STUDIA LINGUISTICA
A JOURNAL OF GENERAL LINGUISTICS

SYNCRETISMS WITH THE NOMINAL COMPLEMENTIZER*

Lena Baunaz & Eric Lander

Abstract. Nominal complementizers (e.g. Eng. *that*, Fr. *que*) often have the same morphophonological form as other grammatical items, such as demonstrative, relative, and wh-pronouns. In this paper we treat such overlaps as instantiations of syncretism, and we discuss the different patterns of syncretism with the nominal complementizer in various languages. We treat the syncretism facts within a nanosyntactic framework (Starke 2009, Caha 2009), meaning that complementizers are not simplex heads of CP (or ForceP/FinP in Rizzi's 1997 sense) but actually composed of multiple features, each feature corresponding to a head in a single functional sequence which is responsible for building demonstratives, complementizers, relativizers, and wh-pronouns (for alternative decompositions of complementizers in Romance, Balkan, and Germanic, see also Baunaz 2015, 2016, to appear Sanfelice & Poletto 2014; and Leu 2015, respectively). Interestingly, moreover, many of the languages under discussion show a bound morpheme appearing as an integral part of the internal morphological makeup of quantifiers. This bound morpheme may also be syncretic with the complementizer (Romance *-que*/*-che*, Serbo-Croatian *-što*, Modern Greek *-ti*, Finnish *-kin*, and Hungarian *ho-*) or not (Germanic *-thing*/*-ting*, for which see also Leu 2005). We call this the 'nominal core' (*n*), and its behavior with regard to syncretism is crucial for determining the hierarchical ordering of the functional sequence.

1. Introduction

Since the 1960s it has been assumed that there is a single syntactic category called C(omp) for *complementizer*. Comps are subordinators which turn clauses into complements (Rosenbaum 1967, Lakoff 1968, Bresnan 1970, 1972, Kiparsky & Kiparsky 1970, Grimshaw 1979, 1991, Lasnik & Saito 1992, Dixon 2006, among many others). Certain

*We would like to thank Alexandra Ioannidou and Metin Bagrıaçık for their judgements on Modern Greek, Genoveva Puskás for her help with the Hungarian data, and Karoliina Lohiniva for her help on the Finnish data. We also thank Boban Arsenijević, Tanja Samardžic and Tomislav Sočanac for their help on the Serbo-Crotian data. We are also very grateful to Liliane Haegeman, Tom Leu, Anna Roussou and an anonymous *Studia Linguistica reviewer* for very interesting comments on earlier versions of this research. We also thank the audience of *Going Romance* 28 (Lisboa, 4-6 December 2014), as well as the audience of the *Variation in C workshop*, Università Ca' Foscari Venezia, October 21- 22, 2014. All errors are our own. Lena Baunaz's research has been supported by the Swiss National Foundation (grant: PA00P1_145313) and FWO project 2009-Odysseus-Haegeman-G091409. Eric Lander's research has been supported by BOF grant 01D30311, FWO project 2009'Odysseus'Haegeman-G091409, and a postdoctoral grant from the University of Gothenburg.

Studia Linguistica ••(•) 2017, pp. 1–34.
Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA

predicates (i.e. V, N, A) select certain kinds of Comp, and Comps in their own right encode certain properties of the complement clauses they introduce (i.e. finiteness and illocutionary force; cf. Rizzi 1997). A great deal of work has been done on the nature of the CP system and its role in language, from selection to clause-typing to extraction phenomena.

In this paper we look at Comp from a different angle, namely in terms of cross-category syncretism. That is, the declarative Comp morpheme in many languages overlaps in morphophonological form with demonstrative pronouns (Dem), relative pronouns (Rel), and interrogative pronouns (Wh) (see below for references). Rather than analyzing these systematic overlaps in form in terms of (coincidental) homophony, we take them to be cases of syncretism, defined as "a surface conflation of two distinct morphosyntactic structures" (Caha 2009:6). In other words, a syncretism is when a single morphophonological form applies in more than one morphosyntactic environment or structure. This will be discussed in detail below.

When Comp is syncretic with (pro)nominal elements, we call this *nominal Comp*. When Comp is syncretic with verbal elements (e.g. 'say', a phenomenon prevalent in Niger-Congo, Sinitic, and Austronesian languages), we call this *verbal Comp*. Also possible is what might be called *prepositional Comp*, when Comp is syncretic with prepositional/case elements (e.g. in Hebrew and Kanuari, for which see Noonan 2007; or Fr. *de* 'to', *à* 'to', *pour* 'for', Eng. *for*, etc.). That is, at least three (and perhaps more) separate inventories of functional features – nominal, verbal, prepositional – can be distinguished on the basis of different categorial syncretisms with Comp crosslinguistically. Though we save discussion of the non-nominal Comp types for another occasion, it is useful to point out that languages do not necessarily choose one or the other species of Comp. English, for instance, has both nominal Comp (1a) and verbal Comp (1b) (see Rooryck 2000, Brook 2014).

(1) a. I thought **that** the computer would be broken forever.
    b. It seemed **like** the computer would be broken forever.

The Comp *like* can be considered syncretic with the comparative verbal construction *be like*, making it a Comp of the verbal species. To take another example, in Ik [Nilo-Saharan] the temporal subordinator 'when, if' is closely related in form to the Dem and Rel pronouns, instantiating a nominal syncretism pattern, but this language also has the verbal Comp and quotative *toimɛna-* which has its origins in the report verb 'say' (Schrock 2014:450, 535). Similarly, French has both nominal Comp (*que*) and prepositional Comp (*de* 'to', *à* 'to', *pour* 'for') available. In other words, it is not the case that a language must 'choose' either the nominal or verbal

features. Ultimately we would like to suggest that all of the feature inventories (nominal, verbal, prepositional) are in fact available in every language.[1] In this paper, though, only the nominal one will be considered.

The paper is structured as follows. In section 2 we present the core data, namely the attested patterns of syncretism with declarative Comp in a number of Indo-European and Finno-Ugric languages. We will show that there is only a single linear ordering which can account for the syncretism patterns in a neat way, specifically in terms of adjacency (that is, only adjacent layers can be syncretic). In section 3, making use of nanosyntactic logic (e.g. Starke 2009, 2011, 2013, Caha 2009, 2010, 2013, Taraldsen 2009, Pantcheva 2011, Rocquet 2013, De Clercq 2013, Vangsnes 2013, 2014), we will show that these features should be understood as unary and additive, making up a functional sequence (fseq). The structures built using this fseq are in superset-subset relations with one another, and the syncretism patterns observed will be shown to follow from nanosyntactic principles of spellout. However, we will see at this point that the merge order of the features in the fseq is not perfectly clear. That is, we do not yet know which feature is the lowest in the fseq (i.e. the first to be merged) and which feature is the highest in the fseq (i.e. the last to be merged). In section 4 we identify a novel strategy for locating the 'smaller' (or 'lower') end of our hierarchy on the basis of syncretism. The strategy builds on Cardinaletti & Starke (1999), who show that clitic pronouns are structurally impoverished compared to weak pronouns (which in turn are structurally impoverished compared to strong pronouns). The ultimate result is the identification of what we call the *nominal core*, which turns out to be a crucial ingredient in the internal morphological structure of other elements as well (most importantly quantifiers). Section 5 concludes the paper.

2. Empirical background

2.1. *The core data: Syncretism patterns*

It is well known that complementizers (Comp) in various Indo-European languages may share the same morphophonological form as demonstrative (Dem), relative (Rel), and interrogative (Wh) pronouns (see Sportiche 2011, Le Goffic 2008 for French; Manzini & Savoia 2003, 2011, among others, for Italian; see Roberts & Roussou 2003, Kayne

[1] However, these fseqs will be lexicalized quite differently crosslinguistically, so in some languages it may not be obvious how a particular fseq is morphologically realized. We are pursuing this in ongoing research, but for now we consider languages like English and French to be clear evidence that different types of Comp coexist within the same language.

Table 1. Syncretism patterns crosslinguistically (neuter/inanimate 3SG forms)

		DEM	COMP	REL	WH
North Gmc	*Swedish*	det	att	som	vad
	Danish	det	at	som	hvad
West Gmc	*English*			which	which
				what (FREE REL)	what
		that	that	that	what
				% as	
	Dutch	dat	dat	dat	wat
	German	das	dass	das	was
	Yiddish	jenc	vos	vos	vos
			az	az	vos
Romance	*French*	ce	que	que	que
	Italian	quello	che	che	che
	Spanish	aquél	que	que	qué
Balkan	*Modern Greek*	ekíno	o-ti	ó-ti (FREE REL)	tí
		ekeí 'there'	pu	(ó-)pu	pú 'where'
	Serbo-Croatian	to	što	što	što
			da		
Finno-Ugric	*Finnish*	tä- 'this'	että	mi-	mi-
	Hungarian	így 'in this manner'	hogy	a-hogy	hogy(-an) 'how'

2008, Leu 2008, 2015 for English and West Germanic in general; see Roussou 2010 for Modern Greek). In this paper we consider data from North Germanic (Swedish, Danish), West Germanic (English, Dutch, German, Yiddish), Romance (French, Italian, Spanish), Balkan (Serbo-Croatian and Modern Greek), and the non-Indo-European languages Finnish and Hungarian. The relevant forms (Dem, Comp, Rel, and Wh) from these languages are gathered and arranged in Table 1. As mentioned above, we consider these overlaps in morphophonological form to be instantiations of syncretism (see below for more details). Syncretic items are shaded in Table 1 (where dark and light gray are used only for visual convenience, i.e. to distinguish the different syncretic patterns). Throughout the table we provide neuter/inanimate third person singular forms; for more discussion on this point see the end of section 2.2.

(i) North Germanic

In North Germanic or Scandinavian, here represented by Swedish and Danish, there is no syncretism between Dem, Comp, Rel, or Wh. Each

layer is lexicalized differently (see Vangsnes 2013, 2014 for more fine-grained layers).[2]

(ii) West Germanic

In all of the West Germanic languages we consider, Comp is syncretic with Rel. In English, Dutch, and German, moreover, Comp is syncretic with distal Dem (Eng. *that,* Du. *dat,* Ger. *das(s)*) but not syncretic with Wh (Eng. *what*, Du. *wat*, Ger. *was*). In Yiddish there are two Comps, *vos* and *az*. *Vos* appears under factive predicates and is syncretic with both Rel and Wh. Elsewhere *az* appears, which can also appear as Rel but not as Wh.

We have also pointed out in Table 1 that English shows a Rel/Wh syncretism in the items *which* and *what*.

(2) a. I read the book **which** you gave me. Rel
b. **Which** book did you donate to the library? Wh

(3) a. You gave me **what** turned out to be a really bad cold. Free Rel
b. **What** did you donate to the library? Wh

This Rel/Wh syncretism is also available in formal Dutch (*de*) *welk*-, German *welch*-, and even Scandinavian, e.g. Sw. *vilk*-.

