RESEARCH

Deconstructing categories syncretic with the nominal complementizer

Lena Baunaz[1] and Eric Lander[2]

[1] University of Zurich, CH

[2] University of Gothenburg, SE

Corresponding author: Lena Baunaz (lena.baunaz@uzh.ch)

This paper investigates the internal structure of categories syncretic with the complementizer from a nanosyntactic perspective (cf. Starke 2009; 2014; Caha 2009). The (emotive factive) *that*-complementizer in Germanic, Romance, Hellenic, Slavic and Finno-Ugric languages has the same morphophonological form as other nominal categories, like demonstrative, interrogative, relative pronouns and indeterminate nouns. We claim that this homophony is not accidental. We also argue that these elements are internally complex and composed of syntactico-semantic features which are hierarchically ordered according to a functional sequence. More specifically, the internal structure can be considered essentially trimorphemic, being composed of (i) a lexical core or base which in our data is nominal (the nominal core called simply *n*), (ii) an inflectional ending (which we label Infl or Φ), and (iii) a functional morpheme which resembles an article of sorts and often (but not always) appears as a prefix (which we label simply F). The *n* and Infl components in the structures studied here are invariant and can be shown to be quite small, while F, on the other hand, crucially varies in size, depending on the function of the relevant morpheme involved (Dem, Comp, Rel, Wh or Indet). Importantly languages may lexicalize each of these components (*n*, Infl, and F) in different ways. Evidence for the fseq we are advocating comes from crosslinguistic patterns of syncretism and morphological containment.

Keywords: complementizer; (cross-categorial) syncretism; morphological containment; nominal core; phrasal spellout

1 Introduction: Cross-categorial syncretism

Declarative complementizers frequently have the same morphophonological form as other categories, like (pro)nouns, prepositions and verbs. In the (pro)nominal domain, for example, this cross-categorial syncretism is observed in English *that*, which can act as a demonstrative pronoun, a complementizer, or an indeclinable relativizer, and also in French *que* and Italian *che*, which can serve as complementizers, relativizers, and interrogative pronouns.[1] Cross-categorial syncretism can also implicate the verbal domain in a wide range of languages. For instance in Akan [Niger-Congo] the element *se*, in addition to being a verb with the meaning 'be like, resemble', has a range of functions: complementizer, quotative marker, purpose marker (i.e. 'in order to'), and similative marker (i.e. 'like, as'). Similarly in Mandarin [Sinitic], the item *shuō* is a verb meaning 'say' but also a complementizer and quotative marker.[2]

---

[1] The items responsible for non-finite complementation in these languages, moreover, appear to involve a cross-categorial syncretism between complementizers and prepositions (French *à, de, pour*; English *for*).

[2] Note that languages do not necessarily choose one or the other type of complementizer. For example, a quick look at English shows verbal, nominal, and prepositional complementizers (i.e. V: *like/be like*, N: *that*, and P: *for*).

The prevalence of this phenomenon suggests that we should not treat it in terms of accidental homophony but rather in terms of a common underlying structure and more properly *syncretism*, defined as "a surface conflation of two distinct morphosyntactic structures" (Caha 2009: 6). Cross-categorial syncretism thus arises when two or more distinct grammatical categories, each with a distinct underlying structure, are spelled out by a single element. In this paper we will show (i) that syncretism involving the nominal complementizer is highly constrained, (ii) that the elements participating in these syncretism patterns can be decomposed further, into a tripartite morphological structure, and (iii) that each of the morphological components in this tripartition have certain basic properties which are stable across languages.

2 Syncretism with the emotive factive complementizer

That-complementizers vary as to what information they lexicalize crosslinguistically (Roussou 2010; Baunaz 2015; 2016; in press; Baunaz & Lander in press; 2017a). Whereas some languages lexicalize only a single form of the complementizer, others show two or even three morphophonological forms. For instance, whereas English has only one nominal Comp (*that*), Modern Greek (MG) has two (*oti* and *pu*).[3] In languages where more than one complementizer appears, the distinction is related to interpretation. In MG for instance, *oti* is a non-factive complementizer and *pu* is a factive complementizer.

In Baunaz & Lander (2017a), we show that the declarative complementizer (Comp) participates in crosslinguistic syncretism patterns involving the demonstrative pronoun (Dem), restrictive relative marker (Rel) (in most languages considered here actually an indeclinable relativizer, labeled Rvz), and interrogative pronoun (Wh).[4] Furthermore we observe that in many cases there is syncretism between these categories and a bound morpheme encoding a bleached meaning like 'thing'. In many of the languages studied here this bound morpheme makes up part of the internal structure of certain quantifier words. We call this element Indet, which stands for *indeterminate noun*.[5]

The data considered in Baunaz & Lander (2017a) came from a sample of 13 Indo-European and Finno-Ugric languages. Our data set in this paper has been expanded to 22 languages. We also make an important distinction in this work which has not been made before, namely that it is the *emotive factive* complementizer (that is, the complementizer used under predicates like 'regret', 'be surprised', 'be happy', etc.) which is the relevant function that overlaps with the functions represented by Dem, Rel/Rvz, and Wh (e.g. Greek *pu* is syncretic with the Rvz function, while *oti* is not). In languages where there is no overt distinction between the factive and non-factive complementizer, we assume that there is syncretism between the two. For English, for instance, we provide *that* as the factive complementizer (even though it happens to also be the non-factive complementizer).

---

[3] We are deliberately leaving *na* aside since its status is not clear, and authors diverge as to its exact identity (complementizer or mood particle). See Roussou (2009; 2010) and Giannakidou (2009) for discussion.

[4] For the general idea concerning a close relationship between pronouns and complementizers, see Le Goffic (2008), Sportiche (2011), Baunaz & Lander (in press; 2017a) for French; Manzini & Savoia (2003; 2011), among others, for Italian; for the comparative Indo-European tradition, see Meyer-Lübke (1890: §613; 1899: §563) on Romance; Roberts & Roussou (2003), Kayne (2008), Leu (2008; 2015) for English and West Germanic in general; Roussou (2010) for Modern Greek; see Kiparsky (1995) on Old Germanic.

[5] For instance, in MG *ká-ti*, the element *ká-* 'each, every' is the distributive operator and *-ti* 'thing' is the indeterminate part (see Baunaz & Lander 2017a). The term *indeterminate pronoun* has also been used, usually to refer to phrases that are generally associated with different operators in Japanese (Kuroda 1965; see also Szabolcsi, Whang & Zu 2014). Kishimoto (2000) and Leu (2005), among others, have also referred to such elements as *light nouns*. Indeterminate nouns are generally invariable for number across languages.

2.1 The data

Table 1 shows the main data we consider in this paper. Languages are grouped into North
Germanic (NGmc), West Germanic (WGmc), Romance (Rom), Finno-Ugric, Hellenic, East
Slavonic (ESlav), West Slavonic (WSlav), and South Slavonic (SSlav). Syncretism is indi-
cated by gray shading, and pronouns are provided in their neuter/inanimate singular
forms throughout. As the reader can see, our columns can be arranged in one particular
order which accounts for all the patterns attested crosslinguistically in terms of strictly
adjacent cells in the paradigm. Exactly this behavior is typical of syncretism (Caha 2009).

Table 1: Syncretism patterns crosslinguistically (neuter/inanimate singular forms).

		DEM$_{PRO}$	COMP$_{FACT}$	REL$_{RESTR}$	WH$_{PRO}$	INDET$_{THING}$
NGmc	Swedish	det^{Pro}	att	som^{Rvz}	vad^{Pro}	-ting
	Danish	det^{Pro}	at	som^{Rvz}	hvad^{Pro}	-ting
	Icelandic	það^{Pro}	að	sem^{Rvz}	hvað^{Pro}	-hvað-
WGmc	English	that^{Pro}	that	that^{Rvz}	what^{Pro6}	-thing
				% as^{Rvz}		
	Dutch	dat^{Pro}	dat	dat^{Rvz}	wat^{Pro}	iets
						wat
	German	das^{Pro}	dass	das^{Rel}	was^{Pro}	-was
	Sw. German	das^{Pro}	dass	wo^{Rvz}	was^{Pro}	-is
	Yiddish	jenc^{Pro}	vos^{Fact}	vos^{Rvz}	vos^{Pro}	-vos
			az	az^{Rvz}		
Rom	French	ce^{Pro}	que	que^{Rvz}	que^{Pro}	-que
	Italian	quello^{Pro}	che	che^{Rvz}	che^{Pro}	-che
	Spanish	aquél^{Pro}	que	que^{Rvz}	qué^{Pro}	N/A[7]
	Romanian	acel^{Pro}	că	ce^{Rvz}	ce^{Pro}	ce-
Finno-Ugric	Hungarian	az^{Pro}	hogy	ami^{Rel}	mi^{Pro}	-mi
	Finnish	tä-^{Pro} 'this'	että	mi-^{Rel}	mi-^{Pro}	mi-
Hellenic	Modern Greek	ekíno^{Pro}	pu^{Fact}	pu^{Rvz}	tí^{Pro}	(-)ti(-)
ESlav	Russian	to^{Pro}	čto	čto^{Rvz}	čto^{Pro}	(-)čto(-)
WSlav	Polish	to^{Pro}	że	co^{Rvz}	co^{Pro}	co-
				% że^{Rvz}		
	Czech	to^{Pro}	že	co^{Rvz}	co^{Pro}	-co
SSlav	Serbo-Croatian	to^{Pro}	što^{Fact}	što^{Rvz}	što^{Pro}	-šta/-što
	Bulgarian	tova^{Pro} 'this'	deto^{Fact}	deto^{Rvz}	kakvo^{Pro}	-shto
	Macedonian	toa^{Pro} 'this/that'	što^{Fact}	što^{Rvz}	što^{Pro}	-što
			deka	deka^{Rvz}		

---

[6] Non-standard English also has relative use of *what* (e.g. *the thing what I can't stand*), giving a Rel/Wh
syncretism.

[7] We do not discuss Spanish in this paper, as Spanish quantifiers happen not to show overt realization of an
indeterminate noun *-que* (cf. *cada* 'each, every', *cada uno* 'each one', *alguno* 'some/someone, somebody',
alguien 'someone, somebody'), noting in passing that *cual-que* 'some' was in use in Old Spanish. See also
Section 4.

Again, more often than not the relative marker at stake is an indeclinable relativizer (Rvz). Note also that some languages with multiple complementizers available will allow one of them to appear under *either* factive or non-factive predicates, while the other one is possible under factive predicates *only* (e.g. SC factive/non-factive *da* vs. strictly factive *što*). In such cases we give the complementizer which is unambiguously factive (e.g. SC *što*), as this is the one that participates in syncretism.[8]

At this point, we briefly summarize the data in Table 1 branch by branch.

