

Lena Baunaz and Eric Lander

Cross-categorical Syncretism and Containment in Balkan and Slavic

Abstract: This paper presents syncretism patterns in the Balkan and Slavic languages between complementizers, (indeclinable) relativizers, and interrogative pronouns. Building on Baunaz and Lander (2017), we claim that complementizers – and the categories they are syncretic with – are internally complex and composed of syntactico-semantic features which are hierarchically ordered according to a functional sequence. The syncretism patterns of Balkan and Slavic are shown to be precisely paralleled by the syncretism data from Romance and Germanic, and are thus compatible with (and additional evidence for) the functional sequence proposed in Baunaz and Lander (2017). Yet, even though the syncretism patterns from Balkan and Slavic support our previous findings, some Slavic languages (Serbo-Croatian, Russian) present a problem in the form of an unexpected morphological containment relation. To account for this *Slavic containment puzzle*, we propose a finer-grained morphological decomposition of some of the items discussed, coupled with a relevant generalization concerning the internal structure of demonstratives.

Keywords: Slavic, syncretism, complementizers, demonstratives, morphological containment

1 Introduction

Syncretism is the phenomenon whereby multiple functions are covered by a single phonological form. For instance, the Latin plural case ending *-īs* has both dative and ablative functions. Rather than positing two separate case endings which happen to be homophonous (dative plural *-īs* vs. ablative plural *-īs*), one posits a single ending *-īs* which is syncretic between dative and ablative (see Baerman, Brown and Corbett 2005, Starke 2009, Caha 2009, among others).

In this paper we discuss a phenomenon which might be called *cross-categorical syncretism*. That is to say, we consider cases where there is a syncretism conflating what are normally considered to be distinct categories: demonstrative,

Lena Baunaz, (University of Zurich)

Eric Lander, (University of Gothenburg)

complementizer, relative pronoun, and interrogative pronoun. Cross-categorical syncretism is present in the (finite) complementizer systems of, for instance, English (*that* is a demonstrative, complementizer, and relative pronoun; see Roberts and Roussou 2003, Kayne 2008, Leu 2015) and French/Italian (*que* and *che* are complementizers, relative pronouns, and interrogative pronouns; see Sportiche 2011 for French, Manzini and Savoia 2003, 2011 for Italian, and Roussou 2010 for similar facts in Modern Greek). The items responsible for non-finite complementation in these languages, moreover, appear to involve a cross-categorical syncretism between complementizers and prepositions (French *à*, *de*, *pour*; English *for*). Cross-categorical syncretism also implicates the verbal domain, as in Akan [Niger-Congo] (*se* is the verb ‘say’, a quotative marker, and a simulative marker ‘like, as if’; see Lord 1993), Mandarin [Sinitic] (*shuō* is the verb ‘say’ and a quotative; see Chappell 2008), and Buru [Austronesian] (*fen(e)* is the verb ‘say’ and a quotative; see Klamer 2000), and even English (*like* is a verb, a quotative marker, and a simulative marker).

In previous work we established – primarily on the basis of facts from Germanic and Romance – a particular underlying functional sequence responsible for building the demonstrative, (nominal) complementizer, relative pronoun, and interrogative pronoun. In this paper we extend our approach to the Balkan and Slavic languages, with interesting consequences. On the one hand, the syncretism patterns from Balkan and Slavic support our previous findings, but on the other hand these languages present a problem in the form of an unexpected morphological containment relation.

The theoretical approach taken is nanosyntactic (Starke 2009, 2011; see also Baunaz et al. 2018), meaning that we consider these items to be complex (cf. Sanfelici and Poletto 2014, Leu 2015, among others), with a fine-grained internal structure consisting of multiple syntactic features. Following the cartographic maxim of ‘one feature – one head’ (Cinque and Rizzi 2008: 50), moreover, these features are merged as heads in a strict, universal hierarchical order (i.e. the functional sequence). One especially important aspect of nanosyntax is that it allows for multiple heads to be spelled out by a single morpheme, that is, phrasal spellout.

The outline of the paper is as follows. In Section 2 we provide some basic background on cross-categorical syncretism in Germanic and Romance and the functional sequence emerging out of the patterns observed. Section 3 is the empirical core of the paper: it presents the relevant syncretism patterns in the Balkan and Slavic languages between complementizer, (the indeclinable) relativizer, and interrogative pronoun. The syncretism patterns are seen to be perfectly compatible with (and thus further evidence for) the functional sequence seen in Section 2. In Section 4, we propose a finer-grained morphological decomposition of some of the items discussed, along with discussion of a relevant generalization

concerning the internal structure of demonstratives. In Section 5 we discuss an interesting puzzle of *morphological containment* which arises in Slavic, for which we provide an account. Section 6 concludes the paper.

2 Background: Syncretisms with the nominal complementizer

Baunaz and Lander (2017) discuss patterns of syncretism with the complementizer in Germanic and Romance (as well as some Balkan and Finno-Ugric, which we omit here). We observe that the declarative complementizer (Comp) in these languages often has the same morphophonological form as demonstrative (Dem), relative (Rel), and interrogative (Wh) pronouns (see work cited above), which we take to be cases of cross-categorical syncretism. The data are summarized in (1).