Finally, for the purposes of illustrating as many of the possible syncretism patterns as possible, we note that certain dialects of English use *as* as a relativizer.

(4) The things **as** I was saying
(Kayne 2008:23, his (203))

In other words, these dialects of English have only a Dem/Comp syncretism, rather than a Dem/Comp/Rel syncretism.

(iii) Romance

The Romance languages considered here are French, Italian, and Spanish. In these languages, the Comp, Rel, and Wh items are all syncretic with each other, but they are not syncretic with Dem. In French, for example, Comp *que* is syncretic with weak Wh *que* and weak Rel *que* (see Sportiche 2011) but not with Dem *ce*. Italian *che* and

[2] We are glossing over various details in Scandinavian that are not directly relevant to our concerns in this paper, for instance the fact that *som* has a variety of different uses (e.g. Sw. *som* can mean 'like, as', along the lines of Eng. *as*), and also the fact that Mainland Scandinavian finite complementizers overlap in form with infinitival markers/complementizers (Sw. *att*, Da. *at*, Icel. *að* 'that, to'). Regarding the latter issue, there is also an unambiguous infinitival marker – Swedish and Norwegian /ɔ/, written <å> in Norwegian; Da. /a/ – which differs from complementizer /at/ and also happens to be homophonous with the reduced form of the conjunction *och*, *og* 'and' (/ɔk/ > /ɔ/ in Swedish/Norwegian). Among others, see Christensen (1983), Platzack (1986), Wiklund (2008). Thanks to a reviewer for highlighting these facts.

Spanish *que* also show a Comp/Rel/Wh syncretism, with non-syncretic Dem (It. *quello*, Sp. *aquél*).[3]

(iv) Balkan

Modern Greek shows a syncretism pattern which is quite similar to Romance. The Comp for non-factive predicates (e.g. *leo* 'say') in Greek is *oti*, which is also syncretic with (Free) Rel *o,ti* 'whatever' and Wh *tí* 'what'.[4] The Comp for epistemic factive (e.g. *thimame* 'remember') and emotive factive (e.g. *lipame* 'regret') predicates is *pu*, which is syncretic with Rel (*ó*-)*pu* 'where/that'[5] and the Wh locative adverb *pu* 'where' (cf. Hungarian; see section 4.5 for more discussion of these facts in Greek). As in Romance, Dem (Gk. *ekíno* 'that' for *oti* and adverbial *ekeí* 'there' for *pu*) does not participate in the syncretism.

In Serbo-Croatian a similar syncretism pattern is found. The Comp used under emotive factive verbs (e.g. *žaliti* 'regret') in Serbo-Croatian is *što*, which is also used as both Rel and Wh. The particle *da* is used to introduce complements of all other types of predicates (epistemic factive verbs, e.g. *sjetiti se* 'remember'; non-factive verbs, e.g. *reći* 'say'; desiderative verbs, e.g. *željeti* 'want') and it is syncretic with the yes-no polarity item *da*, which is apparently derived from the verb 'give' (although this claim is still debated in the literature).[6] Since this paper is concerned with nominal complementizers (i.e. complementizers which are syncretic with nominal items) and not verbal complementizers or mood particles, we will not discuss *da* here, only *što*. See Baunaz (2016, to appear) for more discussion.

(v) Finno-Ugric

In Finnish the stem *mi-* can be used for both Wh and Rel. For the Wh paradigm, *mi-* is the inanimate stem while *ke-* is the animate stem. For the Rel paradigm, *jo-* is normally used for animate antecedents while *mi-* is, again, normally for inanimate referents and is thus the one we cite here (see below). Importantly, *mi-* is often used as a Rel pronoun "when the

[3] Historically, Old French Dem *cil* and *cist* (both later falling together as *ce*; see Carlier & Mulder 2006) and Italian *quello* and *questo* come from the Latin combination *ecce ille*/*ecce iste*, with Spanish *aquél* from the variant *accu ille* (Adams 2013:465-466, 469). In other words, the history of Dem has nothing to do with Wh or Rel pronouns. In Germanic the historical relationship is much closer. In West Germanic the usage of Dem *that* as Comp and Rel is quite obvious (see Roberts & Roussou 2003: §3.4). In fact, even Scandinavian *at(t)* is derived from the loss of the initial dental in Dem **þat*, giving *at* in the Scandinavian languages (Haugen 1976:160).

[4] Desiderative verbs in Greek do not embed the complementizer *oti*. Instead the subjunctive mood particle *na* appears in the complement clauses of desideratives. See Giannakidou (2009).

[5] Where *o-* is a prefix and has its origins in the definite article, as is also the case in *o-ti* (see Roussou 2010 and references cited there; see also section 2.2 and Baunaz & Lander *to appear* a, *under review*).

[6] In fact, *da* in this case is not a complementizer but actually a subjunctive mood particle (cf. Gk. *na* mentioned above). See Sočanac (2011).

reference is to a clause" (Karlsson 1999:150), suggesting that *mi-* belongs in the same paradigmatic system as Comp.

(5) Tuli sade, *mikä* esti matkamme.
'It rained, *which* prevented our trip.' (Karlsson 1999:151)

We consider the possibility of using *mi-* in both functions, Wh and Rel, as a case of syncretism. This syncretism does not extend to Comp *että*, however. Finnish *että* has a somewhat complex history. The *e* component is related to the Uralic proximal Dem (cf. Hungarian *ez* 'this'), while the *-ttä* component is taken to be a modal ending, giving us an original meaning of 'in this way, so' (Keevallik 2008:141) (cf. *hogy* in Hungarian below). Nevertheless, the fact that the history of Finnish *että* is linked to the demonstrative **e-* (*ez* in Hungarian) gives us good reason to say that the Finnish Comp in Table 1 is indeed nominal. The proximal Dem in Finnish, finally, is not a development of the Uralic *e*-root, but is instead *tä-*.

Finally, Hungarian Comp *hogy* is syncretic with both Rel *a-hogy* (where the prefix *a-* has its origins in Dem, i.e. *az hogy*, cf. also Gk. *ó-ti* and *ó-pu* above; see section 2.2 and fn.7 for more details) and the Wh manner adverbial *hogy(-an)* 'how, in which manner', where the suffix *-an* is found in formal registers only, suggesting that it is in the process of being lost. Since Comp *hogy* is syncretic with a Wh adverb, we provide the adverb *így* 'in this manner' for the Dem slot in Table 1. We discuss these facts from Hungarian in more depth in section 4.5.

2.2. *Some methodological assumptions*

As seen above, Comp often adopts the same morphophonological form as pronouns (Dem, Rel, Wh). The phenomenon is found in many Indo-European languages (Romance, West Germanic, Balkan languages), but also in non-Indo-European languages like Finnish and Hungarian. Though more languages can and should be added to the empirical inventory, the prevalence of the phenomenon suggests that the homophony is not accidental and that a unified analysis would be desirable. In this paper, we take the morphophonological overlap between Comp and Dem, Rel, and Wh to be cases of syncretism revealing an underlying hierarchy of features, with more details in the following sections.

At this point we would like to anticipate and dispel some worries about morphological decomposition and the nature of the fine-grained grammatical features at stake in Table 1 and the rest of this paper. First, it is rather obvious that many of the forms provided in Table 1 are not monomorphemic. English *that*, for instance, can be decomposed into smaller morphological parts, i.e. the definite determiner *th(e)* plus the distal *-at* (Chomsky 1995:338, Déchaine & Wiltschko 2002, Leu 2008,

Kayne & Pollock 2010). Indeed, a similar segmentation could be proposed for the closely related German *d-as(s)* and Dutch *d-at*. Similarly, Italian *quello* could be divided into a stem *que-* plus a marker of distality *-l* (compare proximal *que-**st**-o*), plus the definite determiner *-lo*. It is crucial to notice that for our purposes it is not necessary to descend to this level of granularity. Whether or not we decompose Germanic *that*/*dat*/*dass* into *th-at*/*d-at*/*d-ass* or Italian *quello* into *que-l-lo*, the patterns of (non-)syncretism in Table 1 remain the same, and thus our reasoning and analysis on the basis of these patterns are not affected. We certainly agree that more fine-grained decompositions can and should be performed (see Baunaz & Lander *under review*), and also that more than four functional layers are ultimately needed (for instance, there is almost certainly a separate functional layer responsible for free relatives, probably between Rel and Wh), but for the purposes of this paper, where broad classes of syncretism are the main concern, these complications can be abstracted away from.

To take a more concrete case, consider that Hungarian Rel *a-hogy* shows a prefix *a-* which is demonstrative in origin (< *az hogy*[7]), i.e. *a-hogy* is (at least) bimorphemic. At first sight, we can see here a containment relation between Rel and Wh, in that Rel *a-hogy* contains Wh *hogy*. In that sense, it looks like Rel/Wh Fr. ***le**-quel* or It. ***il** quale*, or even Gk. ***ó**-ti*/***ó**-pu*. We suggest that these D markers in Romance, Greek, and Hungarian (Fr. *le-,* It. *il-,* Gk. *ó-,* Hu. *a-*) are agreement morphemes of some sort, which are prefixed to Fr. *quel*, It. *quale*, Gk. *pu*/*ti*, and Hu. *hogy* at some later stage of the syntactic derivation. That is to say, these markers enter the derivation when a relative clause or question is being derived at the *sentential level* (cf. Kayne's 1994 theory of relative clauses wherein D selects CP); in this paper, however, we are concerned with the *word-internal level* (i.e. with the internal featural makeup of words like Eng. *that*, Fr. *que*, etc.), so these markers are irrelevant for our purposes.

Another issue, one which is closely related to decomposition, is the question of the proper place of case and phi features. Many forms in Table 1 involve a specific set of case and phi features, for instance the German complementizer *dass* is syncretic with Dem/Rel N.NOM/ACC.SG *das* (and not with, say, M.NOM.SG *der* or F.NOM.SG *die*). We take this to be a blunt fact about the data. Therefore, in order to ensure that we are comparing apples with apples, we provide only N.NOM/ACC.SG forms throughout (e.g. Ger. *das – dass – das – **was***). Again, we are abstracting away from the finer details of case and phi, in the same way that we abstract away from finer-

[7] Bacskai-Atkari (2016, fn.13) reports that the prefix *a-* is originally "a demonstrative pronoun in the matrix clause (*az* 'that')." Its role was, she claims, to syntactically mark the embedded clause. *Az* may appear on its own or with a lexical head. When *az* and the relative pronoun are adjacent (as in *az ki* 'that who', Bacskai-Atari's example), the matrix pronoun is cliticized onto the relative pronoun, and then eventually reanalyzed as belonging to it: *az ki* > *azki* > *akki* > *aki* 'who'.

grained decompositions like Eng. *th-at* or It. *que-l-lo*. As long as we control for the consistent presence of N.NOM/ACC.SG features throughout the data (the reason we provide Yid. N.NOM/ACC.SG *jenc*, Gk. N.NOM/ACC.SG *ékino*, etc.), we will be able to isolate what we are actually interested in identifying on the basis of syncretism, namely the core features which are responsible for building Dem, Comp, Rel, and Wh. As mentioned above, more featural layers are certainly needed beyond these four, and case and phi features will certainly be among them. See Baunaz & Lander (*to appear* a and *under review*) for a more decompositional approach.