2.1.1 Germanic

We see that in *North Germanic*, there are no (obvious) syncretisms with the complementizer. In *West Germanic*, complementizers are often syncretic with the relativizer and the distal 3SG neuter Dem, but they are not syncretic with Wh. Wh and Indet are syncretic in Icelandic, Dutch, German, and Yiddish (cf. Ice. *eitt-hvað* 'something', Du. *iets* 'something' but also *wat* 'some(thing)', Ger. *et-was* 'something', Yid. *et-vos* 'something').

Note that where German has *et-was*, Swiss German has *öp-is* 'something' (vs. *öp-er* 'someone'), suggesting *-is* (and *-er*) are indeterminate nouns (see also Leu 2016). Note that Swiss German *-is* is not syncretic with the item in the next cell (Wh *was*). The fact that Wh and Indet are very frequently syncretic might suggest that Indet is not really distinct from Wh and that what we have labeled Indet (e.g. in Fr. *quel-que*) is just Wh. To us the Swiss German (and probably Bulgarian) facts prove that Indet and Wh really are separate entities.

2.1.2 Romance

In Romance (minus Romanian), Comp, Rel, Wh and Indet are all syncretic with each other, but these are not syncretic with Dem. Romanian has one declarative complementizer, *că*. Complementizer *că* is the complementizer by default, appearing almost everywhere except under predicates selecting the subjunctive mood (see fn. 8 and also Baunaz & Lander 2017a for more details). *Că* is not syncretic with Rel, Wh, Dem or Indet. Note also that *ce* is used as a relativizer, and that this item is syncretic with the Wh item meaning 'what' (Grosu 1994; Benţea 2010, among others), as well as with the Indet word meaning 'thing' (cf. Ro. *ce-va* 'something', *ori-ce* 'anything').

---

[8] We are acutely aware that there are various complications involving indicative vs. subjunctive (or realis vs. irrealis) in the complementizer systems of many of the languages under discussion here, such as Greek, Balkan, and the dialects of South Italy. For instance, in addition to *oti* and *pu*, Modern Greek also has *na* under desiderative ('wish'-type) verbs. The status of *na* is debated (Roussou 2010 considers it as a complementizer, while Giannakidou 2009, among others, view it as a mood particle), but it is clear that its distribution is different in nature from that of *oti* or *pu*. The same has been observed for Griko (Italiot Greek), with its declarative complementizer *ka* vs. modal complementizer (or particle) *na* (Baldissera 2013 and references cited there). Parallel to Greek, both Bulgarian and Serbo-Croatian also display a subjunctive mood particle (*da*) under certain verbs (see Krapova 1998 on Bulgarian; Baunaz 2016, in press, on Bulgarian and Serbo-Croatian; see also Sočanac 2017 about Slavic more generally). Dual complementizer systems in South Italy appear to be similar in that they often show declarative/epistemic vs. volitional forms (e.g. Sicilian *ca* vs. *chi*) (see Calabrese 1993; Ledgeway 2013; 2015, among others). We will not take a stand here on how mood and modality may or may not intersect with the emotive factive complementizer which we take to be crucial (especially considering that the emotive factive Comp in particular is understudied). Moreover, there is variation to consider: Balkan languages often make use of the indicative complementizer under emotive factive verbs, while Romance languages tend to show subjunctive inflection under emotive factive verbs. On this point – that Romance languages (Standard French, Standard Italian for instance) use the subjunctive mood under emotive factive verbs (subjunctive and indicative complementizer have the same form in these languages, cf. *que/che*) – it is interesting to note that there are data from Manzini & Savoia (2003) pointing to syncretism precisely between the *subjunctive* Comp and Rel (e.g. Ardaùli: Comp$_{IND}$ *ka* – **Comp$_{SUBJ}$ ki** – Rel *ki*, or Làconi: Comp$_{IND}$ *ka* – **Comp$_{SUBJ}$ tʃi** – Rel *tʃi*). Thanks to a reviewer for comments on this topic.

2.1.3 Finno-Ugric

The Hungarian complementizer *hogy* (related to Rel *a-hogy* and Wh manner adverbial *hogy(an)* 'how, in which manner/way') is not syncretic with Dem *a-z*, Rvz *a-mi*, Wh *mi* or Indet *-mi,* though there is syncretism between *a-mi* and *mi* (the *a-* in the Rel marker is likely a D-marker, much like *th-/d-* in West Germanic). For Indet consider *vala-**mi*** 'something, anything', *bár-**mi*** 'anything, whatever'.

The Finnish complementizer *että* is syncretic with neither Rel, Wh and Indet. That Finnish *että* is nominal can be argued on the basis of the fact that it is historically derived from the demonstrative **e-* (see *ez* 'this' in Hungarian). The *-ttä* component is taken to be a modal ending, with the original meaning of 'in this way, so' (Keevallik 2008: 141). We may note that the Finnish proximal demonstrative *tä-* derives from another root than *-ttä*, but that their phonological similarity may perhaps be synchronically analyzable in terms of a nascent Dem/Comp syncretism. As for *mi-*, it is syncretic with Rel, Wh, and Indet (***mi**-kä hyvänsä* 'anything', *ei **mi**-kään* 'nothing'), as in Hungarian. For the Wh paradigm, *mi-* is the inanimate stem (vs. the animate stem *ke-*).

2.1.4 Hellenic

Modern Greek has two different complementizers (though see fn.8 above). *Pu* introduces factive complements, and *oti* introduces non-factive complements. Complementizer *pu* is syncretic with the relativizer *pu*, but not with Dem. We note that the locative Rel pronoun *ó-pu* is bimorphemic, combining interrogative *pu* with the definite article *o-* (cf. Hungarian *a-hogy*, *a-mi* above). *Oti* may also introduce epistemic factive complements (but not emotive factive complements). MG complementizer *pu* is thus the factive complementizer which we single out in our data. It is syncretic with Rel, but not with Dem, Wh, and Indet. Wh and Indet, however, are syncretic (consider Indet in *ká-**ti*** 'something', ***tí**-pota* 'anything').

2.1.5 Slavic

The complementizer by default in Russian is *čto*. *Čto* is syncretic with the Rvz, Wh and Indet (i.e. ***čto**-to* 'something', *ne-**čto*** 'something specific'), though not with Dem *to*.

Polish *że* is not syncretic with anything in the standard language, but it is syncretic with a relativizer which is available in South-Eastern Polish and in some non-standard varieties of Polish. We note that relativizer and the Wh-word *co* are also syncretic with the Indet *-co* (see Po. ***co**-ś* 'something').

The default complementizer *že* in Czech has similar properties as its Polish cognate, though it does not seem to serve as a relativizer in any Czech varieties we know of. Czech *co* also shows Rvz/Wh/Indet syncretism (see Czech *ně-**co*** 'something' for Indet), as in Polish.

Like Modern Greek, Serbo-Croatian and Bulgarian lexicalize two complementizers: *da* and *što* in SC, and *če* and *deto* in Bulgarian. In Serbo-Croatian the complementizer *da* is the complementizer by default. It is not syncretic with Rel, Wh, Indet, or Dem. The use of SC *što* is quite limited: it only appears under emotive factive verbs. It is syncretic with Rvz, Wh and Indet (for which consider SC *ni-**šta*** 'nothing', *ne-**što*** 'something').[9] In addition, just like in Russian, SC *što* is partially syncretic with the 3.SG demonstrative *to*. In Bulgarian, the complementizer *če* appears in most environments, with the notable exception of under emotive factive verbs, where *deto* is used. Comp *deto* is syncretic with Rvz *deto*, but not with Wh *kakvo* 'what' or with Indet *-shto* (see Bg. *ni-**shto*** 'nothing, anything', *ne-**shto*** 'something'). See also fn.8 above.

---

[9] Regional variation as to the use of *što* or *šta* with Comp, Rel and Wh is found amongst SC speakers (Tanja Samardžić & Tomislav Sočanac, p.c.).

Finally, the emotive factive complementizer in Macedonian is *što*, which is also a relativizer, a Wh-pronoun and an Indet (for which consider Mac. *ni-što* 'nothing', *ne-što* 'something'), instantiating a Rvz/Wh/Indet syncretism. Though *deka* is the default complementizer (i.e. not the emotive factive complementizer), we include it here to show that it is syncretic with Rvz *deka* in this language.

2.2 Additional evidence

We note that although most of the evidence in Table 1 comes from Indo-European, the patterns in Hungarian and Finnish (Finno-Ugric) also conform to the sequence. Some further evidence will serve to bolster our generalizations.

First consider our claim that it is the factive complementizer which is at stake in the paradigm above. Consider data from Gungbe, in which factive clauses (1) and relative clauses (2) both make use of the same element, Rvz *ɖĕ*.

(1)   *Gungbe* (Aboh 2005: 266)
        àgásá ɖàxó lɔ́   lέ   **[ɖĕ mí wlé]**.   (factive)
        crab big   DET NUM Rvz 1.PL catch
        'the fact that we caught the [aforementioned] big crabs.'

(2)   *Gungbe* (Aboh 2005: 266)
        àgásá ɖàxó **[ɖĕ mí wlé]** lɔ́   lέ.   (relative clause)
        crab big   Rvz 1.PL catch DET NUM
        'the [aforementioned] big crabs that we caught.'

Similarly in Turkish (Turkic), the factive nominalizer *-DIK* (as opposed to non-factive *-mA/-mAK*; Bağrıaçık & Göksel 2016: 64) is also used for relative clauses (more precisely non-subject relatives, which according to Kornfilt 2008 instantiates the unmarked way to nominalize relative clauses in Turkish). This suggests a factive Comp/Rel syncretism in Turkish as well.