(1) Syncretisms with nominal complementizer (3 SG inanimate/neuter forms)

		DEM	COMP	REL	WH
North Gmc	<i>Swedish</i>	det	att	som	vad
	<i>English</i>	that	that	that	what
	<i>(Non-standard) English</i>	that	that	as	what
West Gmc	<i>Dutch</i>	dat	dat	dat	wat
	<i>German</i>	das	dass	das	was
	<i>Yiddish</i>	jenc	vos	vos	vos
			az	az	
Romance	<i>French</i>	ce(lui)	que	que	que
	<i>Italian</i>	quello	che	che	che
	<i>Spanish</i>	aquél	que	que	qué

Even though we did not perform detailed morphological decompositions of the forms in (1), we still assumed that Dem, Comp, Rel, and Wh elements have a fine-grained internal structure.

As seen in (1), syncretism targets only adjacent cells in the paradigm (as indicated by the shaded areas). The fact that non-adjacent cells are not syncretic is analyzed in terms of the *ABA theorem and can be accounted for by nanosyntactic

principles of spellout (Caha 2009: Section 2.3; see also Bobaljik 2007, 2012). For our purposes here the most important concept is that syncretism reflects structural adjacency, revealing which syntactic heads are merged next to each other in the functional sequence. In other words, the patterns in (1) necessitate a linear order of heads such that the functional layer Dem is next to Comp, which is next to Rel, which is next to Wh: Dem | Comp | Rel | Wh.

While syncretism facts can determine what the linear order of functional heads is in a functional sequence, they do not necessarily determine what the hierarchical order is (i.e. Dem > Comp > Rel > Wh vs. Wh > Rel > Comp > Dem). In Baunaz and Lander (2017) we propose a novel way of handling the hierarchy issue, namely identifying something we call the ‘nominal core’ and studying its behavior with regard to syncretism, a strategy we briefly discuss below. For now, though, observe that the first hierarchy – Dem > Comp > Rel > Wh – lines up with certain findings from (more traditional) cartographic work on the clausal spine (e.g. D > C > Rel in Cinque 2008) and thus this order should be preferred over the other one. In other words, the word-internal or morphological structure we are interested in can be seen to replicate structure at the higher clausal level.

We may assume (following Grimshaw 1991 and later work) that functional structure in the extended projection must be merged on top of a lexical element, in this case a noun (though for our purposes this noun is taken to be ‘lighter’ than a full lexical noun like *house*).

(2) Dem > Comp > Rel > Wh > ... *n*

In nanosyntax, structures are taken to be additive or cumulative. This means that the set of syntactic heads making up the Dem structure is a superset of the set of heads making up the Comp structure, that Comp is in turn a superset of Rel, and so on. The relevant structures are given in (3), with the more abstract labels A, B, C, and D for the heads in order to make the concept of cumulative structure clearer.

(3)	[D	[C	[B	[A	[<i>n</i>]]]]	=	Dem
		[C	[B	[A	[<i>n</i>]]]]	=	Comp
			[B	[A	[<i>n</i>]]]	=	Rel
				[A	[<i>n</i>]]]	=	Wh
					[<i>n</i>]]]	=	nominal core

The cumulative nature of structure is the key to accounting for the *ABA theorem (see Caha 2009: Section 2.3 in particular).

As for the head *n*, it is the smallest element of the structure. Features being cumulative, [*n*] is thus a subset of the structure of the Wh pronoun, which itself is a subset of Rel, and so on. This ‘nominal core’, contained in all of the structures in (3), essentially classifies the morphological items being built as nominal elements. Being the smallest structure that the fseq can build, this item is expected to be semantically vacuous (or at least, semantically light).

An interesting piece of evidence for the existence of the nominal core can be found in the Germanic and Romance languages. In English, for instance, interrogative/quantificational pronouns and (non-D-linked) demonstratives can be decomposed into at least two components: *whi-**ch*** (< Old Eng. *hwi-**lc***; cf. German *we-**lch***, Dutch *we-**lk***, Swedish *vi-**lk***), *ea-**ch*** (< Old Eng. *ǣ-**lc***), *su-**ch*** (< Old Eng. *swi-**lc***). The second component in each form here (i.e. *-(l)ch/-lk*) expresses something along the lines of ‘form’, which makes sense from a historical point of view since these morphemes descend from the Proto-Germanic noun **lik-* ‘body, form’ (see Leu 2015: §6.2.1 and references cited there). This is overt evidence for the light noun being embedded in the structure of the larger pronoun, i.e. *whi-**ch***_{FORM}. Similarly, Romance quantifiers are often built combining an overt operator with the semantically vacuous bound morpheme *-que/-che* (Fr. *quel-**que*** ‘some’, *cha-**que*** ‘each, every’; It. *qual-**che*** ‘some’ and *cias-**che***-*duno* ‘someone’), which is actually syncretic with (non-bound) Comp, Rel, and Wh, as seen in (4).

(4) Romance nominal cores

		DEM	COMP	REL	WH	<i>n</i>
Romance	<i>French</i>	ce	que	que	que	-que
	<i>Italian</i>	quello	che	che	che	-che

In other words, the bound morpheme *-che* is like Gmc. *-(l)ch/-lk* in being a relatively semantically vacuous element which is found in certain nominal environments (e.g. combined with independently built operators like Fr. *quel-*, *cha-*, It. *cias-*, *qual-*). Crucially the Romance nominal core participates in the syncretism patterns we are interested in and for this reason can be considered part of the functional sequence as in (2) and (3) above. The prosodic dependence and relative semantic vacuousness of Gmc. *-(l)ch/-lk* and Rom. *-que/-che* are two reasons to assign it a very small structure (see Cardinaletti and Starke 1999), and since this small structure is syncretic with the Wh-layer, we have evidence for the hierarchy Dem > Comp > Rel > Wh (i.e. the hierarchy with Wh placed at the structurally smaller end).