3. The linear order of Dem, Comp, Rel, and Wh

3.1. *The adjacency effect*

Let us return to the facts presented in section 1.1. Observe that syncretism in Table 1 targets only adjacent cells. It may be charged – at least at first glance – that this is simply a convenient and neat way to arrange the data. However, as observed by Bobaljik (2007, 2012), Starke (unpublished work, e.g. 2011, 2013), and Caha (2009), among others, there is more to this arrangement than meets the eye. In fact, if enough of the possible patterns are attested and taken into account, then only one single linear ordering of the items involved will suffice to capture syncretism in terms of adjacent cells. To be more concrete, consider the patterns in Table 2 (taken from Table 1 above).

Table 2. Three crucial syncretism patterns from Table 1

	DEM	**COMP**	**REL**	**WH**
non-standard English	that	that	as	what
Yiddish	jenc	az	az	vos
Finnish	tä-	että	mi-	mi-

The non-standard English data demand that Dem and Comp be adjacent, the Yiddish data that Comp and Rel be adjacent, and the Finnish data that Rel and Wh be adjacent. The only linear ordering of Dem, Comp, Rel, and Wh which can capture these adjacencies is Dem | Comp | Rel | Wh. If the ordering had been Dem | Rel | Comp | Wh, for instance, then the English data would not be captured, since Dem and Comp are syncretic in this language. If the ordering had been Dem | Comp | Wh | Rel, then the Yiddish data would not be captured, since Comp and Rel are syncretic in this language, and so on. In fact, we will see below how the adjacency effect is actually a reflection of structural

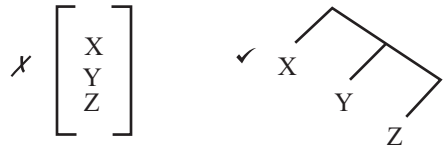
adjacency, meaning that syncretism is a useful tool for probing the internal featural makeup of morphological items.[8]

3.2. *The nanosyntactic theory of syncretism*

Nanosyntax (e.g. Starke 2009, 2011, 2013; Caha 2009, 2010, 2013; Taraldsen 2009; Pantcheva 2011; Rocquet 2013; De Clercq 2013; Vangsnes 2013, 2014; see Baunaz & Lander *to appear* b for an introduction) has developed a successful theory of syncretism which will be useful for us in this paper. Nanosyntax is a direct descendant of cartography and takes a very fine-grained approach to language structure, dealing primarily (at this stage in its development) with word-internal structure. Importantly, the theory makes the case that the same principles which have been identifed for syntax are also operative at the morphological level. That is, there is no principled difference between syntax and morphology.

The nanosyntactic approach is fundamentally based on the reasoning that the general trend of the proliferation of syntactic projections and the atomization of heads – i.e. the view that a single syntactico-semantic feature should correspond to a single, unary head (the "one morphosyntactic property – one feature – one head" maxim; Cinque & Rizzi 2008:50) – have an effect on the architecture and principles of grammar. Since features are heads in the functional spine, nanosyntax assumes that features are never found in unordered bundles (6a) but always arranged in a hierarchically ordered sequence, that is, a tree (6b) (see Dékány 2009:51).

(6) a. Unordered bundle b. Ordered sequence



The highly fine-grained nature of these features/heads, moreover, has important consequences. Morphemes are known to correspond to multiple features at once (e.g. many verbal endings encode both tense

[8] Various grammaticalization paths documented by Heine & Kuteva (2002) provide some supporting evidence for the adjacency idea: *demonstrative* > *complementizer* (Heine & Kuteva 2002:106-107), *w-question* (Wh-word) > (free relative) *complementizer* (Heine & Kuteva 2002:249-250), *w-question* (Wh-word) > *relative* (Heine & Kuteva 2002:251). For the path *demonstrative* > *relative* (Heine & Kuteva 2002:113-115), moreover, we would predict that a transitional stage involving the complementizer is passed through.

and aspect features), and since each feature in nanosyntax is a head, then morphemes must be able to spell out multiple heads at once. In order to allow for multiple heads to be targeted for lexicalization[9] simultaneously, nanosyntax allows for phrasal (XP) spellout. Thus, for instance, it is possible in nanosyntax to spell out the entire chunk $[_{XP}\ X\ [_{YP}\ Y\ [_{ZP}\ Z]]]$ in (6b) as a single phonological exponent. Phrasal spellout contrasts with the traditional approach to spellout, according to which only individual heads (X^0) can be targeted for spellout/morpheme insertion.

The nanosyntactic approach to syncretism and the adjacency effect is crucially based on the idea that features/heads are additive (or cumulative). Consider the tree in (6b) again. If the three features X, Y, and Z are added one by one, this means that there are three possible structures which the syntax can build, as seen in (7).[10]

(7) a. $[_{ZP}\ Z]$
b. $[_{YP}\ Y\ [_{ZP}\ Z]]$
c. $[_{XP}\ X\ [_{YP}\ Y\ [_{ZP}\ Z]]]$

The structure in (7a) is the smallest structure, and it is a subset of the larger structure in (7b), which in turn is a subset of the largest structure in (7c). Each one of these structures, moreover, could be associated with its own phonological exponent. Consider the three imaginary morphemes in (8) (where we define *morpheme* simply as a particular phonological form linked to a particular structure within a lexical entry).

(8) Fseq: X > Y > Z
a. $[_{ZP}\ Z]$ ⇔ *goo*
b. $[_{YP}\ Y\ [_{ZP}\ Z]]$ ⇔ *gah*
c. $[_{XP}\ X\ [_{YP}\ Y\ [_{ZP}\ Z]]]$ ⇔ *gur*

The morpheme *goo* is structurally a subset of *gah*, and *gah* is a subset of *gur*. Moreover, notice that while the morpheme *goo* corresponds only to a single functional layer, namely $[_{ZP}\ Z]$, *gah* and *gur* have more complex internal structures, being composed of multiple layers. This is licit in nanosyntax due to the theory's sanctioning of phrasal spellout.

Imagine now instead that we observe that both $[_{ZP}\ Z]$ and $[_{YP}\ Y\ [_{ZP}\ Z]]$ are lexicalized as *goo*, while $[_{XP}\ X\ [_{YP}\ Y\ [_{ZP}\ Z]]]$ is lexicalized as *gur*.

(9) a. $[_{ZP}\ Z]$ ⇔ ***goo***
b. $[_{YP}\ Y\ [_{ZP}\ Z]]$ ⇔ ***goo***
c. $[_{XP}\ X\ [_{YP}\ Y\ [_{ZP}\ Z]]]$ ⇔ *gur*

[9] In this paper we use *spellout* or *matching* as synonyms for *lexicalization*.

[10] The generative component determines how big the structure will be (i.e. at which layer the tree will stop being built). Structure-building is taken to proceed bottom-up and without skipping features.

In other words, there is a Y/Z syncretism vs. a distinct X form. To account for this pattern, we need to understand how nanosyntax views the process of spellout, namely the matching of syntactic structures (i.e. structures built by the syntax) by lexical structures (i.e. syntactic structures which are stored in the lexicon as part of a lexical entry for a morpheme). Henceforth we abbreviate *syntactic structure* as S(-tree) and *lexical structure* as L(-tree).

There are two crucial principles – the Superset Principle and the Elsewhere Principle – which regulate the process of spellout in nanosyntax (we skip a third principle, the Principle of Cyclic Override, because it is not necessary for our purposes here). The Superset Principle is defined in (10).

(10) The Superset Principle
L can match S if L is a superset (proper or not) of S.

In other words, an L-tree can match a given S-tree if the L-tree is bigger than or the same size as the S-tree. Consider now the lexical entries in (11), which reflect the pattern shown in (9) above.

(11) L1 $<[_{YP}\ Y\ [_{ZP}\ Z]] \Leftrightarrow$ *goo* >
L2 $<[_{XP}\ X\ [_{YP}\ Y\ [_{ZP}\ Z]]] \Leftrightarrow$ *gur* >

Consider now the S-tree in (12) and how it would be lexicalized given the lexical entries in (11).

(12) $[_{XP}\ X\ [_{YP}\ Y\ [_{ZP}\ Z]]] \Rightarrow$ *gur*

L1 corresponds to $[_{YP}\ Y\ [_{ZP}\ Z]]$ and as such is too small to lexicalize the S-tree $[_{XP}\ X\ [_{YP}\ Y\ [_{ZP}\ Z]]]$ by the Superset Principle since it lacks the head X. L2, however, is the same size as the S-tree $[_{XP}\ X\ [_{YP}\ Y\ [_{ZP}\ Z]]]$ and can thus be used to lexicalize this S-tree.

Consider now the S-tree in (13).

(13) $[_{YP}\ Y\ [_{ZP}\ Z]]$

For this tree, either L1 or L2 would be a suitable match by the Superset Principle: L1 $[_{YP}\ Y\ [_{ZP}\ Z]]$ is the same size as $[_{YP}\ Y\ [_{ZP}\ Z]]$, and (L2) $[_{XP}\ X\ [_{YP}\ Y\ [_{ZP}\ Z]]]$ is a superset of $[_{YP}\ Y\ [_{ZP}\ Z]]$. Here the second principle resolves the spellout competition.

(14) The Elsewhere Principle
Choose the L which best fits S.

Given two L-trees which can match the same S-tree by the Superset Principle, the L-tree which has the least amount of superfluous features when compared to the S-tree will be chosen to lexicalize the S-tree. Thus, the S-tree in (15) will be spelled out by L1 *goo*, because L1 $[_{YP}\ Y\ [_{ZP}\ Z]]$ is a perfect fit for the S-tree $[_{YP}\ Y\ [_{ZP}\ Z]]$, while L2

$[_{XP}$ X $[_{YP}$ Y $[_{ZP}$ Z]]] has one extra feature compared to the S-tree $[_{YP}$ Y $[_{ZP}$ Z]].

(15) $[_{YP}$ Y $[_{ZP}$ Z]] => *goo*

With these two principles in mind, the syncretism pattern in (9) can now be derived completely. Let us take each of the three S-trees which can be built by the syntactic component and see how each one is spelled out. Recall that the lexicon is the one shown in (16).

(16) L1 < $[_{YP}$ Y $[_{ZP}$ Z]] ⇔ *goo* >
L2 < $[_{XP}$ X $[_{YP}$ Y $[_{ZP}$ Z]]] ⇔ *gur* >

The first S-tree which can be built by the syntax is shown in (17).