Second consider the fact that the Dem/Comp syncretism appears to be somewhat rare in our data, since only Germanic shows this pattern (see Roberts & Roussou 2003: §3.4; in particular see Longobardi 1991 and Ferraresi 1997; 2005 for Gothic, whose complementizers were case-inflected or relativized demonstratives). However, Heine & Kuteva (2002: 107, citing Lehmann 1982: 64) point out that a similar development (from Dem to Comp, as for Old English *þæt*) has taken place in Welsh *a*, Akkadian *ša* (<*šu*), and Nahuatl *in*. Furthermore, as mentioned above, Finnish demonstrative *tä-* 'this' and complementizer *että* may be approaching syncretism, giving us another potential Dem/Comp syncretism. In Japanese, finally, the (roughly) factive complementizer *ko-to* (original meaning 'thing'; Heine & Kuteva 2002: 295) is apparently made up of *ko-* plus the non-factive complementizer *to* (cf. Kuno 1973, among others). Interestingly, the *ko-* component seems to be the same as the Dem root *ko-* 'this'.[10]

Third, a reviewer brings up the possibility that various North-West Italian dialects systematically defy our generalization above, since demonstrative *kwel* 'that' can also be interrogative 'what' (see Munaro 2001) while Comp and Rel are both *ke*. Thus we seem to have a systematic violation of the adjacency effect seen in Table 1. However, there are various reasons to be careful here. First of all, although Munaro (2001: 282)

___

[10] Note here that it is the proximal, not the neutral, demonstrative at stake. The two are not as different as they might seem at first glance. Lander & Haegeman (2016) have shown that the neutral demonstrative is unmarked in the sense that it can have multiple different readings depending on the context, the proximal is unmarked in the sense that it involves the fewest number of spatial-deictic features.

proposes that Wh *kwe* is historically a reduced form of Dem *kwe(lo/lu)* in these dialects, the synchronic situation according to the *Atlante Italo Svizzero* (1919–1926) was that the two nevertheless were distinct and not syncretic at all (Ligurian: Dem *kwelo/kwelu/kölu* vs. Wh *cos(a)/cose/cusi*, Southern Piedmontese: Dem *lo/lu* vs. Wh *cosa*, Central Piedmontese: Dem *lon/lun* vs. Wh *kwe/kwa*, Northern Piedmontese: Dem *kul(lu)* vs. Wh *kwe*, Valdotian: Dem *(t)sò/sèn* vs. Wh *kye*; Munaro 2001: 282, his (1)). The fact remains that the demonstrative could be used to mean 'what', both in the older *AIS* data as well as in modern North-West dialects, but an important fact is that the complementizer (*che, chi, cu*, etc.) must follow the demonstrative in order for the interrogative reading to emerge (Munaro 2001: 283-284 and elsewhere). This syntactic dimension (and other complications we cannot discuss here for reasons of space) seem to preclude analyzing these forms in terms of straightforward syncretism, at least not in the sense that we mean it.

2.3 Nanosyntactic approach to syncretism

The nanosyntactic approach to syncretism (Caha 2009) is crucially based on the idea that morphosyntactic heads are cumulative, as schematized in (3).[11]

$$
(3) \qquad [_{\text{F1P}}\ F_1\ ]
$$
$$
[_{\text{F2P}}\ F_2\ [_{\text{F1P}}\ F_1\ ]]
$$
$$
[_{\text{F3P}}\ F_3\ [_{\text{F2P}}\ F_2\ [_{\text{F1P}}\ F_1\ ]]]
$$
$$
[_{\text{F4P}}\ F_4\ [_{\text{F3P}}\ F_3\ [_{\text{F2P}}\ F_2\ [_{\text{F1P}}\ F_1\ ]]]]
$$

Each of these structures can also be associated with a phonological exponent. When two (or more) structures are spelled out by the same phonological exponent, we speak of syncretism. Syncretism has been shown to be restricted to adjacent cells/layers in a paradigm, which is known as the *ABA generalization (see Bobaljik 2007; 2012; Caha 2009). Capitalizing on this adjacency restriction, by looking at attested syncretisms across languages it becomes possible to deduce the underlying linear order of functional heads at stake.

The syncretisms in Table 2, for example, necessitate the linear order in (4), where Dem is next to Comp which is next to Rel which is next to Wh which is next to Indet.

Table 2: Six crucial syncretism patterns from Table 1.

	DEM$_{\text{PRO}}$	COMP$_{\text{FACT}}$	REL$_{\text{RESTR}}$	WH$_{\text{PRO}}$	INDET$_{\text{THING}}$
English	that	that	as	what	-thing
Bulgarian	tova	deto	deto	kakvo	-shto
Yiddish	jenc	az	az	vos	-vos
(varieties of) Polish	to	że	że	co	co-
Standard Czech	to	že	co	co	-co
Finnish	tä-	että	mi-	mi-	mi-

---

[11] This idea is central to the Superset Principle, which states that a given lexically stored structure can lexicalize a syntactic structure if the lexical structure's features are a superset (proper or not) of the features in the syntactic structure (see Table 19 below for a concrete example, and Starke 2009; Caha 2009 for details). This theoretical mechanism, along with a version of the Elsewhere Principle, derives the adjacency-constrained syncretism observed.

(4)   Dem | Comp | Rel | Wh | Indet

In Table 2, we see that Bulgarian, Yiddish, and (some varieties of) Polish show that Comp and Rel must be adjacent. Standard Czech and Finnish show that Rel and Wh must be adjacent. (Non-standard) English shows that Dem and Comp must be adjacent. Finally Yiddish, (some varieties of) Polish, Standard Czech, and Finnish show that Wh and Indet must be adjacent. Hence the linear ordering in (4) is the only one which can capture these facts without any *ABA violations.

What syncretism patterns and the *ABA theorem cannot tell us, however, is which *hierarchical* order is correct, that is, whether (5a) or (5b) is correct:

(5)   a.   Dem > Comp > Rel > Wh > Indet
         b.   Indet > Wh > Rel > Comp > Dem

In Baunaz & Lander (in press; 2017a) we present reasons for believing that (5a) is the correct fseq. For the details of the argument we refer to these papers. For the purposes of this paper it is not crucial which fseq in (5) is accurate, but we will assume it to be (5a). This means that the underlying structures for Dem, Comp, Rel, Wh and Indet are the ones given in (6).
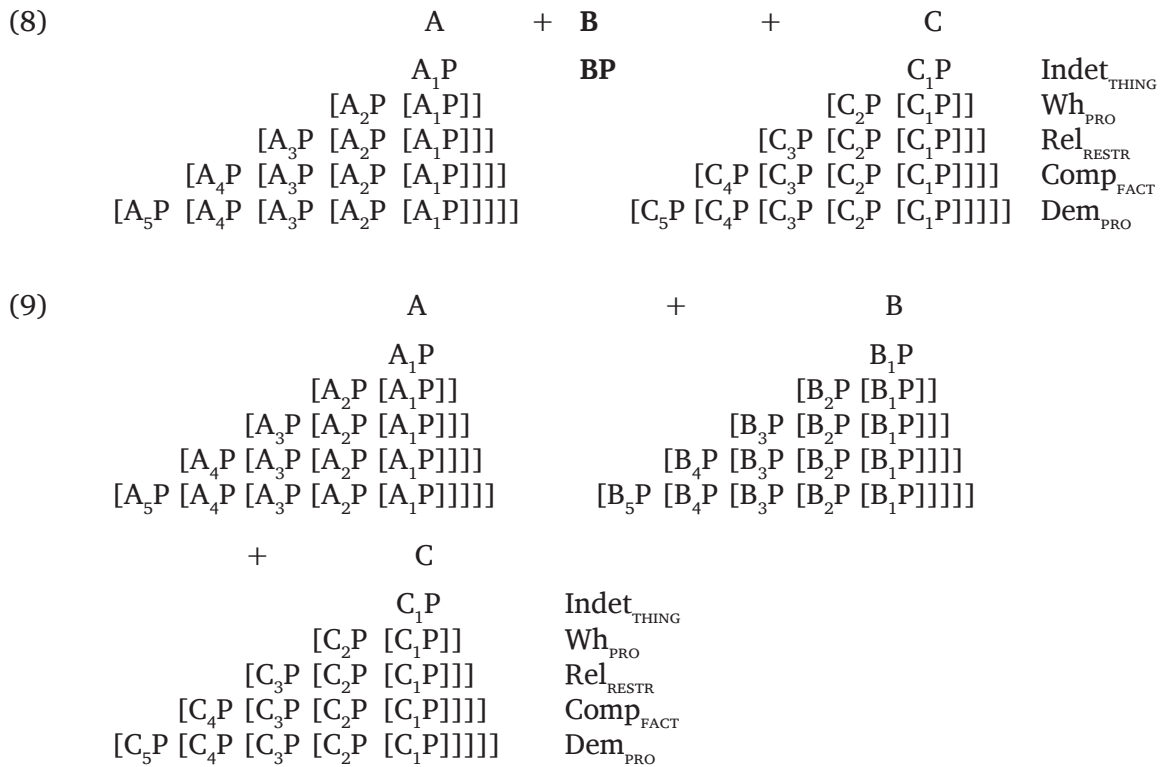
(6)
$$[_{F1P}\ F_1\ ]$$ $\quad$ $\text{Indet}_{\text{THING}}$
$$[_{F2P}\ F_2\ [_{F1P}\ F_1\ ]]$$ $\quad$ Wh_{PRO}
$$[_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1\ ]]]$$ $\quad$ $\text{Rel}_{\text{RESTR}}$
$$[_{F4P}\ F_4\ [_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1\ ]]]]$$ $\quad$ $\text{Comp}_{\text{FACT}}$
$$[_{F5P}\ F_5\ [_{F4P}\ F_4\ [_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1\ ]]]]]$$ $\quad$ Dem_{PRO}

3 A tripartite internal structure

It will be noted that the forms provided in Table 1 have not been overtly decomposed in any way (though the fact that we explicitly take a nanosyntactic approach to syncretism implies that there is a complex internal morphosyntactic structure). In this section we will endeavor to take this next step, performing a radical decomposition on these forms. The ultimate result of this will be an underlying tripartite structure.

The decomposition, furthermore, forces us to change the way we think about the syncretism patterns being tracked in Table 1. There are a number of logical possibilities regarding the nature of this syncretism once a decompositional approach is taken. If there are three morphemes per form, as we will argue below, then we must ask which morpheme participates in the cumulative, superset-subset structure-building that is at the center of the formal analysis of syncretism. It could be that only one morpheme represents the fseq Dem > Comp > Rel > Wh > Indet (e.g. (7)), or that two (e.g. (8)) or even all three (e.g. (9)) of them do, growing and shrinking in sync with each other depending on which structure is being built (Dem, Comp, Rel, Wh, Indet).