3 The core data: Syncretism patterns in Balkan and Slavic

We now extend the approach discussed above to the Balkan and Slavic data. We also make more specific claims about the precise functions implicated in our syncretism patterns. We observe for Balkan and Slavic that declarative complementizers used in (finite) *emotive factive* contexts^{1,2} (labeled Comp_{EF} in the table in (5) below) – that is, under predicates like ‘regret’, ‘be surprised’, ‘be happy’, ‘be sorry’, etc. – are often syncretic with the *indeclinable relativizer* (which we label Rvz instead of Rel ; note that relativizers are sometimes called *relative complementizers*). In some languages (Greek, Russian, Serbo-Croatian), moreover, the Comp/Rvz syncretism also includes the neuter singular Wh-pronoun ‘what’. The neuter singular Dem pronouns cited in (5) are distal ‘that’ unless otherwise noted.

(5) Syncretisms with the nominal complementizer in Balkan and Slavic³

		DEM_{PRO}	COMP_{EF}	RVZ	WH_{PRO}
<i>Modern Greek</i>		ekíno	pu	pu	tí
<i>Romanian</i>		acel	că	ce	ce
<i>Polish</i>		to	że	co	co
West Slavic				% że	
<i>Czech</i>		to	že	co	co
East Slavic	<i>Russian</i>	to	что	что	что

¹ With the exception of Yiddish (where *az* is the semi-factive and non-factive complementizer and *vos* the emotive factive complementizer; Taube 1994), in Germanic and Romance factivity with regard to complementizers is not overtly distinct (i.e. there is a syncretism, e.g. Eng. *that*). The languages discussed here often do make this distinction. In some languages there is a complementizer that can be used in both factive and non-factive contexts (MG *oti*, SC *da*, Bg. *če*, and Ma. *deka*). Certain items, however, are always factive when used (MG *pu*, SC *što*, Bg. *deto*, Ma. *što*). As seen in (5), our Balkan and Slavic data suggest that it is the emotive factive complementizer which participates in syncretism with the relativizer and wh-pronoun. For a more fine-grained account of these facts in terms of veridicality, see Baunaz (2016, under review).

² We also do not consider declarative ‘how’ complementizers like Ro. *cum*, Ru./Bg. *kak*, SC *kako*, and Cz./Po. *jak*, though it is interesting that these are syncretic with Rel and Wh adverbs meaning ‘how’. They also appear to trigger some kind of modality when used.

³ We would like to thank an anonymous reviewer for suggestions on improving our table.

(continued)

		DEM _{PRO}	COMP _{EF}	RVZ	WH _{PRO}
South Slavic	<i>Serbo-Croatian</i>	to	što	što	što
	<i>Bulgarian</i>	tova ‘this’	deto	deto	kakvo
	<i>Macedonian</i>	toa ‘that (close to hearer)’	što	što	što

3.1 The data in more detail

Modern Greek has two complementizers: *pu* and *oti*.⁴ *Pu* introduces epistemic and emotive factive-type of complements (6a, b), while *oti* introduces non-factive complements (6c). *Oti* may also introduce epistemic factive complements (6a), but not emotive factive complements (6b).

- (6) a. *Thimame pu/oti dhjavaze poli.* (Grk)
 remember.1SG that read.3SG much
 ‘I remember that he used to read a lot / I remember him reading a lot.’
 (Roussou 2010: 590, (17))
- b. *O Pavlos lipate pu/*oti i Roxani efije.*
 the Paul is.sad.3SG that the Roxanne left.3SG
 ‘Paul regrets that Roxanne left.’
 (adapted from Giannakidou 2009: 1886, (8))
- c. *O Pavlos ipe oti i Roxani efije.*
 the Paul said.3SG that the Roxanne left.3SG
 ‘Paul said that Roxanne left.’
 (Giannakidou 2009: 1886, (7))

⁴ In addition, Modern Greek displays *pos* ‘that’. *Oti* and *pos* vary freely: *pos* essentially replaces *oti* in everyday usage (Roussou 2000, 2006, 2012). Complementizer *pos* is syncretic with the relative pronoun *o-pos* ‘how’, the free relative *o-pos* ‘whichever way’ and with the *wh*-word *pos* ‘how’. Modern Greek also displays *na* under desiderative ‘wish’-type of verbs. The status of *na* is still debated, with some viewing it as a complementizer (Roussou 2010) and others as a mood particle (Giannakidou 2009, among others). Because *na* is restricted by tense and agreement – it must always, contrary to *oti/pu*, be adjacent to the verb, it can appear in main clauses, and it can co-occur with other complementizers (*pu*) (whereas *oti* and *pu* cannot co-occur) – we follow Giannakidou (2009) and claim that it is a mood particle.

Pu is syncretic with *Rvz*, as seen in (7).

- (7) *O fititi pu sinandises (ine filios mu)* (Grk)
 the student that met.2SG is friend mine
 ‘The student that you met (is my friend).’
 (Roussou 2010: 591, (18a))

Note that *pu* also looks identical to (and is in fact historically derived from) the locative adverb *pu* ‘where’ and the relative adverb *ó-pu* ‘where’. This intersection with locatives (also relevant for Polish and Bulgarian) can be analyzed in terms of syncretism as well, but for reasons of space we cannot discuss this complication here. The important thing to notice is that the kind of *wh*-item we compare across languages in this paper is the interrogative pronoun ‘what’, and thus in Modern Greek we must consider *tí* rather than *pu* under the *Wh*-column. In sum, then, *Comp_{EF}* in Greek is syncretic with *Rvz*, but not with *Dem* or *Wh*.