(17) S1 $[_{ZP}$ Z] => *goo*

This S-tree can be matched by either L1 or L2, since both of these are supersets of $[_{ZP}$ Z]. However, L1 is a better match by the Elsewhere Principle, since it has only one extra feature while L2 has two extra features.

The second S-tree which can be built by the syntax is shown in (18).

(18) S2 $[_{YP}$ Y $[_{ZP}$ Z]] => *goo*

Once again this S-tree can be matched by either L1 or L2 due to the Superset Principle, since L1 is the same size as S2 and L2 is a superset of S2. By the Elsewhere Principle, however, L1 is a better match, since it fits S2 perfectly while L2 has one superfluous feature. Thus the Y/Z syncretism (*goo*) has been captured: thanks to the Superset Principle, L1 matches either the smaller $[_{ZP}$ Z] structure or the larger $[_{YP}$ Y $[_{ZP}$ Z]] structure as *goo* (and the Elsewhere Principle prevents interference from the morpheme *gur*, whose lexical structure is too large). Put differently, a syncretism amounts to a single L-tree which can match multiple S-trees.

The third and final S-tree which can be built is shown in (19).

(19) S3 $[_{XP}$ X $[_{YP}$ Y $[_{ZP}$ Z]]] => *gur*

Only L2 can match this S-tree, since L1 is too small, so the spellout for this S-tree is *gur*.

Importantly, these principles account for the adjacency effect in syncretism patterns. The adjacency effect can be restated in terms of the fact that we do not observe so-called ABA patterns in the data (Bobaljik 2007, 2012; Caha 2009). Table 3 illustrates what is meant by ABA patterns.

Table 3. *ABA

	DEM	COMP	REL	WH
Unattested 1	A	B	A	
Unattested 2		A	B	A
Unattested 3	A	B	B	A

The lack of ABA patterns[11] can be explained by nanosyntactic principles of spellout, as we will explain next (following Caha 2009: §2.3).

Consider the ABA pattern illustrated in (20).

(20) S1 $[_{ZP}$ Z] => *A*
S2 $[_{YP}$ Y $[_{ZP}$ Z]] => *B*
S3 $[_{XP}$ X $[_{YP}$ Y $[_{ZP}$ Z]]] => *A*

According to the *ABA theorem, it should be impossible for both S1 and S3 to spell out as one thing (*A*), while the middle structure S2 spells out as another (*B*). To see why this is so, imagine what an attempt to generate the ABA pattern might look like. We might, for instance, posit the lexical entries in (21).

(21) L4 $<[_{XP}$ XP $[_{YP}$ YP $[_{ZP}$ ZP]]] $\Leftrightarrow$ *A* >
L5 $<[_{YP}$ YP $[_{ZP}$ ZP]] $\Leftrightarrow$ *B* >

L4 can match either the largest structure (S3) or the smallest structure (S1) by the Superset Principle (since the L-tree in L4 is the same size as S3 and is also a superset of S1), while L5 is a perfect match for S2. By only taking the Superset Principle into account, then, it might at first glance appear possible to generate an ABA pattern. However, taking the Elsewhere Principle into account shows why this will not work. While L4 can in principle match S1, L4 is not the best fit for S1, since L5 is also available in the lexicon. While L4 has two extra features compared to S1, L5 has only one. Therefore L5 wins the competition to spell out S1.

(22) S1 $[_{ZP}$ Z] => *B* L5 is a better fit than L4
S2 $[_{YP}$ Y $[_{ZP}$ Z]] => *B* L5 is a perfect fit
S3 $[_{XP}$ X $[_{YP}$ Y $[_{ZP}$ Z]]] => *A* L4 is the only fit

The ABA pattern is thus not derivable and dissolves into an 'BBA' pattern instead.

[11] Though the *ABA generalization is crosslinguistically very robust (Bobaljik 2007, 2012; Caha 2009), there are – as with any generalization – some exceptions which have been discovered. The status of 'gaps' in the functional sequence is currently an issue being debated in nanosyntactic work (see Caha 2009: §9.3, Starke 2013, Vanden Wyngaerd 2014).

With these principles of matching/spellout in mind, consider now what we need for the Dem/Comp/Rel/Wh data discussed above. We need at least four features/heads, where each one corresponds to a layer in the functional sequence, and each element (Dem, Comp, Rel, Wh) corresponds to a progressively larger featural structure. (Recall that these features/heads are at the word-internal level and not at the sentential level, the level at which these items are traditionally studied.[12])

(23)

$[_{F1P}\ F_1]$	=>	*W*	=	Dem	*or*	Wh
$[_{F2P}\ F_2\ [_{F1P}\ F_1]]$	=>	*X*	=	Comp	*or*	Rel
$[_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]]$	=>	*Y*	=	Rel	*or*	Comp
$[_{F4P}\ F_4\ [_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]]]$	=>	*Z*	=	Wh	*or*	Dem

Recall that the only linear order which can neatly account for the syncretism facts is one in which Dem is adjacent to Comp, Comp is adjacent to Rel, and Rel is adjacent to Wh. When it comes to the cumulatively larger structures in (23), this linear order can be satisfied in two possible ways. That is, either Dem is the smallest, Comp is one feature larger than Dem, Rel is one feature larger than Comp, and Wh is the largest structure; or Wh is the smallest, Rel is one feature larger than Wh, Comp is one feature larger than Rel, and Dem is the largest structure. This will be settled in the next section on the basis of empirical material, but for now let us choose the second scenario, shown in (24), just for the sake of illustrating how the nanosyntactic principles of spellout work for the English and French patterns.

(24)

$[_{F1P}\ F_1]$	=>	*W*	=	Wh
$[_{F2P}\ F_2\ [_{F1P}\ F_1]]$	=>	*X*	=	Rel
$[_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]]$	=>	*Y*	=	Comp
$[_{F4P}\ F_4\ [_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]]]$	=>	*Z*	=	Dem

Given the second scenario, the English and French morphemes are provided in (25), with boldface indicating syncretism.

(25)

$[_{F1P}\ F_1]$	=>	Eng. *what*	Fr. ***que***	Wh
$[_{F2P}\ F_2\ [_{F1P}\ F_1]]$	=>	Eng. ***that***	Fr. ***que***	Rel
$[_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]]$	=>	Eng. ***that***	Fr. ***que***	Comp
$[_{F4P}\ F_4\ [_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]]]$	=>	Eng. ***that***	Fr. *ce*	Dem

12 Since we are dealing with word-internal structure, questions about, for instance, the doubly-filled Comp filter or the *that*-trace effect are not directly relevant. The projections involved in our Comp structure are not clausal projections like Spec-ForceP or Spec-FinP (or simply Spec-CP), but rather very fine-grained projections which are internal to the Comp word *that* itself. Note that our nanosyntactic structures do not even make use of specifiers (see Starke 2004) – or at least the (non-)existence of specifiers has no bearing on the main points we are trying to make.

The lexicon for English, consisting of L6 and L7, is given in (26); the one for French, consisting of L8 and L9, is provided in (27).

(26) English lexicon
L6 $< [_{F1P}\ F_1] \Leftrightarrow$ *what* >
L7 $< [_{F4P}\ F_4\ [_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]]] \Leftrightarrow$ *that* >

(27) French lexicon
L8 $< [_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]] \Leftrightarrow$ *que* >
L9 $< [_{F4P}\ F_4\ [_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]]] \Leftrightarrow$ *ce* >

We will now turn to each S-tree and see how English and French spell them out according to what is available in their respective lexicons.

The possible S-trees and their spellouts in English and French are provided in (28–31), with a short explanation for each spellout.

(28) a. $[_{F1P}\ F_1]$ => Eng. *what* L6 is a better match than L7
b. $[_{F1P}\ F_1]$ => Fr. *que* L8 is a better match than L9

(29) a. $[_{F2P}\ F_2\ [_{F1P}\ F_1]]$ => Eng. *that* L7 is the only match
b. $[_{F2P}\ F_2\ [_{F1P}\ F_1]]$ => Fr. *que* L8 is a better match than L9

(30) a. $[_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]]$ => Eng. *that* L7 is the only match
b. $[_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]]$ => Fr. *que* L8 is a better match than L9

(31) a. $[_{F4P}\ F_4\ [_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]]]$ => Eng. *that* L7 is the only match
b. $[_{F4P}\ F_4\ [_{F3P}\ F_3\ [_{F2P}\ F_2\ [_{F1P}\ F_1]]]]$ => Fr. *ce* L9 is the only match

In English, the Dem/Comp/Rel syncretism is captured by lexical entry L7: L7 can match multiple structures by the Superset Principle without any interference from L6, which is too small to spell out anything beyond the smallest S-tree (28). In French, the Comp/Rel/Wh syncretism is captured by lexical entry L8: L8 can, again, match multiple structures by the Superset Principle, and there is no interference from L9 in (28–31) due to the Elsewhere Principle, which guarantees that L8 will be chosen over L9 in (28–31) because L8 is a better fit. For the largest S-tree (31), however, L8 is too small and L9 becomes the only possible match.

Each language discussed in Table 1 will have its own language-specific inventory of lexical entries in its lexicon. Swedish, for instance, will have four separate lexical entries, each one perfectly matching one of the four S-trees in (28–31). In this kind of lexicon there will be no chance for the Superset Principle to operate and thus no chance of syncretism. The nanosyntactic view of spellout leads us to an interesting conclusion about language. All languages share the universal functional sequence ($F_4 > F_3 > F_2 > F_1$), but they do not lexically 'package' the fseq in the same

way. Thus, linguistic variation can be captured in terms of the properties of lexical entries. Different lexical entries will lead to different spellouts of the same underlying S-trees (as we saw for English and French above). Thus we can think of crosslinguistic variation as arising from the fact that languages store differently sized chunks of the functional sequence in their lexicon (see Starke 2011 for more discussion of this idea).

4. Hierarchical order and the nominal core

We have now seen how nanosyntactic principles of spellout can account for the adjacency effect in syncretism patterns, but there is still the outstanding question of whether the hierarchical ordering of our four elements in Tables 1 and 2 should be Dem > Comp > Rel > Wh or the opposite, namely Wh > Rel > Comp > Dem. This question comes down to whether Wh is the smallest structure and Dem the largest, or Wh the largest and Dem the smallest.

This kind of issue can be solved in a number of ways. Often nanosyntacticians turn to morphological containment. For instance, Icelandic Dem *það* could be analyzed as bimorphemic *þ-að*, which contains the Comp morpheme *að*, suggesting that Dem is structurally larger than and therefore above Comp, i.e. Dem > Comp > Rel > Wh. However, the opposite containment relation seems to be observed in Serbo-Croatian Dem *to*, which is apparently contained within Comp/Rel/Wh *što*, suggesting the opposite hierarchy, i.e. Wh > Rel > Comp > Dem. While we are convinced that these morphological containment relations can eventually be cleaned up and clarified with further research (see Baunaz & Lander *to appear* a for the Slavic case), at this point we choose instead to explore a different strategy for determining the merge order of our fseq, since the containment facts are at this stage inconclusive.