(7) $\quad$ **A** $\quad$ + $\quad$ **B** $\quad$ + $\quad$ **C**

$\qquad A_1P$ $\qquad$ **BP** $\qquad$ **CP** $\qquad$ $\text{Indet}_{\text{THING}}$
$[_{A_2P}\ [A_1P]]$ $\qquad\qquad\qquad\qquad\qquad$ Wh_{PRO}
$[_{A_3P}\ [A_2P\ [A_1P]]]$ $\qquad\qquad\qquad\qquad$ $\text{Rel}_{\text{RESTR}}$
$[_{A_4P}\ [A_3P\ [A_2P\ [A_1P]]]]$ $\qquad\qquad\qquad$ $\text{Comp}_{\text{FACT}}$
$[_{A_5P}\ [A_4P\ [A_3P\ [A_2P\ [A_1P]]]]]$ $\qquad$ Dem_{PRO}

(8)　　　　　　　　　　A　　　+　**B**　　　+　　　C

A_1P　　　　　　**BP**　　　　　　　　C_1P　　　Indet$_{THING}$
$[A_2P \ [A_1P]]$　　　　　　　　　　　$[C_2P \ [C_1P]]$　　Wh$_{PRO}$
$[A_3P \ [A_2P \ [A_1P]]]$　　　　　　　$[C_3P \ [C_2P \ [C_1P]]]$　Rel$_{RESTR}$
$[A_4P \ [A_3P \ [A_2P \ [A_1P]]]]$　　　　$[C_4P \ [C_3P \ [C_2P \ [C_1P]]]]$　Comp$_{FACT}$
$[A_5P \ [A_4P \ [A_3P \ [A_2P \ [A_1P]]]]]$　　$[C_5P \ [C_4P \ [C_3P \ [C_2P \ [C_1P]]]]]$　Dem$_{PRO}$

(9)　　　　　　　　　　A　　　　+　　　　B

A_1P　　　　　　　　　　B_1P
$[A_2P \ [A_1P]]$　　　　　　　　$[B_2P \ [B_1P]]$
$[A_3P \ [A_2P \ [A_1P]]]$　　　　　$[B_3P \ [B_2P \ [B_1P]]]$
$[A_4P \ [A_3P \ [A_2P \ [A_1P]]]]$　　　$[B_4P \ [B_3P \ [B_2P \ [B_1P]]]]$
$[A_5P \ [A_4P \ [A_3P \ [A_2P \ [A_1P]]]]]$　$[B_5P \ [B_4P \ [B_3P \ [B_2P \ [B_1P]]]]]$

　　　　+　　　　C

C_1P　　　　　Indet$_{THING}$
$[C_2P \ [C_1P]]$　　Wh$_{PRO}$
$[C_3P \ [C_2P \ [C_1P]]]$　Rel$_{RESTR}$
$[C_4P \ [C_3P \ [C_2P \ [C_1P]]]]$　Comp$_{FACT}$
$[C_5P \ [C_4P \ [C_3P \ [C_2P \ [C_1P]]]]]$　Dem$_{PRO}$

Indeed, the complexity of the situation increases greatly once multiple morphemes, each with their own potential internal structure and behavior, are taken into account. Not even all logical possibilities are presented in (7–9), moreover. For example, the different sequences (A and C in (8), for instance) do not necessarily have to be dependent on each other (e.g. if A builds up to A_3 then C must also build up to C_3) but could be partially or totally independent of one another. In any case, the point should be clear that allowing for multiple morphemes complicates the clean picture of syncretism presented in Table 1. Fortunately in this case, we will argue that the simplest of these scenarios – the one sketched in (7), with a single fseq and two invariant bits of structure – is at stake for the data in Table 1.

3.1 Uncovering the basic tripartition

3.1.1 Germanic

The elements in Table 1 can be decomposed further, strongly suggesting that they have a complex internal structure. Taking a look at English, it is clear that the Dem items can be segmented into two parts, as seen in (10). (For now we provide forms in their standard orthography, but more phonologically precise representations are given later in the paper).

(10)　　*English*
　　　　Dem　　**th-**at
　　　　　　　th-is
　　　　　　　th-ese
　　　　　　　th-ose
　　　　　　　th-ere
　　　　　　　th-us

This is also the case for the Dem forms of the other Germanic languages (except Yiddish, for which see below). Some examples are given in (11).

(11)   a.   *Icelandic*
         Dem **þ**-að   'that'
                þ-etta   'this' (NEUT.SG)
                þ-essi   'this' (MASC/FEM.SG)

      b.   *Swedish*
         Dem **d**-et   'that' (NEUT.SG)
                d-en   'that' (COMMON.SG)
                d-etta 'this' (NEUT.SG)
                d-enna 'this' (COMMON.SG)
                d-essa 'these'
                d-om   'those'

      c.   *Dutch*
         Dem **d**-at   'that' (NEUT.SG)
                d-ie   'that' (COMMON.SG)
                d-it   'this' (NEUT.SG)
                d-eze 'this' (COMMON.SG)
                d-eze 'these'
                d-ie   'those'

      d.   *German*
         Dem **d**-as   'that' (NEUT.SG)
                d-er   'that' (MASC.SG)
                d-ie   'that' (FEM.SG)
                d-ie   'those'

There is a common trend of analyses which argue in favor of the initial element in Germanic being an instantiation of definiteness/the definite article (D) (see Déchaine & Wiltschko 2002; Kayne 2005; Kayne & Pollock 2010; Roehrs 2010; Leu 2015, among others). This would mean that D is contained within Dem.

    Furthermore, a *wh*-marker is also found throughout Wh-items in Germanic, as seen in (12).

(12)   a.   *English*
         Wh **wh-**at
                wh-ich
                wh-o
                wh-en
                wh-ere

      b.   *Icelandic*
         Wh **hv**-að   'what'
                hv-aða   'which'
                hv-er   'who'
                hv-ernig 'how'
                hv-ar   'where'
                hv-enær 'when'

      c.   *Swedish*
         Wh **v**-ad   'what'
                v-ilket 'which'
                v-em   'who'
                v-ar   'where'

    d.   *Dutch*
Wh   **w**-at   'what'
        w-elk-   'which'
        w-ie   'who'
        w-aar   'where'
        w-anneer 'when'

    e.   *German*
Wh   **w**-as   'what'
        w-elch-   'which'
        w-er   'who'
        w-ie   'how'
        w-o   'where'

    f.   *Yiddish*
Wh   **v**-os   'what, which'
        v-er   'who'
        v-i   'how'
        v-u   'where'
        v-en   'when'

Focusing now on the specific elements which participate in the syncretisms being tracked in Table 1 – i.e. the neutral neuter/inanimate singular forms (Eng. *that/what,* Du. *dat/wat,* Ger. *das/was,* etc.) – we can note that the prefix is variable, that is, alternating between D and Wh morphemes depending on function, while the rest of the form is (in most languages) invariant, that is, not alternating depending on function (e.g. Icel. *-að,* Du. *-at,* Ger. *-as*). In other words we can speak of a basic bimorphemic structure as in Table 3, where the first part displays a morphological alternation and the second part remains stable. In Table 3 we label these two components F (for 'functional') and Base.

The decomposition in Table 3, however, is only an approximation. It can be seen that Base in Table 3 actually contains (at least) two elements: a vowel and a consonant. Taking Dutch and German, the consonant (Du. *-t,* Ger. *-s*) can be identified with frozen inflectional/agreement (Φ) morphology (cf. German strong adjective ending for neuter nominative/accusative, i.e. *-es/-s*; see also Leu 2015, for instance). This leaves the vowel (Du. **-ɑ-,** Ger. **-a-**) as the "true" realization of what we have called Base.

Obviously, since the non-decomposed Base in Table 3 was invariant, each of the individual components resulting from decomposition in Table 4 inherit this property of invariance. That is, neither Du. **-ɑ-** / Ger. **-a-** nor Du. *-t* / Ger. *-s* show any kind of morphophonological alternation throughout the paradigm.

Even languages with (apparently) more complex systems can be seen to fall into a tripartite structure with invariant Base and Inflection components. The relevant Icelandic forms are given in Table 5, where we note that *sem* is very likely a portmanteau that is not overtly decomposable (though the fact that *að* can cooccur with it might put this part

Table 3: Bimorphemic structure in Germanic.

	F	Base
Icelandic	θ- ~ kʰv-	-að
Dutch	d- ~ ʋ-	-ɑt
German	d- ~ v-	-as

of the relativizer on a par with the other **-a-** cores identified in the table). Here and in the tables that follow gray shading will be used to highlight portmanteau elements.

First, *-ð* is uncontroversially the neuter singular inflectional marker in Icelandic, and not surprisingly it remains stable in Table 5. As for the Base, besides *sem* (which we have taken to be a portmanteau), the only obstacle to Icelandic showing us another instantiation of an invariant Base seems to be the long vowel in the Wh/Indet pronoun (i.e. Wh /kʰv-**aː**-ð/ vs. short /-a-/ elsewhere). However, this is a non-issue since vowel length is a non-contrastive, predictable (hence allophonic) property in Icelandic. Thus there is no real phonological reason stopping us from equating this long **-aː-** with the short **-a-** seen elsewhere in the paradigm.

Yiddish too, with some additional investigation, can be seen to fall in line with the expectation of an invariant Base and Inflection. We will begin with the Base. In Table 6, we see that **-o-** is a plausible candidate for the realization of an invariant Base in Yiddish. Once again, the apparently deviant form, this time the demonstrative pronoun *jenc*, can

Table 4: Trimorphemic decomposition in Dutch and German.

		F	Base	Inflection
Dutch	**Dem**	d-	**-ɑ-**	-t
	Comp	d-	**-ɑ-**	-t
	Rel	d-	**-ɑ-**	-t
	Wh	ʋ-	**-ɑ-**	-t
	Indet	ʋ-	**-ɑ-**	-t
German	**Dem**	d-	**-a-**	-s
	Comp	d-	**-a-**	-s
	Rel	d-	**-a-**	-s
	Wh	v-	**-a-**	-s
	Indet	v-	**-a-**	-s

Table 5: Trimorphemic decomposition in Icelandic.

		F	Base	Inflection
Icel.	**Dem**	θ-	**-a-**	-ð
	Comp	∅	**a-**	-ð
	Rel	sɛːm ~ sɛm (+ að)		
	Wh	kʰv-	**-aː-**	-ð
	Indet	kʰv-	**-aː-**	-ð

Table 6: Trimorphemic decomposition in Yiddish.

		F	Base	Inflection
Yiddish	**Dem**	*jen(t)-*		-s
	Comp	v-	**-o-**	-s
	Rel	v-	**-o-**	-s
	Wh	v-	**-o-**	-s
	Indet	v-	**-o-**	-s

plausibly be analyzed as a portmanteau *jent-* (consisting of both F and Base). Note in particular that the ending */-ts/* in Dem *jenc* could be seen to deviate from the regular *-s* in the rest of the paradigm. Interestingly, Jacobs (2005: 112) writes that "[s]urface phonetic affricates are frequently found between consonants *l*, *n*, and a following sibilant" (but then immediately following with "though this might be seen as a historical process"). In other words, depending on one's analysis of such affricates in Yiddish, one could defend the view that */-ts/* should be analyzed simply as */-s/* in this case, with the (predictable) insertion of the stop */-t-/* in this particular environment (i.e. */jen-s/* > *jenc* [jents]), bringing this ending in line with the rest of the paradigm as well.

Even if the process of *t*-insertion is historic and not a part of the active phonological processes, one can still say that the *-t-* is part of the portmanteau.[12] Thus *jent-* is not decomposable and thus cannot be expected to show the Base.