Romanian has one declarative complementizer, namely *că*. This item appears almost everywhere (but not under predicates selecting the subjunctive mood). In (8) we provide an emotive factive example using *că*.^{5, 6}

- (8) *Ion regretă că Maria e bolnavă.* (Rom)
 Ion regrets that M. is.IND sick
 ‘John regrets that Mary is sick.’

Romanian *că* is not syncretic with anything in the table above. However, there is a syncretism between *Rvz* and *Wh* in the form of *ce*, as illustrated in (9) (Grosu 1994, Bențea 2010, among others).⁷

- (9) a. *Am citit cartea ce a publicat-o*
 have.1SG read book.the that has published-CL.FEM.SG
Paul anul trecut. (Rom)
 Paul last year.
 ‘I read the book that Paul published last year.’
 (Bențea 2010: 175, (30b))

⁵ Romanian subjunctive clauses may be introduced by *ca* when an XP is topicalized. If no topic is present, *ca* is also absent. We do not discuss this here.

⁶ *De* may also be classified as a declarative complementizer in Romanian (see Hill 2002). However, not only does it cover the same ground as Fr. *de* and It. *di*, namely *non-finite* complement clauses, it also extends to interrogative and conditional contexts. Because of its broad usage we leave out *de* for the time being.

⁷ As a relativizer, *ce* is quite rare and can only be used with subject and direct object relatives (see Grosu 1994, Bențea 2010).

- b. **Ce** *ai* *auzit?*
 What have.2SG heard
 ‘What did you hear?’
 (Bențea 2010: 173, (25a))

Thus there is a Rvz/Wh syncretism in Romanian, an important fact in its own right as it establishes the adjacency of the Rvz and Wh layers.

The default complementizer in Polish is *że*, with an emotive factive example given in (10a). Hansen et al. (2016: 205–206) write that Comp *że* is historically derived “from the masculine form of the Proto-Slavonic interrogative pronoun *jbže* ‘which’ which is no longer in use in Polish.” This means that the Polish complementizer *że* is historically – but not synchronically – related to an interrogative pronoun, just like Modern Greek *pu*. Note also that Comp *że* can be suffixed with *-by* in irrealis contexts (which agrees in number and person with the subject). For some speakers *że* is syncretic with the relativizer *że* (10b). In this function it is important to note that Po. *że* is possible only in certain contexts, as noted by Murelli (2011: 195), namely when there is some nuance of consecutivity or finality (the relativizer by default being *co* ‘that’).⁸ In addition, Murelli (2011: 195) writes: “In South-Eastern Polish dialects the relative particle *że* is used not only in constructions with a consecutive nuance, but has generalized to an all-purpose (relative) particle.” This is illustrated in (10c).

- (10) a. *Maria jest zadowolona że wyjeżdżasz* (Pol)
 M. is happy that leave.2SG
 ‘Maria is happy that you’re leaving.’
- b. % *Takiego człowieka że-by spał z otwartymi*
 such man that-IRR sleep.PAST.MASC with open
oczami, jeszcze nie widziałem.
 eyes I.have yet.not seen
 ‘I still haven’t seen a man that slept/would sleep with his eyes open.’
- c. *Ten chłopak, że-smy go wczoraj spotkali.* (SE Pol)
 That lad that-we.are CL.M.ACC.3SG yesterday met
 ‘The lad we met yesterday.’
 (Laskowski 1991: 275, cited in Murelli 2011: 195, (4.113))

To sum up, Polish Comp_{EF} *że* is not syncretic with anything in the standard language, but is syncretic with the relativizer in southeastern Polish and in some

⁸ Unfortunately Murelli (2011) does not provide clear examples for (non-standard) Polish. Our informants are very reluctant to accept *że* as a relativizer, but one of our speakers provided us with (10b).

non-standard varieties. Note also that the default relativizer *co* ‘that’ is syncretic with the *wh*-pronoun *co* ‘what’.

There is one complementizer in Czech, namely *že*. It is the complementizer by default, but in (11) we once again provide the relevant context for our proposal, namely *že* under an emotive factive predicate.

- (11) *Maruška je šťastná, že Honza odešel.* (Cz)
 Mary.NOM is.3SG happy.FEM that John leave.PAST.MASC
 ‘Mary is happy that John left’

Že is not syncretic with any other items considered here.⁹ Just like in Polish, though, there is a syncretism between Rvz *co* ‘that’ and the Wh-pronoun *co* ‘what’.

The default complementizer in Russian is *čto*. In (12a) we provide an emotive factive example. As in Polish, the suffix *-by* (from auxiliary ‘be’) can appear in irrealis/subjunctive contexts (see Hansen et al. 2016). Contrary to Polish, though, *-by* does not agree in number and person with the subject in Russian. Moreover, *čto* is syncretic with Rvz (12b, c) and Wh (12d).

- (12) a. *Mne zhal’, čto ty obidel Ivana.* (Ru)
 1DAT.SG pity that you hurt Ivan
 ‘I regret (lit. it’s a pity to me) that you hurt John.’
 b. *Eti bol’shie kartiny, čto visiat na stene, privezli*
 these big pictures that are on wall have.been.brought
iz drugogo goroda.
 from another city
 ‘These big pictures that are on the wall have been brought from another city.’
 c. *On uvidel staryi dom, čto postroil ego ded.*
 he has.seen old house that had.built his grandfather.
 ‘He has seen the old house that his grandfather had built.’
 d. *Čto vy budete zakazyvat’?*
 what you.NOM.2PL will.AUX.2PL order
 ‘What would you like to order?’

Čto is partially syncretic (in a sense to be made more precise below) with the distal 3SG demonstrative *to*.