We will propose to solve the hierarchy issue by adopting a line of thinking from Cardinaletti & Starke (1999), who show on the basis of semantic, syntactic, morphological, and prosodic properties that there is a tripartite typology of pronouns (strong, weak, and clitic). For instance, semantically speaking, strong pronouns must be referential, while weak and clitic pronouns are unspecified for refentiality and can even be expletive. When they refer, weak and clitic pronouns need to be associated with a prominent discourse antecedent, unlike strong pronouns (see Cardinaletti & Starke 1999: §2.5 for a more thorough definition of referentiality). In addition, clitic pronouns are more morphologically and prosodically deficient than weak pronouns, which are more deficient than strong pronouns. They propose that strong pronouns have the full structure [C [Σ [I]]], weak pronouns the smaller structure [Σ [I]], and clitic pronouns the even smaller structure [I] (where C = referential and human features, Σ = prosodic features, I = phi features).

For our purposes, it could be imagined that a morphological item exists which participates in the basic syncretism patterns identified in Table 1, and which is also a bound morpheme. If this hypothetical element is a bound morpheme – that is, morphologically and prosodically weak/dependent – then its structure should also be very small. By this logic, then, depending on which end of the Dem | Comp | Rel | Wh spectrum the bound morpheme is syncretic with, we would be able to locate the 'smaller' end of the hierarchy. The two options are sketched in Tables 4 and 5.

In Table 4, the bound morpheme *-W* is syncretic with the Wh morpheme *W*, meaning that Wh is at the structurally smaller end of the hierarchy,

Table 4. Possibility 1 – Wh is the lowest/smallest => Dem > Comp > Rel > Wh

	DEM	COMP	REL	WH	Bound morpheme
Possibility 1	Z	Y	X	W	-W

Table 5. Possibility 2 – Dem is the lowest/smallest => Wh > Rel > Comp > Dem

	WH	REL	COMP	DEM	Bound morpheme
Possibility 2	W	X	Y	Z	-Z

giving Dem > Comp > Rel > Wh (> Bound morpheme). In Table 5, the bound morpheme *-Z* is syncretic with the Dem morpheme *Z*, meaning that Dem is at the structurally smaller end of the hierarchy, giving Wh > Rel > Comp > Dem (> Bound morpheme). Below we will see data from a number of languages that the first option in Table 4 is the correct answer.

In this section we will give empirical arguments in favor of the existence of what we term the *nominal core*. We start from the observation that quantifiers like Fr. *cha-que* 'each, every' and *quel-que* 'some' are built using a bound morpheme (*-que*) which appears to be identical to the complementizer (Fr. *que*) (see also Szabolcsi et al. 2014 for discussion of the internal structure of quantifiers in Hungarian and Japanese). This kind of morpheme is what we call the nominal core, and we propose that it is a semantically bleached, non-referential functional element which can be found in certain nominal environments (e.g. combined with independently built operators like Fr. *cha-*, *quel-*). The semantic vacuousness of this element can, in addition to its prosodic weakness, be considered another reason to assign it a very small structure.

We assume that *n* is not a full lexical noun (i.e. of category N, such as Fr. *garçon* 'boy'), but rather a functional category which we call the

nominal core (i.e. little *n*, as opposed to big N). There are reasons to think that *n* itself has some internal functional structure. For the purposes of this paper, however, 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 the different kinds of *n* (on which see Baunaz & Lander 2017).[13]

The behavior of the nominal core with regard to syncretism patterns, furthermore, is crucial for determining the direction of our fseq. As sketched in Tables 4 and 5, if the nominal core is found to syncretize with the Dem-end of the fseq, then we know that Dem is at the lower end of the sequence; if the nominal core is found to syncretize with the Wh-end, then we know that Wh is at the lower end of the sequence. In the next sections we discuss the relevant items and patterns according to language group.

4.1. *The nominal core in Germanic*

An argument in favor of the existence of nominal cores in natural languages can be found in the guise of various North and West Germanic interrogative, indefinite, and (non-D-linked) demonstrative pronouns like Eng. *which*, *each*, *such*.

(32) a. Eng. whi-**ch** (< Old Eng. *hwi-**lc***)
b. Ger. we-**lch**-
c. Du. we-**lk**-
d. Sw. vi-**lk**-

(33) a. Eng. ea-**ch** (< Old Eng. *ǣ-**lc***)
b. Du. e-**lk**-

(34) a. Eng. su-**ch** (< Old Eng. *swi-**lc***)
b. Ger. so-**lch**-
c. Du. zu-**lk**-
d. Nor. s-**lik**-

As seen in (32–34), these items can be decomposed into at least two components. The first component we take to be a quantificational operator. The second component (Eng. *-ch*, Ger. *-lch,* Du. *-lk*, Sw. *-lk*), on the other hand, expresses something along the lines of 'form', which makes sense from a historical point of view since these morphemes descend from the Old Germanic noun **līk*- 'body, form' (see Leu 2015: §6.2.1 and references cited there). This component, then, can be thought of as a nominal core, more specifically n_{FORM}.

[13] The term *indeterminate* pronoun could also be used, which refers to phrases that are generally associated with different operators in Japanese Grammar (Kuroda 1965). Indeterminate pronouns are generally invariable for number across languages. They are also sometimes called *light nouns* (Kishimoto 2000, Leu 2005, among others).

Another kind of nominal core in Germanic can be identified by looking at quantifiers. Eng. *every-thing*, *some-thing*, *no-thing* and Sw. *all-ting*, *nå(go)n-ting*, *ingen-ting*, for instance, show the nominal core n_{THING}. Eng. *every-one*/*-body*, *some-one*/*-body*, *no-one*/*-body*, etc. show the nominal core n_{BODY}. Note here that n_{BODY} and n_{THING} are invariant forms, in the sense that they do not inflect. They are thus not to be equated with big N Eng. *body* and *thing* (or Sw. *grej*/*sak* 'thing'). As seen in (35) and (36), Eng. *thing* and *body* and Sw. *grej*/*sak* show agreement for plural.

(35) a. We already have that thing.
b. We already have all those things.
c. My body can't handle this stress.
d. Our bodies aren't built for such forces.

(36) a. Vi har den grej-en / sak-en redan.
we have that thing-the thing-the already
b. Vi har alla dom grej-er-na / sak-er-na redan.
we have all those thing-PL-*the thing*-PL-*the already*

Big N, in other words, inflects in a way that little *n* does not.

Now, in terms of syncretism, the nominal core in Germanic is of little help in identifying the direction of the fseq. This is because *-ch*/*-lch*/*-lk* and *-thing*/*-ting*, *-body*/*-one* are not syncretic with Dem, Comp, Rel, or Wh. See Tables 6 and 7.

Table 6. Nominal core is not syncretic with Wh

		DEM	COMP	REL	WH	*n*
North Gmc	*Swedish*	det	att	som	vad	-lk -ting
West Gmc	*English*	that	that	that	what	-ch -thing / -body
	Dutch	dat	dat	dat	wat	-lk
	German	das	dass	das	was	-lch

Table 7. Nominal core is not syncretic with Dem

		WH	REL	COMP	DEM	*n*
North Gmc	*Swedish*	vad	som	att	det	-lk -ting
West Gmc	*English*	what	that	that	that	-ch -thing / -body
	Dutch	wat	dat	dat	dat	-lk
	German	was	das	dass	das	-lch

Purely on the basis of the Germanic nominal core, then, we cannot decide in which direction our hierarchy is built. Let us therefore move on to other language groups, in order to see if their syncretism patterns with the nominal core are more informative.

4.2. *The nominal core in Romance*

In Romance, the nominal core can be identified in quantifiers like Fr. *cha-que* 'each', *quel-que* 'some' and It. *cias-c-uno* 'each (one)', *qual-che* 'some'. In (37) we provide the relevant forms in French, and in (38) we provide the relevant forms in Italian.[14]

(37) a. cha-**que** 'each/every'
b. cha-**c**-un (des N) 'each/every (of.the Ns)'
c. quel-**que** 'some'
d. quel-**qu**'un (des N) 'someone/somebody (of.the Ns)'
e. quel-**que** chose 'something'

(38) a. cias-**c**-uno ((dei/delle) N) 'each(one)/every(one) ((of.the) Ns)'
b. cias-**che-**duno ((dei/delle) N) 'each (one)/every(one) (of.the) Ns)'
c. qual-**che** 'some'
d. qual-**c-**uno (dei/delle N) 'some(one) ((of.the) Ns)'
e. qual-**che-**duno 'someone'
f. qual(-**che**) cosa[15] 'something'

Again, we take the first component to be a quantificational operator (Fr. *cha-*, *quel-* and It. *cias-*, *qual-*), and the second component (Fr. *-que* or reduced *-c* and It. *-che* or reduced *-c*)[16] to be the nominal core.[17] These morphemes derive from the same Latin source – namely the morpheme *-que* in *quis-que* 'every/each' and *qualis-que* 'some' – suggesting that these elements have been syncretic throughout history (see also Buck 1949 and Leu 2015). Note here that Spanish followed a different path, considering

[14] A similar observation is made by Leu (2015: 109), though he considers bound *-que* to be an instance of Comp *que* (where both are hosted by the C head).

[15] *Qualcosa* is a contraction of *qualche* + *cosa* (Patota 2002). Similarly to Germanic *-thing*/*-ting* and *-body* above, there is a difference in inflectional capabilities between 'functional' or 'semi-lexical' Fr. *chose*/It. *cosa* in *quelque chose*/It. *qual-che cosa* 'something' and 'lexical' Fr. *chose*/It. *cosa*, where the former is invariant while the latter can agree in gender and number (see Leu 2005 and references cited there for discussion). However, in Romance the nominal core is Fr. *-que*/It. *-che* (not Fr. *chose*/It. *cosa*). Thus the functional *chose*/*cosa* does not correspond to the structure of little *n*, but may instead be thought of as a structurally impoverished kind of big N, for instance a lexical N which cannot project NumP (thanks to a reviewer for this suggestion). See also Leu (2009).

[16] Note that depending on its phonological environment this morpheme is pronounced /k/ in French and Italian, with elision of the schwa in French and of /e/ in Italian (cf. Grammont's 1894 *Law of Three Consonants*).

[17] We consider elements like Fr. *bien que*, It. *benché* 'although' or It. *perché* 'why' to be complex Comps, that is, they are the result of combining an adverb (*bien*, *ben-*) or preposition (*per-*) with the full Comp. Thanks to a reviewer for raising this issue.

Sp. *cada* 'each/every', *alguno* 'some', *alguien* 'someone'. Interestingly, however, Old Spanish had a version of *algun* in which *-que* was used, i.e. *cual-que* (cf. It. *qual-que*).