3.1.2 Slavic

The discussion of Yiddish *jen(t)-* leads us quite smoothly into Slavic, where a packaging similar to *jent-* is on full display. In order to set the background, we start from SC *što* and Ru. *čto* (we leave Bulgarian and Macedonian for future work), where we observe the same tri-morphemic template as in West Germanic: SC */ʃ-t-o/* and Ru. */ʂ-t-o/*. This is shown in Table 7. Historically, */ʃ/* and */ʂ/* derive from palatalization of the *wh*-morpheme *k-* before a front vowel (i.e. Proto-Balto-Slavic *ki-to > Proto-Slavic *čь-to 'what'). The second consonant *t-* is the demonstrative root, and *-o* is the neuter singular inflection.

Polish and Czech, however, do not fit in as neatly. While Dem *t-o* can be decomposed just as in Serbo-Croatian and Russian, Comp *że* */ʐɛ/* (Polish), *že/ʒe/* (Czech) and Rel/Wh/Indet *co* /t͡so/ (both Polish and Czech) are less straightforward: not only are the consonants different, but in Comp the vowel is also divergent (i.e. Comp */-e/* vs. */-o/* in Dem, Rel, Wh, and Indet). We think a natural approach for Polish and Czech would be to analyze the initial affricate /t͡s-/ in Rel/Wh/Indet *co* /t͡so/ as a portmanteau of F and Base (like Yiddish *jen-*), as shown in Table 8.

Furthermore, the alternation between *-e* and *-o* in Table 8 also falls out in a completely regular way, since *-e* is in fact an allophone of *-o* after "soft" consonants (e.g. Po. *ż-* /ʐ/ and Cz. *ž-* /ʒ/). This means that Polish Comp *ż-e* and Czech complementizer *ž-e* have exactly

Table 7: Trimorphemic decomposition in Serbo-Croatian and Russian.

		F	Base	Inflection
Serbo-Croatian	**Dem**	(∅)[13]	t-	-o
	Comp	ʃ-	-t-	-o
	Rel	ʃ-	-t-	-o
	Wh	ʃ-	-t-	-o
	Indet	ʃ-	-t-	-o
Russian	**Dem**	(∅)	t-	-o
	Comp	ʂ-	-t-	-o
	Rel	ʂ-	-t-	-o
	Wh	ʂ-	-t-	-o
	Indet	ʂ-	-t-	-o

---

[12] We would like to thank a reviewer for this suggestion.

[13] Note the lack of the F ingredient in the Slavic Dem forms (e.g. SC *t-o* vs. *š-t-o* elsewhere). This is the case in Serbo-Croatian, Russian, Polish, and Czech, but not, crucially, in Macedonian and Bulgarian. We refer to Baunaz & Lander (in press) for details and fuller discussion of this so-called "Slavic containment puzzle".

the same basic structure as Serbo-Croatian and Russian Comp/Rel/Wh/Indet *što/čto* above, with the same neuter singular ending as well.

In sum, in Polish and Czech both Comp *ž-e* and Rel/Wh/Indet *c-o* are underlyingly tripartite structures, with the initial affricate being analyzed as a portmanteau morpheme.

3.1.3 Romance

If we start by looking at French, though, it is clear that it is amenable to an analysis like the one provided for Slavic above, where /k-/ in *que* /kə/ is a bigger chunk of structure, composed of both F and the Base component. Indeed, the same can be posited for *ce* /sə/, as shown in Table 9.[14]

If we had taken /s-/ or /k-/ to be the Base, then the F column would require a null morpheme, which we take to be an undesirable analysis, preferring to see these consonants as portmanteau morphemes encoding both the F ingredient and the Base.[15] However, an unfortunate result of the portmanteaux in French is that the Base *cannot be overtly identified*, since decomposition is not possible.

Turning now to Italian, the same analysis can, simply for the sake of parsimony, be given for It. *che* /k-e/. As for Dem *quello* /kwello/, on the other hand, there is reason to

Table 8: Trimorphemic decomposition in Polish and Czech.

		F	Base	Inflection
Polish	Dem	(∅)	t-	-o
	Comp	ʐ-		-e
	Rel	t͡s-		-o
	Wh	t͡s-		-o
	Indet	t͡s-		-o
Czech	Dem	(∅)	t-	-o
	Comp	ʒ-		-e
	Rel	t͡s-		-o
	Wh	t͡s-		-o
	Indet	t͡s-		-o

Table 9: Trimorphemic decomposition in French.

		F	Base	Inflection
French	Dem	s-		-ə
	Comp	k-		-ə
	Rel	k-		-ə
	Wh	k-		-ə
	Indet	k-		-ə

---

[14] A slight complication is that M.SG *ce* /sə/ becomes *cet* /sɛt/ when the next word begins with a vowel (e.g. *ce garçon* but *cet ami*). Whatever the ultimate account of this alternation should be, we note that it resembles the *a* vs. *an* alternation in English, which not only is conditioned by a following vowel-initial word but also may involve a difference in vowel quality (*a* /ə/, /ʌ/, or /eɪ/ vs. *an* /ən/, /ʌn/, or /æn/).

[15] We have argued elsewhere that the Dem forms *t-o* in Slavic truly lack an article (so *t-* is not a portmanteau of F and Base, but rather just a realization of Base alone). Baunaz & Lander (in press) argue that it is the non-availability of a definite article which allows these forms to exist in precisely those Slavic languages without definite articles. Since French does have a definite article, however, *c-e* cannot be treated in the same way.

think that it should be segmented slightly differently, i.e. /kwe-llo/. Indeed, as seen in (11), Italian demonstratives overtly contain the definite article.

(11)   *Dem*   *Def*

       M.SG   quel-**lo**   **lo**   (+ word-initial *sC*- or *z*-)
                [quel]   il   (+ word-initial other *C*-)
       F.SG   quel-**la**   **la**
       M.PL   que-**gli**   **gli**   (PL of *lo*)
                que-**i**   **i**   (PL of *il*)
       F.PL   quel-**le**   **le**

Now, if we identify Def with the F slot, then we can keep Italian on a par with much of Germanic by putting -*(l)lo* here as well. The leftover morpheme, namely /kwe-/, appears to be quite large. We can offer two possibilities for /kwe-/: either it can be segmented as /kw-/ plus /-e/, or it is a portmanteau made up of both Base and Inflection. The first option is more interesting: not only does it leave us with a regular /-e/ ending in the right-hand column throughout the paradigm, but more importantly it means that the nominal core emerges in the Dem form (recall that it remained hidden in French above) as **kw-**. Therefore we have shown this first option in Table 10 below. The Base emerges in Dem in Italian.

    Finally we can take a look at the Romanian forms, where it seems possible to uncover a realization of the Base. First of all, we might assume that Comp *că* /kə/ has the same underlying structure as in French and Italian, namely that the initial stop is a portmanteau of F and Base. However, since the vowel is different in Comp (schwa rather than /e/), it would be better to analyze the entire form as an idiomatic portmanteau of F, Base, and Inflection. This approach allows us to posit that the vowel /e/ is a completely regular morpheme in the rest of the paradigm. Furthermore, Dem contains the definite article in Romanian, as it does in Italian. See Tables 11–12 (from Savu & Bican-Miclescu 2012).

Table 10: Trimorphemic decomposition in Italian.

		F	Base	Inflection
	Dem	-(l)lo	**kw-**	-e
	Comp	k-		-e
Italian	**Rel**	k-		-e
	Wh	k-		-e
	Indet	k-		-e

Table 11: Romanian Dem 'that'.

	M.SG	**F.SG**	**M.PL**	**F.PL**
NOM/ACC	a.ˈtʃe**l**	a.ˈtʃe̯**a**	a.ˈtʃe**j**	a.ˈtʃe.**le**
GEN/DAT	a.ˈtʃe.**luj**	a.ˈtʃe.l**ej**	a.ˈtʃe.**lor**	a.ˈtʃe.**lor**

Table 12: Romanian Def 'the'.

	M.SG	**F.SG**	**M.PL**	**F.PL**
NOM/ACC	-ul	-a	-j	-le
GEN/DAT	-luj	-ej	-lor	-lor

This means that *-l* in *(a-)cel* /tʃ-e-l/ needs to be put in the F column. Finally we abstract away from *a-* in *acel* since it is droppable in the vernacular. As seen in Table 13, this leaves us with a nominal core **tʃ-**.

As we will explain below, /tʃ-/ in Rel, Wh, and Indet is technically *not* an overt realization of the Base; rather, in these forms /tʃ-/ must be a portmanteau of both F and Base. Only in the Dem form is /tʃ-/ an overt realization of the Base in Romanian. See section 5.3 for further details.

3.1.4 Back to English

We now turn to English, with its even more complex (and seemingly problematic) paradigm. In Table 14 we see the English forms, and despite what seems to be a high degree of decompositionality, there is some variation in what is expected to be the invariant Base ingredient: **-æ-** or **-ʌ-/-ɒ-**?

One interesting fact relevant for our purposes here concerns the marking of distal/neutral vs. proximal. Note that this distinction in English is actually accomplished by the contrast /-æt/ vs. /-ɪs/ (rather than, say, /-æ-/ vs. /-ɪ-/).[16] In other words, there is reason to want to keep /-æt/ in *that* as a unit rather than splitting it up into two parts (i.e. /-æ-/ plus /-t/). For Comp and Rel, however, which are not meant to encode spatial deixis, this reasoning does not apply.

There is also an important detail about Comp and Rel that sets them apart from Dem, in that they have reduced forms that Dem does not (i.e. /ðət/, as in *The fact* /ðət/ *he bought new clothes…* or *the man* /ðət/ *came to dinner* but not */ðət/ girl is my daughter*). Interestingly, the first part of this reduced form is homophonous with the definite article in English, i.e. *the* /ðə/. (Note that, at least for Rel, affixation of the definite article – or at least something syncretic with the definite article – is crosslinguistically common: Fr. *le-quel*, Hu. **a-mi**, MG ó-*pu* and even Comp **o-ti**.) In other words, we have reason to group /ðə-/ as a unit in Comp and Rel (in the same way we grouped /-æt/ as a unit for Dem). Indet is not syncretic with anything: because it is not overtly decomposable and also quite large (i.e. about the same size as the lexical noun *thing*), we propose that it is a portmanteau of F, Base, and Inflection. This brings us to the revised table for English, in Table 15.

Table 13: Trimorphemic decomposition in Romanian.

		F	Base	Inflection
	Dem	-l	**tʃ-**	-e
	Comp	kə		
Rom.	**Rel**	tʃ-		-e
	Wh	tʃ-		-e
	Indet	tʃ-		-e

Table 14: Trimorphemic decomposition in English 1.0.