⁹ For some speakers, Czech *že* is syncretic with the relativizer that occurs in the same contexts as Romanian relative *că* and Polish relative *że*, that is, when the relative relation exhibits a nuance of consecutivity or finality (Murelli 2011: 195). Our Czech informants do not accept *že* in this context.

Like Greek, Serbo-Croatian has two complementizers: *da* and *što*. While *da*¹⁰ is the default complementizer, the use of *što* – interestingly for our purposes – is limited to appearing under emotive factive verbs (13a). Crucially, furthermore, Comp *što* is syncretic with both Rvz (13b) and Wh (13c). Note that regional variation as to the use of *što* or *šta* is found among SC speakers (T. Samardžić and T. Sočanac, p.c.).

- (13) a. *Žalim što si povrijedio Ivana.* (SC)
 regret.1SG that AUX.PAST.2SG hurt.PAST.PART John
 ‘I regret that you hurt John.’
- b. *Profesor što predaje istoriju ima veliki nos.*
 professor that teaches history have big nose
 ‘The professor who/that teaches history has a big nose.’
 (Mitrović 2012: 1, (1))
- c. *Što radish?*
 what do.2SG
 ‘What do you do?’

Thus SC Comp *što* is syncretic with both Rvz and Wh. In addition, just like Russian *čto*, SC *što* is partially syncretic with the distal 3SG Dem *to*.

Bulgarian exhibits two declarative complementizers: *če* and *deto*. Comp *če* appears everywhere, for instance in semi-factive contexts (14).¹¹

- (14) *Interesno e če tuk e zapazen*
 interesting be.PRES.3SG that here be.PRES.3SG stored
edinstvenijat original (Bulg)
 sole-DEF.M.NOM original
 ‘It’s interesting that the only original is stored here.’
 (Hansen et al. 2016: 212, (134))

¹⁰ There is a debate in the literature concerning the status of SC *da*. The general trend nowadays is that there are two homophonous items with the form *da*: declarative *da* vs. modal *da*. The distinction between the two has been established on the basis of (i) their distribution and (ii) their historical development. In particular modal *da* would be historically derived from 2nd and 3rd person singular imperatives of the verb *dati* ‘give’, namely **dadjъ* (Old Church Slavonic *daždъ*). The origin of ‘declarative’ *da* is unclear (though it might have been adverbialized from a demonstrative). The reader is referred to Todorović (2012) and Sočanac (2017) and references cited in these works for details.

¹¹ It is interesting for our purposes to note that in earlier stages of the language *če* had a relativization function (see Sonnenhauser 2015 for examples, as well as references cited there).

Under emotive factive verbs, both *če* and *deto* can be used (Krapova 2010), as shown in (15a).¹² The use of one or the other complementizer triggers a subtle change in meaning: the presupposition of the embedded clause gets somewhat stronger when *deto* is selected than when *če* is selected (see Baunaz 2016, under review). Comp_{EF} *deto* is syncretic with Rvz *deto* ‘that’, as shown in (15b).¹³

- (15) a. *Sážaljavam, če/deto ne možax da doйда.* (Bulg)
 regret.1SG that not could.1SG MOD come.1SG
 ‘I regret that I couldn’t come.’
 (Krapova 2009: 1240, (1a))
- b. *Tova e čovekāt, deto (go) snimax včera.*
 This is man.the that him.CL.ACC photographed.1SG yesterday
 ‘This is the man that I photographed yesterday.’
 (Krapova 2009: 1240, (1b))

To sum up, Bg. Comp_{EF} *deto* is syncretic with Rvz (but not Wh or Dem).

Finally, the Macedonian Comp_{EF} is *što* (16a), which is syncretic with Rvz (16b) and Wh (16c).

- (16) a. *Mi e milo što si otide.* (Mac)
 me.DAT is glad that AUX left
 ‘I’m glad that he (has) left.’
- b. *Profesorot što predava istorija ima golem nos.*
 professor.DEF that teaches history has big nose
 ‘The professor that teaches history has a big nose.’
- c. *Što e ova?*
 what be.3SG this.NEUT.SG
 ‘What is this?’
 (Tomić 2006: 419, (fn.2, (ii)))

The default complementizer in Macedonian is *deka*, as it appears with non-factive verbs and semi-factive verbs. As it is not the emotive factive complementizer, we do not include it in our table.

¹² Some speakers accept *deto* also with semi-factive verbs like ‘remember’. When that is the case, the same presupposition shift as the one described here is in order, i.e. the presupposition of the embedded clause is somewhat stronger with *deto* than with *če* (see Baunaz 2016, 2018 for more details).

¹³ Historically the relative *kādeto* is derived from the interrogative adverb *kāde* ‘where’ by adding the definite/demonstrative *-to* morpheme. *Deto* is synchronically (partially) syncretic with the relative pronoun *kādeto* ‘where, which, whom’.

All of the attested syncretisms in (5) are restricted to adjacent cells: Modern Greek Comp *pu*, Russian Comp *čto*, SC Comp *što*, Bulgarian Comp *deto*, and (non-standard) Polish Comp *że* are all syncretic with at least Rvz and sometimes also with Wh, but never only with Wh to the exclusion of Rvz. If we look a little bit closer at the tables in (1) and (5) we see that the linear ordering of Dem, Comp, Rvz, and Wh can only be the one given in (17), which captures the relevant adjacencies: Bulgarian (and some varieties of Polish) show that Comp and Rvz must be adjacent, and the Czech data demand that Rvz and Wh also be adjacent. Drawing from our previous work, we include (non-standard) English data to show that Dem and Comp need to be adjacent as well (18) (since none of the Balkan/Slavic data happens to show syncretism with Dem).