It should be clear that the nominal core in Romance is syncretic with Comp. Not only is the nominal core syncretic with Comp, but also with Rel and Wh. Thus we have a syncretism pattern which locates the small end of our hierarchy at Wh, since there is a Wh/*n* syncretism observed, as shown in Table 8.

Table 8. Nominal core is syncretic with Wh

		DEM	COMP	REL	WH	*n*
Romance	*French*	ce	que	que	que	-que
	Italian	quello	che	che	che	-che
	Spanish	aquél	que	que	qué	(-que)

The syncretism pattern in which the nominal core participates in Romance, then, indicates that the hierarchy should be Dem > Comp > Rel > Wh > *n*.

4.3. *The nominal core in Modern Greek*

The nominal core in Greek participates in a syncretism pattern which is very similar to the one in Romance. In (39) and (40) we provide the relevant forms, where *(-)ti(-)* and *(-)pu(-)* are bound morphemes which are syncretic with Comp/Rel/Wh *(o)ti* and *(o)pu*.

(39) a. ká-**ti** 'something'
b. **tí-**pota 'anything'
c. **TÍ-**POTA 'nothing'

(40) a. ká-**pu** 'somewhere'
b. **pu-**thená 'anywhere'
b. **PU**-THENÁ 'nowhere'

In (39a) and (40a) there is an overt existential operator *ká-* which attaches to the nominal core. In (39b-c) and (40b-c) there is an adverbial element (**-***thená* and *-pota*/*poté*) which is attached to the nominal core (note that stressed *TÍ-POTA* and *PU-THENÁ* are interpreted as Neg-words; see Giannakidou 2012).[18]

[18] Some Modern Greek dialects use *tipote* (cf. Medieval Greek *tipote* 'nothing', with final *e* turning into *a* possibly by analogy with other adverbs, e.g. *kodá* 'near', *makrjá* 'far'). Synchronically, *pota*/*poté* is the polarity adverb '(n)ever'. Thanks to M. Bağrıaçık (p.c.) and A. Ioannidou (p.c.) for their help with Greek etymology. See also Holton et al. (1997). For *puthená* it is possible that, at least historically, there are three morphemes involved: *pu-* 'where' plus the ablative case ending *-then* 'from' plus the adverbial particle *-a*. Thanks to M. Bağrıaçık (p.c.) for discussion.

Observe here that the bound functional element *(-)ti(-)* '-thing' (i.e. n_{THING}) must be distinguished from the lexical item *prágma* 'thing' in Greek, since *prágma* inflects for number, as seen in (41).

(41) a. to pragma pu mu pires
the thing that me bought
'the thing you bought for me'
b. ta pragma-ta pu mu pires
the.PL *thing*-PL *that me bought*.2SG
'the things you bought for me'

Number inflection is impossible in the invariant nominal core *-ti*, which according to us corresponds simply to n_{THING}.

The nominal core *(-)pu(-)*, found in *ká-pu* 'somewhere' and *pu-thená* 'anywhere', corresponds to n_{PLACE}. We may ask at this point why we still attribute a nominal, rather than an adverbial, category to an item with the meaning 'place'. We assume that this is so on the basis of the syncretism patterns in which they participate. That is, the fact that Gk. *(-)pu(-)* participates in a nominal syncretism pattern reveals that it is actually a nominal core (even though on the surface it would appear to be an 'adverbial core'). See also our discussion of Hungarian in section 4.5.

The Greek nominal core syncretisms are in agreement with our findings from Romance. The nominal cores *(-)ti(-)* and *(-)pu(-)* are syncretic with Comp, Rel, and, crucially, Wh.

Table 9. Nominal core is syncretic with Wh

	DEM	COMP	REL	WH	*n*
Modern Greek	ekíno	o-ti	ó-ti	tí	(-)ti(-)
	ekeí	pu	(ó-)pu	pú	(-)pu(-)

Since these bound morphemes overlap with the Wh-element, we know that Wh is at the small/low end of the fseq.

4.4. *The nominal core in Serbo-Croatian*

As seen in Table 1, SC *što* instantiates a Comp/Rel/Wh syncretism. The element *što* also appears as a bound morpheme *-što* with the meaning 'thing' in quantifiers such as those given in (42).[19]

(42) a. sva-**šta** 'everything'
b. ni-**šta** 'nothing'
c. ne-**što** 'something'
d. bilo-**šta** 'anything'

[19] There is variation as to the use of *što* or *šta* with Comp, Rel, and Wh items among Serbo-Croatian speakers (T. Samardžic and T. Sočanac, p.c.).

Again, functional *-što* '-thing' (i.e. n_{THING}) must be distinguished from lexical *stvar* 'thing' in Serbo-Croatian. While lexical 'thing' can inflect for number, as shown in (43) for *stvar*, functional 'thing' cannot.

(43) a. Imamo već tu stvar.
have.1PL *already that thing*
'We already have that thing.'
b. Imamo sve ove stvar-i.
have.1PL *all these thing*-PL
'We have all these things.'

In other words, there is a clear distinction between the lexical N *stvar* 'thing' and the functional n_{THING} *-što*.[20]

The nominal core n_{BODY} in Serbo-Croatian can be seen in (44).

(44) a. sva-**(t)ko** 'everyone'
b. ni-**ko** 'no one'
c. ne-**ko** 'someone'
d. bilo-**ko** 'anyone'

The nominal core *-ko* is in fact syncretic with the Rel/Wh stem *(t)ko-* 'who, which'.[21] Interestingly, it also appears in the interrogative Comp *a-ko* 'if'.[22] Thus both n_{BODY} and n_{THING} participate in a Comp/Rel/Wh/*n* syncretism. This is seen in Table 10.

Table 10. Nominal core is syncretic with Wh

	DEM	COMP	REL	WH	*n*
Serbo-Croatian	to	što	što	što	-što
		a-ko 'if'	ko-	ko-	-ko

We conclude that the syncretism pattern in Serbo-Croatian agrees with the Romance and Greek patterns from the previous sections. Since *n* is syncretic with Wh in Serbo-Croatian, we have yet more evidence that Wh is located at the small end of the fseq.

[20] The two are not historically related either: *stvar* is etymologically a nominalization of the verb 'make' (i.e. 'the result of making'), while *što* derives from the Wh-morpheme *k-*, a front mid vowel (which palatalized *k-*), and the N.SG Dem *to* (B. Arsenijević, T. Sočanac, p.c).

[21] Relevant here is Heine & Kuteva's (2002:250-251) grammaticalization path *w-question* (Wh-word) > *indefinite pronoun*.

[22] We note that *-ko* also appears within *kako* 'how'.

4.5. *The nominal core in Hungarian*

Hungarian has generalized the use of *hogy* 'how' as its Comp, for both factive and non-factive predicates. Comp *hogy* is partially syncretic with both Rel and Wh pronouns of manner.

Table 11. Nominal core is syncretic with Wh

	DEM	COMP	REL	WH
Hungarian	így 'in this manner'	hogy	a-hogy	hogy(-an)

Recall from section 2.2 that *a-* in Rel *a-hogy* is a D-item of some kind (as is the case for relativization in a lot of languages, cf. MG ***ó-*** in ***ó-****ti*/***ó-****pu* and Italian ***il-*** in ***il-****quale*, etc.) and that the suffix *-an* in Wh *hogy(-an)* is in the process of being lost.

After our discussion of Romance, Greek, and Serbo-Croatian, it will come as no surprise that one possible nominal core in Hungarian is *-hogy*, as seen from the adverbial quantifiers in (45).

(45) a. vala-**hogy**(-an) 'somehow, anyhow'
b. minden-**hogy**(-an) 'everyhow'
c. se-**hogy**(-an) 'no-how'

The kind of nominal core seen in (45) is more precisely what we label n_{MANNER}.

In fact, it seems likely that n_{MANNER} *hogy* is not monomorphemic and can be decomposed further, into *ho-gy*. The first item, *ho-*, has its own paradigm in Hungarian (e.g. locative *ho-l* 'where', goal *ho-va* 'to where', source *ho-nnan* 'from where'; Szabolcsi et al. 2014:11, fn.3) and seems to be quite similar to Gk. *pu* 'where', i.e. n_{PLACE}. It is this item, *ho-*, which seems to be the true nominal core. The second item, *-gy*, appears to be responsible for the manner reading (e.g. *ú-gy* 'in that manner, so', *í-gy* 'in this manner, so', *na-gy* 'big in that way, so big', etc.).[23]

Taking into account the nominal core, the syncretism pattern observed in Hungarian is the following.

Table 12. Nominal core *ho-* is syncretic with Wh

	DEM	COMP	REL	WH	*n*
Hungarian	**í**-gy 'in this manner, so'	**ho**-gy	a-**ho**-gy	**ho**-gy(-an)	**ho**-gy(-an)

[23] Thanks to G. Puskás (p.c.) for discussion and examples.

Thus, even though the meaning of the nominal core is at first glance adverbial, we choose to analyze it as a nominal element n_{MANNER}. This is based on a comparison of the syncretism pattern in Hungarian with the ones seen in Indo-European, which are clearly nominal. Since Hungarian shows a syncretism pattern implicating Comp, Rel, and Wh – which is extremely similar to Romance, Greek, and Serbo-Croatian – we have chosen to analyze the Hungarian pattern, deep down, as being nominal as well.

Indeed, a similar type of pattern can even be observed in Germanic. As seen in Table 13, English, Dutch, and German display embedded tensed clauses introduced by the Comp 'how' (Eng. *how*, Du. *hoe*, and Ger. *wie*). It has been shown that these clauses are propositions (Melvold 1991). Importantly, *how*-clauses and *that*-clauses are roughly similar in meaning, although *how*-clauses seem to be related to factivity (see Legate 2010; Nye 2010, 2012, 2013 for developments and references).

Table 13. Declarative *how*-complementizers in West Germanic

		DEM	COMP	REL[24]	WH	n_{MANNER}
WGmc	*English*	so	how	how	how	-how
	Dutch	zo	hoe	hoe	hoe	---
	German	so	wie	wie	wie	-wie

In a similar way to Hungarian, there seems to be a nominal core n_{MANNER} in Germanic as well,[25] considering examples like Eng. *every-how* (archaic), *some-how*, *no-how* or Ger. *irgend-wie* 'somehow' (but not quite Du. *hoe dan ook* 'somehow'). Indeed, nominal cores (to which quantificational operators attach) commonly show Wh-morphology in Germanic: Eng. *some-**what***, *any-**where***, *no-**where***, West Flemish *een-t-**wat*** 'something', *een-t-**wien*** 'someone', *een-t-**waarom*** 'somehow' (L. Haegeman, p.c.), Ger. *irgend-**wie*** 'somehow', *et-**was*** 'something', etc. In Germanic, then, nominal cores very often resemble (in fact are syncretic with) Wh-items. This is yet more evidence that Wh and *n* must be adjacent layers, and since we know that *n* is very small – something we

[24] These Comps are syncretic with some types of relative pronouns, as shown by the acceptability of Eng. *the way how*... or Du. *de wijze hoe*....