		F	Base	Inflection
	Dem	ð-	**-æ-**	-t
	Comp	ð-	**-æ-**	-t
English	**Rel**	ð-	**-æ-**	-t
	Wh	(h)w-	**-ʌ-** (Am.)	-t
			-ɒ- (UK)	

---

[16] However, the case could be made for overt tripartition in the *plural*, since /-z/ serves to mark plural in both distal/neutral and proximal: /ð-oʊ-**z**/ (Am.), /ð-əʊ-**z**/ (UK) vs. /ð-iː-**z**/.

Table 15: Trimorphemic decomposition in English 2.0.

		F	Base	Inflection
English	**Dem**	ð-	-æt (**vs.** -ɪs)	
	Comp	ðæ- ~ ðə-		-t
	Rel	ðæ- ~ ðə-		-t
	Wh	(h)w-	**-ʌ-** (Am.)	-t
		(h)w-	**-ɒ-** (UK)	-t
	Indet	-thing		

As seen in Table 15, the Base now reveals itself in the interrogative pronoun, namely the vowel **-ʌ-** in American varieties and **-ɒ-** in UK varieties.

3.1.5 Interim conclusion

We have argued for a tripartite underlying structure for Dem, Comp, Rel, and Wh items. Crucially, in various cases it can be shown with some further investigation that languages which do not seem to fit the tripartite mold at first actually are not counterexamples to our hypothesis at all. Instead, these languages just require a more careful approach, where complicating factors about phonological processes or structural considerations concerning possible portmanteau morphemes (which are typically treated as "phonological idioms" in nanosyntax) are kept in mind.

Since the Base is invariant, it will not grow and shrink in the typical subset-superset manner that nanosyntactic theory makes use of in its account of syncretism (Caha 2009) – but if we are correct about the Base, then there is no syncretism to account for anyway. We will take a similar stance towards the frozen set of Φ features (Inflection) as well, namely that it is a constant, invariant piece of structure. This leaves us with F. That is, F is the source of the syncretism patterns we have identified in a number of languages in Table 1 above. It is this part which involves cumulative structure-building in the nano-syntactic sense. In the languages we have considered here, then, we end up with an F that alternates based on the particular function (Dem, Comp, Rel, Wh, or Indet) and two invariant ingredients, Base and Infl.

3.2 The underlying structure

For the items discussed so far we can propose the basic merge order in (12a), which entails the specific structures in (12b).[17]

---

[17] A reviewer expresses some worries about morphological decomposition and the nature of the fine-grained grammatical features at stake here: what does it mean for a demonstrative pronoun to be composed of Comp, Rel, and Wh features? Though we do not have a full answer to this question, we do have two relevant comments to make. (i) Our data seem to point to a view in which these elements are at their core actually much simpler, syntactico-semantically speaking, than usually thought, basically made up of a noun with some functional architecture added. Additional syntactic functions and semantic readings probably result from these nominal elements entering into relations during the syntactic derivation with other elements (e.g. the relationship established between an antecedent and its relative pronoun which can be analyzed in terms of movement and agreement). Fleshing this idea out more, it appears that we are actually tracking D features rather than operator features. Icelandic *hv-að* 'what' vs. *eitt-hv-að* 'something', for example, show that *hv-* is not an operator in both since in the latter form *eitt-* 'some-' takes on this role. This would mean that *hv*-words are more rightly indefinites, as in Japanese, i.e. they get their quantificational meaning via an additional quantificational particle of some kind (which can perhaps be null in Icelandic). Thus the features responsible for building *hv-* seem to be different from the features responsible for building quantificational particles. The "Wh" feature we are looking at, then, does not necessarily trigger an interrogative meaning but is the necessary basis for an operator to enter into the configuration. (ii) On the other hand, the range of functions which we are investigating should not be *over*estimated. In fact, we have isolated only one thin "slice" of the syntactico-semantic realm to which these elements belong. Dem, Comp, Rel, and Wh do not overlap in every direction, but happen to do so when these particular functions (again, distal or

(12)   a. F-domain > Base

b.
$$F_1 > n \qquad \text{Indet}_{\text{THING}}$$
$$F_2\ F_1 > n \qquad \text{Wh}_{\text{PRO}}$$
$$F_3\ F_2\ F_1 > n \qquad \text{Rel}_{\text{RESTR}}$$
$$F_4\ F_3\ F_2\ F_1 > n \qquad \text{Comp}_{\text{FACT}}$$
$$F_5\ F_4\ F_3\ F_2\ F_1 > n \qquad \text{Dem}_{\text{PRO}}$$

While Infl can be identified with a set of Φ features, we still have not said anything explicit about Base. As seen in (12), we suggest that Base is a (semi-)lexical category on top of which Φ[18] and the functional features used for building Dem, Comp, Rel, Wh, and Indet items are merged. In other words, Base is a noun of sorts, which we will label the *nominal core* (*n*).

We have argued that *n* is invariant. Because there is no syncretism to account for in *n*, then, there is no reason to think that its structure will grow and shrink in the subset-superset way typical of syncretism (Caha 2009). We propose to take a similar stance towards the Inflection part, that is, the frozen set of Φ features: it is a constant, invariant piece of structure. F, on the other hand, is apparently the source of the syncretism patterns we have identified in a number of languages above (see Table 1). It is this part which involves cumulative structure-building.

3.2.1 Germanic

These results can now be represented in Tables 16–19 below, where F is now represented with a larger cell in the Dem row, and then gradually smaller down to Indet. Reflecting the basic facts in Table 1, F is either syncretic with Dem, Comp, Rel (German, Dutch), or with Comp, Rel, Wh, Indet (Yiddish), pointing to the hierarchy Dem > Comp > Rel > Wh > Indet.

Table 16: Identifying the nominal core in German and Dutch.

			F			*n*	Infl
German	**Dem**	d-				**-a-**	-s
	Comp		d-				
	Rel			d-			
	Wh				v-		
	Indet					v-	
Dutch	**Dem**	d-				**-ɑ-**	-t
	Comp		d-				
	Rel			d-			
	Wh				ʋ-		
	Indet					ʋ-	

---

neutral Dem, emotive factive Comp, indeclinable restrictive Rel, and interrogative pronoun Wh, all with neuter/inanimate singular inflection) are considered. Thus there is a bigger question of how functional sequences intersect with one another to create such multidimensional paradigms, with syncretism patterns going in multiple directions (for relevant discussion see Vanden Wyngaerd to appear).

[18] If we want to keep the basic merge order of F-domain > *n*, then we must allow for the constituent of Φ features to be merged in different spots in the structure. In English, *n* and Φ can form a portmanteau, this implies a constituent made up of *n* and Φ to the exclusion of F, meaning Φ is merged between F and *n*: [F… **[Φ [*n*]]**]; in other languages F and *n* make up a portmanteau morpheme to the exclusion of Φ, meaning that Φ must be merged on top of F: [Φ **[F… [*n*]]**].

The nominal core in these five languages has also been revealed as a vowel: Ger. **-a-**, Du. **-ɑ-,** Icel. **-a-**, Yid. **-o-** and English **-ʌ-** (Am.) and **-ɒ-** (UK). Inflection is overtly realized as the 3rd person neuter morpheme of these languages.

3.2.2 Slavic

As for German and Dutch, decomposing the relevant forms into three parts in Serbo-Croatian and Russian does indeed result in a plausible candidate for the "genuine" nominal core (SC/Ru. **-t-**). Syncretism in the F column is with Comp, Rel, Wh and Indet (just like in Yiddish above). Following our analysis from section 3.1.2 above, we propose to represent SC and Russian as in Table 20.

Recall that in Polish and Czech, the initial affricate /t͡s-/ in Rel/Wh/Indet *co* /t͡so/ is a portmanteau of F and Base (like Yiddish *jen-*). The only instance where the Base is overtly realized is with Dem *t-o* (see also fn. 15) in both languages. This is represented in Table 21.

3.2.3 French and Italian

In section 3.1.3 we argued for an analysis for French amenable to the one provided for Slavic, where /k-/ in *que* /kə/ is a bigger chunk of structure, composed of both F and *n*, i.e. /k-/ is a portmanteau. In French, thus, the nominal core *cannot be overtly identified*,

Table 17: Identifying the nominal core in Icelandic.

			F		*n*	Infl
Icel.	Dem		θ-			
	Comp		∅			
	Rel		sɛːm ~ sɛm			
	Wh		kʰv-		(-)a(ː)-	-ð
	Indet			kʰv-		

Table 18: Identifying the nominal core in Yiddish.

		F		*n*	Infl
Yiddish	Dem	jen(t)-			-s
	Comp		v-	-o-	
	Rel		v-		
	Wh		v-		
	Indet		v-		

Table 19: Identifying the nominal core in English.

		F		*n*	Infl
English	Dem	ð-		-æt (**vs.** -ɪs)	-t
	Comp		ðæ- ~ ðə-		
	Rel		ðæ- ~ ðə-		
	Wh	(h)w-		-ʌ- (Am.)	
				-ɒ- (UK)	
	Indet			-thing	

Table 20: Identifying the nominal core in Serbo-Croatian and Russian.

			F		n	Infl
Serbo-Croatian	**Dem**			(∅)	**(-)t-**	-o
	Comp		ʃ-			
	Rel		ʃ-			
	Wh		ʃ-			
	Indet		ʃ-			
Russian	**Dem**			(∅)	**(-)t-**	-o
	Comp		ʂ-			
	Rel		ʂ-			
	Wh		ʂ-			
	Indet		ʂ-			

Table 21: Identifying the nominal core in Slavic.

		F	n	Infl
Polish	**Dem**	(∅)	**t-**	-o/-e
	Comp		ʐ-	
	Rel		t͡s-	
	Wh		t͡s-	
	Indet		t͡s-	
Czech	**Dem**	(∅)	**t-**	-o/-e
	Comp		ʒ-	
	Rel		t͡s-	
	Wh		t͡s-	
	Indet		t͡s-	

since decomposition is not possible. The same can be posited for *ce* /sə/, as shown in Table 22.

Recall that we argued for a similar analysis for Italian, with Dem *quello* /kwello/ segmented slightly differently, i.e. /kwe-llo/. Morphologically we argued that *-(l)lo* occupies the F position, while /kwe-/, being quite large, should be segmented as /kw-/ plus /-e/: not only does it leave us with a regular /-e/ ending in the right-hand column throughout the paradigm, but more importantly it means that the nominal core emerges in the Dem form (recall that it remained hidden in French above) as **kw-**. This is shown in Table 23.