(17) Dem | Comp | Rvz | Wh

(18) Four crucial syncretism patterns from (1) and (5)

	DEM	COMP	RVZ	WH
Non-standard English	that	that	as	what
Bulgarian	tova	deto	deto	kakvo
% Polish	to	że	że	co
Czech	to	že	co	co

The linear order in (17) is the only one that can capture the data in (1) and (5) accurately. Any other ordering would disrupt this empirical reality: if Comp and Dem were not contiguous, then the English data would not be captured. If the ordering had been Dem | Comp | Wh | Rvz, then the Bulgarian and Polish data would not be captured either, since Comp and Rvz are syncretic in this language, and so on. In Baunaz and Lander (2017: §2.1), we came to the same conclusion, so we can straightforwardly say that the same ordering posited on the basis of data from Germanic and Romance also holds for Balkan and Slavic.

As in Romance, nominal cores can be identified in Slavic and Balkan. In particular, there is a nominal core in Serbo-Croatian, Russian, and Modern Greek that is syncretic across the Comp, Rvz, and Wh layers: SC *ne-što* ‘something’, *sva-šta* ‘everything’, *ni-šta* ‘nothing’, *bilo-šta* ‘anything’; Russian *čto-to* ‘something’ and *ne-čto* ‘something (specific)’. Similarly, Czech and Polish have a nominal core syncretic with Rvz and Wh (but not with Comp): Cz. *-co* ‘-thing’, as in *ně-co* ‘something’, and Po. *co-* as in *co-ś* ‘something’. Modern Greek can also form its quantifiers with a bound morpheme which is syncretic with Wh *tí* ‘what’ (but not with Comp *pu*). Quantifiers like *ká-ti* ‘something’ and *tí-pota* ‘anything’ overtly

display the nominal core *-ti-* ‘-thing’. Finally, Romanian also has a nominal core syncretic with Rvz and Wh (but not with Comp): *-ce-* as in *ce-va* ‘something’ or *ori-ce* ‘anything’.

We claim here that these bound morphemes are the Slavic and Balkan counterparts of Romance *-que/-che* and Gmc *-(l)ch/-lk* discussed earlier: they are semantically quite vacuous and are only found in combination with operators like ‘every-’, ‘some-’, etc. As such, these items in Balkan and Slavic present strong supporting evidence for the nominal core hypothesis and the reasoning (based on Cardinaletti and Starke 1999) that they are realizations of the lowest (i.e. smallest) bit of structure than can be built using our functional sequence.

- (19) Dem > Comp > Rvz > Wh > ... n
 [DCBA...n] [CBA...n] [BA...n] [A...n] [n]

Thus Balkan and Slavic confirm that the linear ordering Dem | Comp | Rvz | Wh can be assigned the hierarchical order Dem > Comp > Rvz > Wh (rather than Wh > Rvz > Comp > Dem) on the basis of syncretism with the nominal core.

3.2 An emerging puzzle

The tables in (1) and (5) differ in one crucial way. In (5) there are no syncretisms with Dem. Because of this one might wonder if Dem is even relevant to these particular languages and if we should not, perhaps, remove this column from the table completely. On the one hand, this would seriously undermine the universality of our fseq, something to be avoided on general principles. There is also empirical evidence to support keeping the Dem column for Balkan and Slavic: Comp, Rel, and Wh all contain the Dem element *to* (SC *š-to* and Ru. *č-to*). The fact that these items all make use of the same basic morphological ingredients when they are constructed is evidence that they belong to the same ‘paradigm’.

- (20)
- | | Dem | Comp | Rvz | Wh |
|----|-----------|-------------|-------------|-------------|
| Ru | to | č-to | č-to | č-to |
| SC | to | š-to | š-to | š-to |
- = [Comp/Rvz/Wh č- / š- [Dem to]] (?)

Though the morphological containment shown in (20) is evidence that the functional template Dem > Comp > Rvz > Wh also applies to Slavic, there is something unexpected about the relationship between Dem *to* and Comp/Rvz/Wh *čto/što*. According to our fseq, Dem is the most complex (i.e. the largest)

structure, with Comp being the next biggest structure, then Rvz, and finally Wh. However, the containment relation in (20) suggests the exact opposite, namely that Dem is structurally smaller than Comp, Rvz, and Wh, since it is overtly contained *within* these structures. Our fseq predicts instead that if Dem happens to be involved in a morphological containment relation, then the Dem structure should be the one containing Comp, Rvz, or Wh rather than the other way around.

In Section 4, we provide an account of this ‘Slavic containment puzzle’ (as we call it). We extend our analysis to Bulgarian, Polish, and Czech. Our analysis calls for further decomposition of the morphemes discussed so far, a possibility readily afforded to us by the nanosyntactic approach.

4 Decomposition, demonstratives, and definite markers

We now turn to further decomposition of the Germanic, Romance, Balkan, and Slavic data, showing how some of these data fit the approach developed so far, while others are at first glance problematic.

4.1 Decomposing further in Germanic

In Baunaz and Lander (2017) we mention that many of the forms in the table in (1) are obviously (at least) bimorphemic (e.g. Eng. *th-at*) but that syncretism can be studied without full-fledged decomposition down to the smallest level. While this is true, in this paper we in fact continue to decompose the forms in (1). For instance, if we take English, at least the demonstrative and *wh-* items are straightforwardly decomposable.

- | | | | | |
|------|-----|-------|---|---|
| (21) | Dem | th-at | = | /ð-æt/ |
| | Wh | wh-at | = | / (h)w-ʌt/ (North America), / (h)w-ɒt/ (UK) |

The *th-*prefix in the demonstrative form can be put on a par with other such prefixes in Germanic.