[25] Interestingly, Vangsnes (2008) discusses syncretisms with *how*$_{\text{MANNER}}$ in Scandinavian and shows that they include morphemes that are etymologically derived from a direction noun. For instance, Icelandic *hvernig* 'how$_{\text{MANNER}}$' is derived from the nominal phrase *hvern **veg*** 'what way (ACC)' (Vangsnes 2008:126). He proposes that this is part of an abstract template (i.e. WH – (WAY) – s – (en)) which can account for a variety of forms, including forms which have no overt direction noun. We thank a *Studia Linguistica* reviewer for pointing this out to us.

deduce on the basis of its prosodic weakness and semantic vacuousness – we know that Wh is at the small end of the functional hierarchy.

4.6. *The nominal core in Finnish*

In Finnish there is a Rel/Wh syncretism in *mi-*, with non-syncretic Comp (*että*) and Dem (*tä-*) forms. If Finnish is like Romance, Greek, Serbo-Croatian, and Hungarian in having a Wh/*n* syncretism, then we would expect the nominal core(s) in Finnish to look like the Rel/Wh stem *mi-*. However, the nominal core in Finnish is not *mi-*, as seen from the forms given in (46) and (47) (from Karlsson 1999:142).

(46) a. jo-kin ‘something’
b. kumpi-kin ‘each one (of two)’
c. ku-kin ‘each one, everyone’

(47) a. (ei) ku-kaan ‘no one’
b. (ei) mi-kään ‘nothing’
[*ei* = NEG]

We take the invariant particles *-kin* and *-kaan* ~ *-kään* (alternating according to vowel harmony) to be nominal cores. Crucially, they are not syncretic with either Wh or Dem.

Table 14. Nominal core is not syncretic with Wh

	DEM	COMP	REL	WH	*n*
Finnish	tä-	että	mi-	mi-	-kin -kaan ~ -kään

Table 15. Nominal core is not syncretic with Dem

	WH	REL	COMP	DEM	*n*
Finnish	mi-	mi-	että	tä-	-kin -kaan ~ -kään

Thus the Finnish nominal cores are no help in determining the hierarchy of our elements. However, the evidence from Romance, Greek, Serbo-Croatian, and Hungarian are more than enough to establish the hierarchy Dem > Comp > Rel > Wh > *n*.

4.7. *Germanic quantifier formation and *ABA*

Taking a few steps back, let us return to a salient difference between Germanic and Romance. Had we not known anything about the syncretism patterns and quantifier formation discussed above, it would

be reasonable to wonder why Romance, for instance, makes use of a morpheme which resembles Comp in its quantifiers (Fr. *cha-**que***, *quel-**que***; It. *cias-**c**-uno*, *qual-**che***) while in Germanic this is completely ungrammatical: Eng. **every-**that***, **some-**that***, **no-**that***; Sw. **all-**att***, ***nå(go)n-**att***, ***ingen-**att***.

This difference can be understood in terms of syncretism and the *ABA theorem. Recall from above that Romance displays a large span of syncretism from Comp all the way down to *n* (Table 16), while Germanic does not, that is, English shows a Dem/Comp/Rel syncretism and Swedish shows no syncretism (Table 17).

Table 16. Syncretism in Romance

		DEM	COMP	REL	WH	*n*
Romance	*French*	ce	que	que	que	-que
	Italian	quello	che	che	che	-che

Table 17. Syncretism in Germanic

		DEM	COMP	REL	WH	*n*
Germanic	*English*	that	that	that	what	-thing
	Swedish	det	att	som	vad	-ting

In Romance the nominal core is syncretic with Comp, giving the impression that Comp participates in quantifier formation (in reality Comp is a large structure composed of multiple functional heads while the nominal core is a much smaller structure corresponding to the head *n*). In Germanic, on the other hand, the nominal core is not syncretic with Comp, so there will be no impression that Comp participates in quantifier formation. Ruling out forms like Eng. **every-**that*** or Sw. **all-**att***, then, boils down to the *ABA theorem. As seen in (48), the pattern in English is *that – what – -thing* (ABC).

(48) $\underbrace{\text{Dem} > \text{Comp} > \text{Rel}}_{\text{Eng. } that} > \underbrace{\text{Wh}}_{what} > \underbrace{n_{\text{THING}}}_{(\text{-})thing}$

An **ABA** pattern (***that*** – *what* – ***that***) would be needed to produce the form **every-that*, so the *ABA theorem neatly rules this out. Similarly, in Swedish the pattern is *att – som – vad – -ting* (ABCD).

(49) Dem > Comp > Rel > Wh > n_{THING}

Sw. *att* — *som* — *vad* — *(-)ting*

An **ABCA** pattern (***att*** – *som* – *vad* – ***att***) would be needed to produce a form like **all-att*, so the *ABA theorem rules out this form as well.

5. Conclusion

This paper proposes that certain complementizers (e.g. Eng. *that*, Fr. *que*) are fundamentally of a nominal nature and have a complex internal structure. This is based on attested patterns of syncretism with nominal Comp. That is, it is common for nominal Comp to be syncretic with Dem, Rel, and Wh pronouns. Working within a nanosyntactic framework, we propose that these syncretisms can be accounted for by positing a single functional sequence responsible for building Dem, (nominal) Comp, Rel, and Wh elements. This hierarchy is seen in (50).

(50) Dem > Comp > Rel > Wh > *n*

The hierarchy in (50) has a number of important properties which we summarize here.

First, as we have emphasized throughout the paper, this is not a functional spine at the clausal or sentential level but rather a functional spine at the level of individual words. The layers in (50) should be interpreted in terms of unary and additive features which give rise to five different structures which the generative component can construct. The hierarchy in (50) is therefore more accurately presented as in (51).

(51) a. **(Op +)** $[_{n\text{P}}\ n]$ => nominal core
b. $[_{\text{F1P}}\ \text{F}_1\ [_{n\text{P}}\ n]]$ => Wh pronoun
c. $[_{\text{F2P}}\ \text{F}_2\ [_{\text{F1P}}\ \text{F}_1\ [_{n\text{P}}\ n]]]$ => Rel pronoun
d. $[_{\text{F3P}}\ \text{F}_3\ [_{\text{F2P}}\ \text{F}_2\ [_{\text{F1P}}\ \text{F}_1\ [_{n\text{P}}\ n]]]]$ => Complementizer
e. $[_{\text{F4P}}\ \text{F}_4\ [_{\text{F3P}}\ \text{F}_3\ [_{\text{F2P}}\ \text{F}_2\ [_{\text{F1P}}\ \text{F}_1\ [_{n\text{P}}\ n]]]]]$ => Dem pronoun

These structures are in superset-subset relations with one another. The structure of the so-called nominal core (51a) corresponds simply to the layer *n*. The structure of a Wh pronoun (51b) corresponds to *n* and F_1. The structure of Rel (51c) corresponds to *n*, F_1, and F_2. The structure of Comp (51d) corresponds to *n*, F_1, F_2, and F_3. Finally, the structure of Dem (51e) corresponds to *n*, F_1, F_2, F_3, and F_4.

Second, the hierarchy in (50) has at its core a non-referential, semantically vacuous nominal element which we have called the nominal

core (*n*) (e.g. Germanic *-thing*/*-ting*, Romance *-que*/*-che*, Gk. *-ti*, SC *-što*, Hungarian *ho-*, Finnish *-kin*) and which essentially classifies the items built in (50)/(51) as nominal elements.[26] In all of these languages the nominal core can be merged with an operator (**Op** in (51)) to create quantificational elements like It. *qual-**che*** or SC *ne-**što***.[27] Importantly, the nominal core is actually syncretic with Wh (as well as Comp and Rel) in Romance, Greek, Serbo-Croatian, and Hungarian.

On the basis of syncretism it is only possible to determine the *linear order* of elements (such that only adjacent elements are syncretic). Another diagnostic is needed to determine the *hierarchical order* of elements. We proposed to study the behavior of the so-called *nominal core* in relation to Dem, Comp, Rel, and Wh in order to determine the merge order of the functional sequence. We reasoned (following ideas in Cardinaletti & Starke 1999) that since the nominal core is semantically quite empty and also a bound morpheme (i.e. morphologically and prosodically deficient), it must be structurally minimal; thus its behavior with regard to syncretism will help us identify structures which have sizes comparable to *n*. Since *n* is very often syncretic with Wh (but not with Dem), this means that Wh must be quite small as well, placing it at the smaller end of the fseq (while Dem corresponds to the largest structure at the other end of the fseq, as in (50)/(51)).

References

Adams, J.N. 2013. *Social variation and the Latin language*. Cambridge: Cambridge University Press.

Baunaz, L. 2015. On the various sizes of complementizers. *Probus* 27(2): 193–236.

Baunaz, L. 2016. Deconstructing complementizers in Serbo-Croatian, Modern Greek and Bulgarian. *Proceedings of NELS 46, Vol. 1*, ed. C. Hammerly & B. Prickett, 69–77.

Baunaz, L. to appear. Decomposing complementizers: The fseq of French, Modern Greek, Serbo-Croatian and Bulgarian complementizers. *Exploring Nanosyntax*, eds. L. Baunaz, K. de Clercq, L. Haegeman, & E. Lander. New York: Oxford University Press.

Baunaz, L. & Lander, E. 2017. The internal structure of ontological categories. Talk given at the Incontro di Grammatica Generative 43. IUSS Pavia. Feb 16–17.

Baunaz, L. & Lander, E. To appear a. Cross-categorial syncretism and the Slavic containment puzzle. *Balkan syntax and (universal) principles of*

[26] In ongoing work we have identified other categorial cores, such as the verbal core.