3.2.4 Romanian

Romanian Comp *că* /kə/ has been analyzed as an idiomatic portmanteau of F, *n*, and Infl. The vowel /e/ in Dem *acel*, Rel, Wh and Indet *ce* is a completely regular morpheme in the rest of the paradigm. Also Dem contains the definite article in Romanian, as it does in Italian, meaning that *-l* in *(a-)cel* /tʃ-e-l/ needs to be put in the F column. As seen in Table 24, this seems to leave us with a nominal core **tʃ-**.

Note here that the /tʃ-/ elements appearing in Rel, Wh, and Indet are technically *not* overt realizations of the nominal core, but portmanteau morphemes of both F and *n*. If they were not realizations of both F and *n* and only represented *n*, then the F column would be left empty for no (apparent) reason in Rel, Wh, and Indet.

Table 22: Nominal core not retrievable in French.

		F	*n*	Infl
French	Dem		s-	-ə
	Comp		k-	
	Rel		k-	
	Wh		k-	
	Indet		k-	

Table 23: Nominal core emerges in Dem in Italian.

		F	*n*	Infl
Italian	Dem	-(l)lo	kw-(?)	-e
	Comp		k-	
	Rel		k-	
	Wh		k-	
	Indet		k-	

Table 24: Nominal core emerges in Dem in Romanian.

		F	*n*	Infl
Rom.	Dem	-l	tʃ-	-e
	Comp		kə	
	Rel		tʃ-	-e
	Wh		tʃ-	-e
	Indet		tʃ-	-e

In the Dem form specifically, however, /tʃ-/ is in fact an overt realization of the nominal core. Consider (13).

(13) Using the lexical entry for /tʃ-/ to spell out [*n*]

a. Lexical entries $< [F_4 [F_3 [F_2 [F_1]]]] + [n] \Leftrightarrow /tʃ-/ >$
$< [F_5 [F_4 [F_3 [F_2 [F_1]]]]] \Leftrightarrow /-l/ >$

b. Structures $[F_4 [F_3 [F_2 [F_1]]]]$ + $[n]$ => /tʃ-/[19] (Comp)
$[F_3 [F_2 [F_1]]]$ + $[n]$ => /tʃ-/ (Rel)
$[F_2 [F_1]]$ + $[n]$ => /tʃ-/ (Wh)
$[F_1]$ + $[n]$ => /tʃ-/ (Indet)
$[F_5 [F_4 [F_3 [F_2 [F_1]]]]]$ $[n]$ (Dem)
=> /-l/ => /tʃ-/

As shown in (13b), the affricate /tʃ-/ spells out both F and the nominal core in Comp, Rel, Wh, and Indet. In Dem, however, F is spelled out as the definite article /-l/, leaving the nominal core [*n*] still without a spellout. If there had been a specific lexical entry for the structure [*n*], then it would be expected to surface here. However, apparently the

---

[19] Though recall that this will later be overwritten by a specific idiomatic lexical entry for /kə/, once Inflection has been added.

Romanian lexicon lacks such an entry, and has to make do with the lexical entry for /tʃ-/ in order to spell out [*n*] (a perfectly legal option afforded by the Superset Principle). Thus /tʃ-/ is an overt realization of the Romanian nominal core in precisely the Dem form.

4 The "silent category" hypothesis

Kayne (2005) develops the idea that certain functional categories (PLACE, THING, YEARS, MUCH, VERY, COLOR, among others) may be unpronounced/non-overt while nevertheless being universally present in the syntax.[20] Consider the functional noun HOURS (small capitals indicate non-pronunciation), which is required to be pronounced in French but cannot be overt in English.

(14)   What time is it?
          a.   It's 3 HOURS.
          b.   Il est 3 *(heures).   (cf. Kayne 2005: 258–260)

Another example from Kayne is that English locative *here* and *there* are "simply the demonstrative *here* and *there* that are embedded in a larger DP with unpronounced noun and determiner" (2005: 67), i.e. THIS *here* PLACE and THAT *there* PLACE.

   Kayne's (2005) influential approach to modeling crosslinguistic variation in terms of the same underlying structure with variation reducing to different elements receiving overt pronunciation from language to language is a mission we are, of course, highly sympathetic to, and various overlaps can be found between our two approaches. We also think the two approaches yield different predictions, one of which we will delve into in this section.

   It has been suggested before (Garzonio & Poletto 2012; 2017) that nominal classifiers such as THING can be found inside quantifiers, and that these are sometimes overtly realized (e.g. Italian *qualche cosa, qualcosa*, etc.). It is important to recognize that we have identified an even smaller component in our (classifier-like) nominal core (*n*). Recall from above that It. *che* was decomposed as $[_{F + n}$ k- $[_{Infl}$ -e]], meaning that *n* is part of the initial consonant /k-/ and apparently not even part of *cosa* at all. Thus our understanding of *n* is that it must be distinct from – and in fact much smaller than – the nominal classifiers discussed in other work. Our interpretation is that *n* is a layer in a very fine-grained functional structure that serves to classify an element as THING, PERSON, etc.[21] However, nominal restrictions like It. *cosa* or Fr. *chose* are larger, consisting of both *n* and (semi-) lexical N. N in this case is more rightly described as semi-lexical, since it has limited inflectional capacities.

   Furthermore, the Kaynean approach is unclear on what exactly the difference between, say, It. *che* and *che cosa* (both meaning 'what') or Eng. *that* (as a standalone pronoun) and *that thing* might be. The important observation for researchers like Kayne and Garzonio & Poletto appears to be that the options with *cosa* and *thing* reveal an underlying classifier which is otherwise non-overt, providing support for underlying structures like *what* THING or *that* THING. While we agree that this is basically the case and therefore an important general point in favor of positing a more abstract underlying structure, there is more to the story. Notice that in our approach it is not possible to model both *che* and

---

[20] We would like to thank an anonymous reviewer for comments which led to the writing of this section.

[21] We believe that *n* itself has some internal functional structure as well. For the purposes of this paper it will be sufficient to assume that *n* can simply come in different "flavors" (e.g. n_{FORM}, n_{BODY}, n_{THING}, n_{PLACE}, etc.) rather than elaborating a full functional hierarchy relating to the different kinds of *n* (but see Baunaz & Lander 2017b for a nanosyntactic perspective on "dummy" nouns or *ontological categories*, as they are called in the functional literature; see Haspelmath 1997; Diessel 2003; Cysouw 2004).

che cosa as parallel structures of the same size. For *che* we are already committed to the idea that *ch-* spells out F and *n*, and that *-e* spells out Infl; if we now have *che cosa*, we still need F, *n*, and Infl for *che*, leaving no available structure for *cosa*. We are forced, then, to give *che* and *che cosa* different structures, with one possibility (with *cosa* corresponding to a constituent made up of *n* and semi-lexical N) in (15).

(15) a. $[_{\text{F} + n}\ ch\text{-}\ [_{\text{Infl}}\ \text{-}e]]$
      b. $[_{\text{F} + n}\ ch\text{-}\ [_{\text{Infl}}\ \text{-}e]]$ $[_{n + \text{N}}\ cosa]$

If *che* and *che cosa* instantiate (related but) distinct structures, as necessitated by our approach, then there must also be an interpretive difference between the two. While more research is required, this prediction appears to be borne out. In (16) we have provided two ways to ask 'What did you do?', one with *che* (16a) and one with *che cosa* (16b).

(16) a. **Che** hai fatto?
      b. **Che cosa** hai fatto?
          'What did you do?'

Interestingly, (16b) with *che cosa* is odd in an out-of-the-blue context, whereas *che* (16a) is fine in this context. More specifically, the option with *che cosa* presupposes that something has necessarily been done (Ciro Greco, p.c.), along the lines of 'What is/are the thing(s) that you have done?'. This supports our general line of reasoning, according to which *che* and *che cosa* must be (at least slightly) different, structurally speaking. A nice consequence of being forced to posit a double structure for *che cosa* with a semi-lexical N, then, is that this may start to explain the presuppositional reading which is present in (16b) but absent in (16a).

5 Concluding remarks

In this paper we have endeavored to show that Dem, Comp, Rel, Wh, and Indet elements not only participate in syncretism patterns but also have a tripartite structure. By looking in more detail at the internal structure of these elements, we have shown that they are composed of an F domain that can grow and shrink in the typical nanosyntactic subset-superset manner, an invariant base (*n*, also called *nominal core*), and an invariant inflection.

Based on Baunaz & Lander (2017a), we have claimed that languages may lexicalize some or all the parts of this fseq in different ways: if the Base *n* is compulsory in all languages, in some languages the F domain can either be missing or form a portmanteau with the Base *n*, or it can realize a morpheme on its own. In languages where Infl is realized, it is almost always frozen and suffixed to the compulsory Base *n*. Base *n* and Infl can also form a portmanteau constituent to the exclusion of F. Overall, the Base *n* appears to be the fundamental structural building block used in the construction of Dem, Comp, Rel, Wh, and Indet elements.

We have one final remark. Since Base and (we have assumed – cf. for example our brief discussion of Yiddish /-s/ above) Inflection are invariant, it turns out that what we have been tracking in our syncretism data is in fact F, which often appears as a functional prefix in many of the languages we have considered here. This is an interesting result, especially since our hierarchy parallels findings from more traditional cartographic work on the clausal spine (e.g. **D** > **C** > **Rel** in Cinque 2008; Force > Int > Foc (i.e. **C**-domain) > **Wh** in Rizzi 2001). This parallelism suggests that the word-internal or morphological structure we are interested in is replicated at the higher clausal level. In fact, it would appear that the bigger the F-structure is, the higher the entire complex of $[\text{F} + n + \Phi]$ ends up

being merged (cf. De Clercq & Vanden Wyngaerd 2016; 2017 on merging nano-structures in the clausal spine). This tells us, furthermore, that syncretism really can help us map out "macro-syntax" above the word level (though one has to be careful to determine precisely which morphological ingredients are responsible for the syncretism patterns emerging).

Abbreviations

1 = first person, 3 = third person, *AIS = Atlante Italo Svizzero*, Am. = American, Bg. = Bulgarian, C(omp) = complementizer, *C* = consonant, Cz. = Czech, D = definiteness, Dem = demonstrative pronoun, DET = determiner, Du. = Dutch, ESlav = East Slavonic, F = functional (feature), Fr. = French, F(EM) = feminine gender, FACT = factive, Ger. = German, Hu. = Hungarian, Icel. = Icelandic, IND = indicative, Indet = indeterminate noun, Infl / Φ = inflectional material, It. = Italian, M(ASC) = masculine gender, MG = Modern Greek, *n* = nominal core, N = noun, NEUT = neuter gender, NUM = number, NGmc = North Germanic, P = preposition, PL = plural, Po. = Polish, PRO = pronoun/pronominal, Rel = relative pronoun, RESTR = restrictive, Ro. = Romanian, Rom = Romance, Ru. = Russian, RVZ = relativizer, SC = Serbo-Croatian, SG = singular, SSlav = South Slavonic, SUBJ = subjunctive, V = verb, WGmc = West Germanic, Wh = interrogative pronoun, WSlav = West Slavonic, Yid. = Yiddish

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Aboh, Enoch Oladé. 2005. Deriving relative and factive clauses. In Laura Brugè, Giuliana Giusti, Nicola Munaro, Walter Schweikert & Giuseppina Turano (eds.), *Contributions to the Thirtieth Incontro di Grammatica Generativa, Venice*, February 26–28, 2004, 265–285. Venice: Università Ca'Foscari.