- | | | | | | |
|------|---------|-----|------|---|-----------|
| (22) | Swedish | Dem | det | = | /d-e:(t)/ |
| | Dutch | Dem | d-at | = | /d-at/ |
| | German | Dem | d-as | = | /d-as/ |

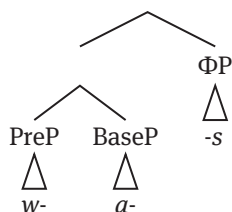
More specifically this prefix has been argued to be an instantiation of the definite article (Def) appearing as a subcomponent of the demonstrative (see Déchaine and Wiltschko 2002, Kayne and Pollock 2010, Leu 2015 (and previous work), Roehrs 2010, among others).

The *wh*-operator is instantiated by *wh*-morphology in many languages. Consider the Germanic forms in (23).

- (23) Swedish Wh v-ad = /v-ɑ:(d)/
 Dutch Wh w-at = /v-ɑt/
 German Wh w-as = /v-as/
 Yiddish Wh v-os = /v-ɔs/

In some of these languages, moreover, the remaining element when the prefix is removed (Du. *-at*, Ger. *-as*, etc.) is the same as the element which remains when the D-prefix is removed. In German (and Slavic, as we see below) there is also evidence for a third component in the structure, namely agreement (in German the strong adjective ending for neuter nominative/accusative, namely *-(e)s*), which we simply label Φ P here.

- (24) Trimorphic structure for German *was*

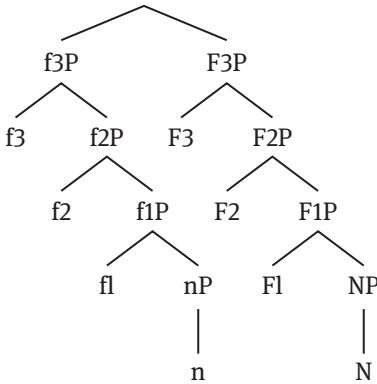


All in all, then, we have a tripartite structure, where the leftmost position (PreP, here corresponding to the morpheme *w-*) is the locus of our fseq (Dem > Comp > Rvz > Wh). The other constituents (BaseP and Φ P) are also assumed to have complex internal structure. Note that we are making crucial use of phrasal spell-out here (a ubiquitous trait of nanosyntactic proposals).

For NP in particular, we note that Čaha and Pantcheva (2014) and Taraldsen (2018) have both proposed that prefixes are independent constituents merged as complex specifiers outside of the main extended projection. They also reason that the functional heads making up such prefixes, moreover, must be merged on top of a lexical category, and that this lexical category is a classifier-like noun (distinct from the main lexical N). This is abstractly shown in (25), where the

constituent f3P is the prefix structure (PreP in (24)), with the classifier-like noun *n* at the bottom, and the constituent F3P is the base structure (BaseP in (24)), with lexical N at the bottom.

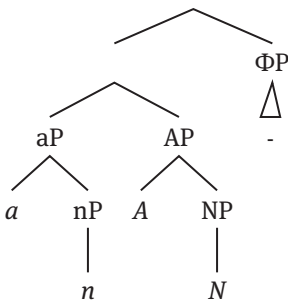
(25)



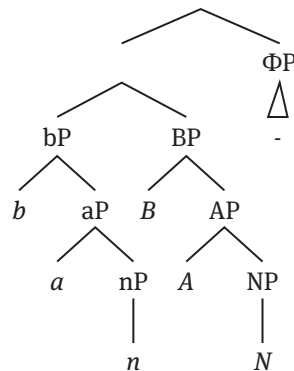
Both Caha and Pantcheva (2014) and Taraldsen (2018) provide interesting evidence from Bantu in favor of such a binominal structure. Without going into detail, we note that our structure closely matches the structure already argued for by these authors on independent grounds (except for Φ P, of course, which is an additional constituent we have proposed for the structures we consider here).

We remind the reader that structures are cumulative, and that this also applies to our double structure made up of PreP and BaseP. This is illustrated in (26).

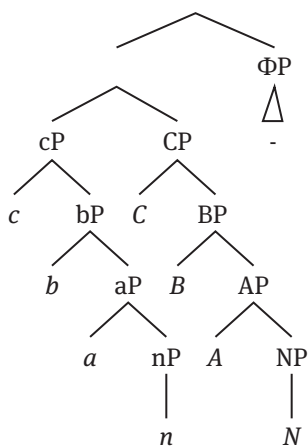
(26) a. Wh



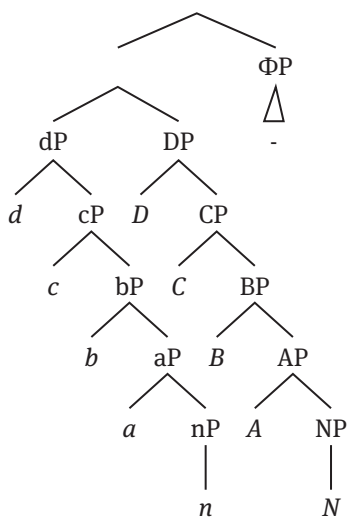
b. Rel



c. Comp



d. Dem



With two fseqs (one for PreP and one for BaseP), we expect that two distinct syncretism patterns should also be possible. Indeed, this is exactly what we observe with, for instance, Dutch Dem/Comp/Rvz /d-/ vs. Wh /v-/ in PreP, as opposed to total syncretism of /-at/ in the BaseP sequence, as shown in (27).