[27] It is true, of course, that not all quantificational elements are decomposable into a clear Op component plus the nominal core: Fr. *personne* 'no one', *rien* 'nothing'; SC *sve* 'all'; Du. *iets* 'something', among others. Here the possibility of using phrasal spellout in nanosyntax becomes a great advantage. Since it is possible to spell out XPs of varying sizes in nanosyntax, we can straightforwardly posit that these items are portmanteaus whose lexical structures correspond to entire constituents like [[Op] *n*] (as opposed to two separate morphemes, one for [Op] and one for [*n*]). See Starke (2011: 6) for relevant discussion.

grammar (provisional title), eds. I. Krapova & B. Joseph. Berlin: Mouton de Gruyter.
Baunaz, L. & Lander, E. To appear b. Nanosyntax: the basics. *Exploring Nanosyntax*, eds. L. Baunaz, K. de Clercq, L. Haegeman, & E. Lander. New York: Oxford University Press.
Baunaz, L. & Lander, E. Under review. deconstructing categories syncretic with the nominal complementizer. Ms., University of Zürich and University of Gothenburg.
Bacskai-Atkari, J. 2016. On the diachronic development of a Hungarian declarative complementiser. *Transactions of the Philological Society* 114(1): 95–116.
Bobaljik, J. 2007. On comparative suppletion. Ms., University of Connecticut.
Bobaljik, J. 2012. *Universals in comparative morphology: Suppletion, superlatives, and the structure of words*. Cambridge, MA: MIT Press.
Bresnan, J.W. 1970. On complementizers: Toward a syntactic theory of complement types. *Foundations of Language* 6: 297–321.
Bresnan, J.W. 1972. *Theory of complementation in English syntax*. PhD, MIT.
Brook, M. 2014. Comparative complementizers in Canadian English: Insights from early fiction. *University of Pennsylvania Working Papers in Linguistics* 20(2). Available at: http://repository.upenn.edu/pwpl/vol20/iss2/2.
Buck, C.-D. 1949. *A dictionary of selected synonyms in the principal Indo-European languages: A contribution to the history of ideas*. Chicago: University of Chicago Press.
Caha, P. 2009. *The nanosyntax of case*. PhD, University of Tromsø.
Caha, P. 2010. The parameters of case marking and spell out driven movement. *Linguistic Variation Yearbook* 10: 32–77.
Caha, P. 2013. Explaining the structure of case paradigms by the mechanisms of Nanosyntax. The Classical Armenian nominal declension. *Natural Language and Linguistic Theory*, 31(4): 1015–1066.
Cardinaletti, A. & Starke, M. 1999. The typology of structural deficiency. A case study of the three classes of pronouns. *Clitics in the languages of Europe*, ed. H. van Riemsdijk , 145–233. Berlin: Mouton De Gruyter.
Carlier, A. & Mulder, W. 2006. Du démonstratif à l'article défini: Le cas de *ce* en français moderne. *Langue française* 152: 96–113.
Chomsky, N. 1995. *The minimalist program*. Cambridge, MA: MIT Press.
Christensen, K.K. 1983. The categorial status of Norwegian infinitival relatives. *Papers from the Seventh Scandinavian Conference of Linguisitcs*, ed. F. Karlsson, 9–20. Department of General Linguistics, University of Helsinki.
Cinque, G. & Rizzi, L. 2008. The cartography of syntactic structures. *CISCL Working Papers on Language and Cognition* 2: 43–59.
Déchaine, R.-M. & Wiltschko, M. 2002. Decomposing pronouns. *Linguistic Inquiry* 33: 409–443.
De Clercq, K. 2013. *A unified syntax of negation*. PhD, Ghent University.
Dékány, É. 2009. The nanosyntax of Hungarian postpositions. *Nordlyd* 36 (Special issue on Nanosyntax), 41–76.
Dixon, R.M.W. 2006. Complement clauses and complementation strategies in typological perspective. *Complementation: A cross-linguistic typology*, eds. R.M.W. Dixon & A.Y. Aikhenvald, 1–48. New York: Oxford University Press.
Giannakidou, A. 2009. The dependency of the subjunctive revisited: Temporal semantics and polarity. *Lingua* 119: 1883–1908.
Giannakidou, A. 2012. The landscape of Greek quantifiers. *Handbook of quantifiers in natural language (Studies in Linguistics and Philosophy 90)*, eds. E.L. Keenan & D. Paperno, 285–346. Dordrecht: Springer.

Grammont, M. 1894. La loi des trois consonnes. *Mémoires de la Société de Linguistique de Paris* 8: 53–90.
Grimshaw, J. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10 (2): 279–326.
Grimshaw, J. 1991. Extended projection. Ms., Brandeis University.
Haugen, E. 1976. *The Scandinavian languages*. Cambridge, MA: Harvard University Press.
Heine, B. & Kuteva, T. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
Holton, D., Mackridge, P. & Philippaki-Warburton, I. 1997. *Greek: A comprehensive grammar of the modern language*. London: Routledge.
Karlsson, F. 1999 [1983]. *Finnish: An essential grammar*. London & New York : Routledge.
Kayne, R. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
Kayne, R. 2008. Why isn't this a complementizer? Ms., New York University.
Kayne, R. & Pollock, J.-Y. 2010. Notes on French and English demonstratives. *Structure preserved: Studies in syntax for Jan Koster*, eds. J.W. Zwart & M. de Vries, 215–228. Amsterdam: John Benjamins.
Keevallik, L. 2008. Conjunction and sequenced actions: The Estonian complementizer and evidential particle *et*. *Crosslinguistic studies of clause combining: The multifunctionality of conjunctions*, ed. R. Laury, 125–152. Amsterdam: John Benjamins.
Kiparsky, P. & Kiparsky, C. 1970. Fact. *Progress in linguistics*, eds. M. Bierwisch & K. Heidolph, 143–173. The Hague: Mouton.
Kishimoto, H. 2000. Indefinite pronouns and overt N-raising. *Linguistic Inquiry* 31(3): 557–566.
Kuroda, S.-Y. 1965. *Generative grammatical studies in the Japanese language*. PhD, MIT.
Lakoff, R. 1968. *Abstract syntax and Latin complementation*. Cambridge, MA: MIT Press.
Lasnik, H. & Saito, M. 1992. *Move Alpha: Conditions on its application and output*. Cambridge, MA: MIT Press.
Legate, J.A. 2010. On how *how* is used instead of *that*. *Natural Language and Linguistic Theory* 28: 121–134.
Le Goffic, P. 2008. *Que* complétif en français: Essai d'analyse. *Langue française* 158: 53–68.
Leu, T. 2005. Something invisible in English. *University of Pennsylvania Working Papers in Linguistics* 11(1): 143–154.
Leu, T. 2008. *The internal syntax of determiners* (GAGL 47). PhD, New York University.
Leu, T. 2009. From Greek to Germanic: Poly-(*in)-definiteness and weak/strong adjectival declension. *Merging features: Computation, interpretation and acquisition*, ed. A. Gavarró, J.M. Brucart & J. Solà. Oxford: Oxford University Press.
Leu, T. 2015. *The architecture of determiners*. Oxford: Oxford University Press.
Manzini, M.R. & Savoia, L.M. 2003. The nature of complementizers. *Rivista di grammatica generativa* 28: 87–110.
Manzini, M.R. & Savoia, L.M. 2011. *Grammatical categories*. Cambridge: Cambridge University Press.
Melvold, J. 1991. Factivity and definiteness. *More Papers on Wh-movement*. *MIT Working Papers in Linguistics* 5: 97–117.
Noonan, M. 2007. Complementation. *Language typology and syntactic description, Vol. II: Complex constructions*, ed. T. Shopen, 52–150. Cambridge: Cambridge University Press.

Nye, R. 2010. The categorial status of Dutch and English declarative *hoe-*/*how*-complement clauses. Paper presented at the 19th International Postgraduate Linguistics Conference on Language Variation and Change, University of Manchester.
Nye, R. 2012. The categorial status of Dutch and English declarative *hoe-*/*how*-complement clauses. *Salford Working Papers in Linguistics and Applied Linguistics*, ed. M. van Hattum, J. Morris & D. Hoffman 2, 28–47. University of Salford.
Nye, R. 2013. *How complement clauses distribute: Complementiser*-how *and the case against clause-type*. PhD, Ghent University.
Pantcheva, M. 2011. *Decomposing path: The nanosyntax of directional expressions*. PhD, University of Tromsø.
Patota, G. 2002. *Lineamenti di grammatica storica dell'italiano*. Bologna: Il Mulino.
Platzack, C. 1986. COMP, INFL, and Germanic word order. *Topics in Scandinavian syntax*, eds. L. Hellan & K. Koch Christensen, 185–234. Dordrecht: D. Reidel Publishing.
Rizzi, L. 1997. The fine structure of the left periphery. *Elements of grammar*, ed. L. Haegeman, 281–337. Dordrecht: Kluwer Academic Publishers.
Roberts, I. & Roussou, A. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.
Rocquet, A. 2013. *Splitting objects: A nanosyntactic account of direct object marking*. PhD, Ghent University.
Rooryck, J. 2000. *Configurations of sentential complementation: Perspectives from Romance languages*. London: Routledge.
Rosenbaum, P. S. 1967. Phrase structure principles of English complex sentence formation. *Journal of Linguistics* 3(1): 103–118.
Roussou, A. 2010. Selecting complementizers. *Lingua* 120(3): 582–603.
Sanfelici, E. & Poletto, C. 2014. On the nature of complementizers: Insights for Italian subject relative clauses. Paper presented at Going Romance 28, Lisbon.
Schrock, T.B. 2014. *A grammar of Ik (Icé-tód)*. PhD, University of Leiden.
Sočanac, T. 2011. Subjunctive in Serbian/Croatian. *Generative Grammar in Geneva* 7: 49–70.
Sportiche, D. 2011. French relative *qui*. *Linguistic Inquiry* 42: 83–124.
Starke, M. 2004. On the inexistence of specifiers and the nature of heads. *Structures and beyond*, ed. A. Belletti, 252–268. New York: Oxford University Press.
Starke, M. 2009. Nanosyntax: A short primer to a new approach to language. *Nordlyd* 36 (Special issue on Nanosyntax), 1–6.
Starke, M. 2011. Towards an elegant solution to language variation: Variation reduces to the size of lexically stored trees. LingBuzz/001183.
Starke, M. 2013. Auxiliaries and structural gaps: Current issues in Nanosyntax. Lecture series presented at CRISSP, Hogeschool-Universiteit Brussel.
Szabolcsi, A., Whang, J. & Zu, V. 2014. Quantifier words and their multi-functional(?) parts. *Language and Linguistics* 15(1): 115–155.
Taraldsen, K.T. 2009. The nanosyntax of Nguni noun class prefixes and concords. LingBuzz/000876.
Vanden Wyngaerd, G. 2014. Pointers and the ABA diagnostic. Paper presented at Nanosyntax workshop at SLE 47, Poznań, Poland.
Vangsnes, Ø.A. 2008. Decomposing manner *how* in colloquial Scandinavian. *Studia Linguistica* 62(1): 119–141.
Vangsnes, Ø.A. 2013. Syncretism and functional expansion in Germanic wh-expressions. *Language Sciences* 36: 47–65.

Vangsnes, Ø.A. 2014. Indexicals by nanosyntax: Wh and D items apart. Paper presented at Nanosyntax workshop at SLE 47, Poznań, Poland.
Wiklund, A.-L. 2008. Tense/mood/aspect-doubling. *Microvariation in syntactic doubling*, eds. S. Barbiers, O. Koeneman, M. Lekakou & M. der van Ham, 171–188. Bingley: Emerald Group.

Received October 22, 2015
Accepted May 5, 2017

Lena Baunaz
University of Zürich
Romanisches Seminar
Zürichbergstrasse 8
CH–8032 Zürich Switzerland
lena.baunaz@uzh.ch

Eric Lander
University of Gothenburg
Box 200
40530 Göteborg
Sweden
eric.lander@svenska.gu.se