Bağrıaçık, Metin & Aslı Göksel. 2016. Greek and Turkish influences in the clausal complements of Cunda Turkish. In Mine Güven, Didar Akar, Balkız Öztürk & Meltem Kelepir (eds.), *Exploring the Turkish linguistic landscape: Essays in honor of Eser Erguvanlı-Taylan*, 57–80. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/slcs.175.05bag

Baldissera, Valeria. 2013. *The Griko dialect of Salento: Balkan features and linguistic contact*. Venice: Ca' Foscari University of Venice dissertation.

Baunaz, Lena. 2015. On the various sizes of complementizers. *Probus* 27(2). 193–236. DOI: https://doi.org/10.1515/probus-2014-0001

Baunaz, Lena. 2016. Deconstructing complementizers in Serbo-Croatian, Modern Greek and Bulgarian. In Christopher Hammerly & Brandon Prickett (eds.), *Proceedings of NELS 46* 1. 69–77.

Baunaz, Lena. In press. Decomposing complementizers: The fseq of French, Modern Greek, Serbo-Croatian and Bulgarian complementizers. In Lena Baunaz, Karen De

Clercq, Liliane Haegeman & Eric Lander (eds.), *Exploring nanosyntax* (Oxford Studies in Comparative Syntax). New York: Oxford University Press.

Baunaz, Lena & Eric Lander. 2017a. Syncretisms with the nominal complementizer. *Studia Linguistica*. Online ISSN: 1467-958. DOI: https://doi.org/10.1111/stul.12077

Baunaz, Lena & Eric Lander. 2017b. The internal structure of ontological categories. Paper presented at the *Incontro di Grammatica Generative 43*, 16–17 February. Pavia: IUSS Pavia.

Baunaz, Lena & Eric Lander. In press. Cross-categorial syncretism and the Slavic containment puzzle. In Iliana Krapova & Brian Joseph (eds.), *Balkan syntax and (universal) principles of grammar* (provisional title). Berlin: Mouton de Gruyter.

Benţea, Anamaria. 2010. On restrictive relatives in Romanian: Towards a head-raising analysis. *Generative Grammar in Geneva* 6. 165–190.

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

Bobaljik, Jonathan. 2012. *Universals in comparative morphology. Suppletion, superlatives, and the structure of words*. Cambridge, MA: MIT Press.

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

Calabrese Andrea. 1993. The sentential complementation of Salentino: A study of a language without infinitival clauses. In Adriana Belletti (ed.), *Syntactic theory and the dialects of Italy*, 28–98. Turin: Rosenberg & Sellier.

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

Cysouw, Michael. 2004. Interrogative words: An exercise in lexical typology. Handout.

Déchaine, Rose-Marie & Martina Wiltschko. 2002. Decomposing pronouns. *Linguistic Inquiry* 33. 409–443. DOI: https://doi.org/10.1162/002438902760168554

De Clercq, Karen & Guido Vanden Wyngaerd. 2016. Adjectives and the ban on double negation. Paper presented at *SinFonIJA 9, 15–17 September*. Brno, Czech Republic.

De Clercq, Karen & Guido Vanden Wyngaerd. 2017. Why affixal negation is syntactic. In Aaron Kaplan, Abby Kaplan, Miranda McCarvel & Edward Rubin (eds.), *Proceedings of WCCFL 34*, 151–158. Sommerville, MA: Cascadilla Press.

Diessel, Holger. 2003. The relationship between demonstratives and interrogatives. *Studies in Language* 27(3). 635–655. DOI: https://doi.org/10.1075/sl.27.3.06die

Ferraresi, Gisella. 1997. *Word order and phrase structure in Gothic*. Stuttgart: University of Stuttgart dissertation.

Ferraresi, Gisella. 2005. *Word order and phrase structure in Gothic*. Leuven: Peeters.

Giannakidou, Anastasia. 2009. The dependency of the subjunctive revisited: Temporal semantics and polarity. *Lingua* 119. 1883–1908. DOI: https://doi.org/10.1016/j.lingua.2008.11.007

Grosu, Alexander. 1994. *Three studies in locality and case*. London: Routledge. DOI: https://doi.org/10.4324/9780203427132

Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.

Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511613463

Jacobs, Neil G. 2005. *Yiddish: A linguistic introduction*. Cambridge: Cambridge University Press.

Kayne, Richard. 2005. *Movement and silence*. New York: Oxford Uuniversity Press. DOI: https://doi.org/10.1093/acprof:oso/9780195179163.001.0001

Kayne, Richard. 2008. Why isn't *this* a complementizer? Ms. New York University.

Kayne, Richard & Jean-Yves Pollock. 2010. Notes on French and English demonstratives. In Jan-Wouter Zwart & Mark de Vries (eds.), *Structure preserved: Studies in syntax for Jan Koster*, 215–228. Amsterdam: John Benjamins.

Keevallik, Leelo. 2008. Conjunction and sequenced actions: The Estonian complementizer and evidential particle *et*. In Ritva Laury (ed.), *Crosslinguistic studies of clause combining: The multifunctionality of conjunctions*, 125–152. Amsterdam: John Benjamins.

Kiparsky, Paul. 1995. Indo-European origins of Germanic syntax. In Adrian Battye & Ian Roberts (eds.), *Clause structure and language change*, 140–169. Oxford: Oxford University Press.

Kishimoto, Hideki. 2000. Indefinite pronouns and overt N-raising. *Linguistic Inquiry* 31. 557–566. DOI: https://doi.org/10.1162/002438900554451

Kornfilt, Jaklin. 2008. Agreement: Subject case correlations in Turkish and beyond. Ms. Leipzig Spring School on Linguistic Diversity: Topics in Turkic Syntax.

Krapova, Iliana. 1998. Subjunctive complements, null subjects and case checking in Bulgarian. *University of Venice Working Papers in Linguistics* 8(2). 73–93.

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

Lander, Eric & Liliane Haegeman. 2016. The nanosyntax of spatial deixis. *Studia Linguistica*, 1–66. DOI: https://doi.org/10.1111/stul.12061

Ledgeway, Adam. 2013. Testing linguistic theory and variation to their limits: The case of Romance and its dialects. *Corpus* 12. 271–327.

Ledgeway, Adam. 2015. Reconstructing complementiser-drop in the dialects of the Salento: A syntactic or phonological phenomenon? In Theresa Biberauer & George Walkden (eds.), *Syntax over time: Lexical, morphological, and information-structural interactions*, 146–62. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199687923.003.0009

Le Goffic, Pierre. 2008. Que complétif en français: Essai d'analyse. *Langue française* 158. 53–68.

Leu, Tom. 2008. *The internal syntax of determiners* (GAGL 47). New York, NY: University of New York dissertation.

Leu, Tom. 2015. *The architecture of determiners*. New York: Oxford University Press.

Leu, Tom. 2016. Pronominal ingredients of indefinite pronouns. Talk given at the *9th Days of Swiss Linguistics*, June 2016. Geneva: University of Geneva.

Longobardi, Giuseppe. 1991. Alcune riflessioni informali sulla posizione del verbo in gotico e le prospettive di una sintassi comparata dei complementatori generici. Ms. University of Venice.

Manzini, Maria Rita & Leonardo Maria Savoia. 2003. The nature of complementizers. *Rivista di Grammatica Generativa* 28. 87–110.

Manzini, Maria Rita & Leonardo Maria Savoia. 2011. *Grammatical categories*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511974489

Meyer-Lübke, Wilhelm. 1890. *Grammatik der Romanischen Sprachen* I. Lautlehre. Leipzig: Fues's Verlag (R. Reisland).

Meyer-Lübke, Wilhelm. 1899. *Grammatik der Romanischen Sprachen* III. Syntax. Leipzig: O.R. Reisland.

Munaro, Nicola. 2001. Free relatives as defective Wh-elements: Evidence from the North-Western Italian dialects. In Yves D'hulst, Johan Rooryck & Jan Schroten (eds.), *Romance languages and linguistic theory 1999: Selected papers from 'Going Romance' 1999, Leiden*, 9–11 December, 281–306. Amsterdam/Philadelphia: John Benjamins.

Rizzi, Luigi. 2001. On the position "Int(errogative)" in the left periphery of the clause. In Guglielmo Cinque & Giampaolo Salvi (eds.), *Current studies in Italian syntax*: *Essays offered to Lorenzo Renzi*, 287–296. Amsterdam: Elsevier.

Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486326

Roehrs, Dorian. 2010. Demonstrative-reinforcer constructions. *The Journal of Comparative Germanic Linguistics* 13(3). 225–268. DOI: https://doi.org/10.1007/s10828-010-9038-4

Roussou, Anna. 2009. In the mood for control. *Lingua* 119. 1811–1836. DOI: https://doi.org/10.1016/j.lingua.2008.11.010

Roussou, Anna. 2010. Selecting complementizers. *Lingua* 120(3). 582–603. DOI: https://doi.org/10.1016/j.lingua.2008.08.006

Savu, Carmen & Sebastian Bican-Miclescu. 2012. Romanian demonstratives. Ms. nanosyntax weblab "The ingredients of demonstratives" convened by Michal Starke.

Sočanac, Tomislav. 2017. *Subjunctive complements in Slavic languages: A syntax-semantics interface approach*. Geneva: University of Geneva dissertation.

Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. In Peter Svenonius, Gillian Ramchand, Michal Starke & Tarald Taraldsen (eds.), *Nordlyd: Special issue on nanosyntax* 36. 1–6.

Starke, Michal. 2014. Towards elegant parameters: Language variation reduces to the size of lexically-stored trees. In M. Carme Picallo (ed.), *Linguistic variation in the minimalist framework*, 140–152. Oxford: Oxford University Press.

Szabolcsi, Anna, James Doh Whang & Vera Zu. 2014. Quantifier words and their multifunctional(?) parts. *Language and Linguistics* 15(1). 115–155. DOI: https://doi.org/10.1177/1606822X13506660

Vanden Wyngaerd, Guido. In press. The feature structure of pronouns: A probe into multidimensional paradigms. In Lena Baunaz, Karen De Clercq, Liliane Haegeman & Eric Lander (eds.), *Exploring nanosyntax*. New York: Oxford University Press.