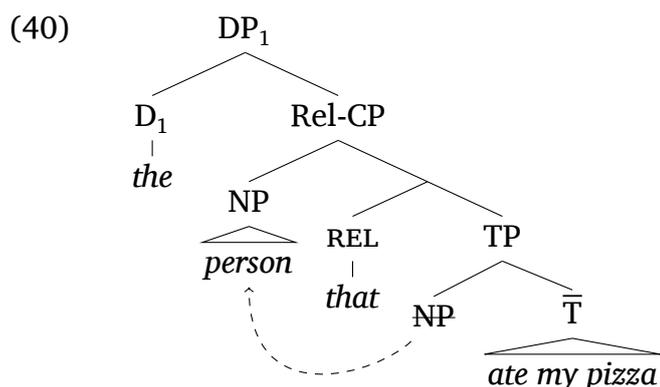
In contrast, *nú* in Basaá is a genuine DP$_2$-internal demonstrative and the relativizer, the head of the relative clause, is null. If so, then the offending paradigm in Basaá should be rewritten as:

(41)

	DEM$_{DEF}$	COMP	REL	WH	DEM$_{INDF}$
Basaá:		∅	∅	kíí	nú

9.7 Summary and conclusion

Cross-categorial syncretisms with the declarative complementizer discussed in Baunaz and Lander (2017; 2018a; 2018b) indicate that the wh-pronoun, the relativizer, the complementizer, and the definite demon-

strative pronoun form an fseq. Thus, morphological containment of non-definite demonstrative pronouns in the structure of the wh-pronoun, the relativizer, and the complementizer in languages like Russian poses a problem for such an fseq in that it does not apply uniformly to languages with and without definiteness marking. This can be fixed by inserting non-definite demonstratives at the bottom of this fseq to the effect that the definite demonstrative is a category which syntactically ranges from the non-definite demonstrative, through WH, REL, COMP, and is closed up by a high DEF. This result is possible to achieve if the underlying representation of these categories is simplified to a mono-branch and its partition into multiple morphemes is solely a result of the spell-out mechanism, not the geometry of an underlying representation.

Also, inserting non-definite demonstratives into the fseq that covers the syncretisms with the complementizers provides the first step in resolving an apparent *ABA violation in Basaá, a solution completed by differentiating genuine relativizers (like e.g. in English) from DP-internal demonstratives (or, descriptively speaking, 'demonstrative operators') disguised as relativizers in non-wh-relative clauses in Basaá.

Acknowledgment

References

Bassong, Paul. 2010. *The structure of the left periphery in Basaá*. University of Yaounde I MA thesis. `http://ling.auf.net/lingbuzz/001692`.

Baunaz, Lena & Eric Lander. 2017. Syncretisms with the nominal complementizer. *Studia Linguistica* 71(1). 1–34.

Baunaz, Lena & Eric Lander. 2018a. Cross-categorial syncretism and the Slavic containment puzzle. In Brian Joseph & Iliyana Krapova (eds.), *Balkan syntax and (Universal) principles of Grammar*. Berlin: Mouton de Gruyter.

Baunaz, Lena & Eric Lander. 2018b. Deconstructing categories syncretic with the nominal complementizer. *Glossa: a journal of general linguistics* 3 (1). 31. `http://doi.org/10.5334/gjgl.349`.

Bobaljik, Jonathan. 2007. On comparative suppletion. `http://ling.auf.net/lingbuzz/000443`.

Carstens, Vicki. 2005. Agree and EPP in Bantu. *Natural Language and Linguistic Theory* 23. 219–279.

Cysouw, Michael. 2004. Interrogative words: An exercise in lexical typology. ZAS Berlin. `http://cysouw.de/home/manuscripts_files/cysouwQUESTION_handout.pdf`.

Jenks, Peter, Emmanuel-Moselly Makasso & Larry Hyman. 2017. Accessibility and demonstrative operators in Basaá relative clauses. In Gratien Gualbet Antindogbe & Rebecca Grollemund (eds.), *Relative clauses in Cameroonian languages,* 17–46. Berlin: Mouton de Gruyter. `http://linguistics.berkeley.edu/~jenks/images/BasaaRelativeClauses.August2015.pdf`.

Kayne, Richard. 1994. *The antisymmetry of syntax.* Cambridge, MA: MIT Press.

Lander, Eric & Liliane Haegeman. 2016. The nanosyntax of spatial deixis. *Studia Linguistica* 71. 1–66.

Lawal, Adenike S. 1991. Yoruba *pé* and *kí*: verbs or complementizers. *Studies in African Linguistics* 22(1). 73–84.

Leu, Thomas. 2015. *The architecture of determiners.* New York: Oxford University Press.

Makasso, Emmanuel-Moselly. 2010. Processus de relativisation en Básaá: de la syntaxe á la prosodie. In L. Downing, A. Rialland, J.-M. Beltzung, S. Manus, C. Patin & K. Riedel (eds.), *Papers from the Workshop on Bantu relative clauses. ZAS Papers in Linguistics 53,* 145–158. Berlin: ZAS.

Ngonyani, Deo. 2001. Evidence for head raising in Kiswahili relative clauses. *Studies in African Linguistics* 30(1). 59–73. `https://journals.linguisticsociety.org/elanguage/sal/article/download/1349/1349-2000-1-PB.pdf`.

Starke, Michal. 2009. Nanosyntax: a short primer to a new approach to language. *Nordlyd* 36. 1–6. `http://ling.auf.net/lingbuzz/001230`.

Starke, Michal. 2018. Complex left branches, spellout, and prefixes. In Lena Baunaz, Karen De Clercq, Liliane Haegeman & Eric Lander (eds.), *Exploring nanosyntax,* 239–249. Oxford: Oxford University Press.

Wiland, Bartosz. 2018. A note on lexicalizing 'what' and 'who' in Russian and in Polish. Ms., Adam Mickiewicz University in Poznań.

Wiland, Bartosz. in prep. *The spell-out algorithm and lexicalization patterns: slavic verbs and complementizers*. Ms., Adam Mickiewicz University in Poznań.