<p>(27) <i>PreP</i></p> <p>[d [c [b [a [n]]]] => d-</p> <p style="padding-left: 2em;">[c [b [a [n]]] => d-</p> <p style="padding-left: 4em;">[b [a [n]] => d-</p> <p style="padding-left: 6em;">[a [n]] => v-</p>	<p style="text-align: right;"><i>BaseP</i></p> <p>[D [C [B [A [N]]]] => -at</p> <p style="padding-left: 2em;">[C [B [A [N]]] => -at</p> <p style="padding-left: 4em;">[B [A [N]]] => -at</p> <p style="padding-left: 6em;">[A [N]] => -at</p>
---	---

Another example is English, where PreP has the same basic pattern as in Dutch (Dem/Comp/Rvz *th-* /ð-/ vs. Wh *wh-* /(h)w-/), while BaseP is realized as Dem/Comp/Rvz /-æt/ vs. Wh /-ʌt/.

<p>(28) <i>PreP</i></p> <p>[d [c [b [a [n]]]] => ð-</p> <p style="padding-left: 2em;">[c [b [a [n]]] => ð-</p> <p style="padding-left: 4em;">[b [a [n]] => ð-</p> <p style="padding-left: 6em;">[a [n]] => (h)w-</p>	<p style="text-align: right;"><i>BaseP</i></p> <p>[D [C [B [A [N]]]] => -æt</p> <p style="padding-left: 2em;">[C [B [A [N]]] => -æt</p> <p style="padding-left: 4em;">[B [A [N]]] => -æt</p> <p style="padding-left: 6em;">[A [N]] => -ʌt</p>
---	--

4.2 Containment of Def within Dem

Importantly for our purposes here, PreP for demonstratives is in many languages an instantiation of the definite article (Def), e.g. Eng. *th(e)* in *th-at*. That is to say, Def is an integral component in the internal structure of Dem. There is in fact an abundance of crosslinguistic evidence outside of Germanic for the claim that Def is contained within Dem. In Klallam [Salish], for instance, demonstrative forms contain a distinct morpheme encoding definiteness, the suffix *-nił*. This is seen in (29) (where PROX and DIST refer to distance from the speaker).

- (29) a. *tiəw' -nił* (Klallam)
 PROX-Def
 = Dem 'this'
- b. *təsə -nił*
 DIST-Def
 = Dem 'that'
- (Montler 2007: 411)

In Tahitian [Austronesian], moreover, the definite article is *te*. Once we decompose the long *ē* of medial and distal *tē-* into *ee*,¹⁴ we see that the definite marker *te* can be isolated in each of Tahitian's three demonstratives.

¹⁴ “[Long vowels] are the same in quality as [short vowels], but are pronounced as if they were a double vowel. [t ... a long vowel is regarded as a sequence of two identical vowels.]” (Tryon 1970: 2).

- (30) a. **te-ie** (Tahitian)
 Def-PROX
 = Dem ‘this’
- b. *tēna*
te-ena
 Def-MED
 = Dem ‘that (near the person addressed)’
- c. *tēra*
te-era
 Def-DIST
 = Dem ‘that (not near the speakers)’
 (Tryon 1970: 9, 24)

In Koromfe [Niger-Congo], an item which “is very similar to the definite article in English” (Rennison 1997: 81) can be appended with a deictic marker, giving the form “more demonstrative or deictic force” and making it “more akin to the English demonstratives *this* and *that*” (Rennison 1997: 234, 81).¹⁵

- (31) a. HU.SG HU.PL (Koromfe)
hoŋ-go **bɛŋ-gɛ**
 Def-DX Def-DX
 = Dem = Dem
- b. Non-HU.SG Non-HU.PL
koŋ-go **hɛŋ-gɛ**
 Def-DX Def-DX
 = Dem = Dem
 (Rennison 1997: 234–235)

In Romanian, the generalization concerning the containment of Def within Dem is straightforwardly instantiated. As seen in (32), the entire Romanian demonstrative paradigm shows morphological containment of the definite article.

- (32) Romanian (Savu and Bican-Miclescu 2012)
 a. Dem ‘that’

	M.SG	F.SG	M.PL	F.PL
NOM/ACC	a.ˈtʃel	a.ˈtʃea	a.ˈtʃej	a.ˈtʃele
GEN/DAT	a.ˈtʃe.luʃ	a.ˈtʃe.leʃ	a.ˈtʃe.lor	a.ˈtʃe.lor

¹⁵ Though, as Rennison notes, Koromfe has additional dedicated deictic items with even more “deictic force” (Rennison 1997: 234, 81).

b. Def ‘the’

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

The definite article is not a prefix in Romanian, but this can be accounted for in terms of movement.

Consider now the Italian forms for ‘that’ in (33).

(33)	<i>Dem</i>	<i>Def</i>	
M.SG	quel- lo [quel]	lo il	(+ word-initial sC- or z-) (+ word-initial other C-)
F.SG	quel- la	la	
M.PL	que- gli que- i	gli i	(PL of <i>lo</i>) (PL of <i>il</i>)
F.PL	quel- le	le	

The Italian forms follow our basic generalization, though with some minor complications: (i) there is no M.SG *que-il*, for instance (*quel* being used instead, perhaps simply for phonological reasons) and (ii) the Def element is not actually a prefix but looks more like a suffix (which we assume can be accounted for in terms of movement). Nevertheless, the generalization discussed above regarding Def being morphologically contained within Dem very clearly holds for Italian, with both singular and plural Def being overtly contained within the Dem forms.¹⁶

As a final set of evidence for our generalization regarding Def-containment within Dem, consider Greek. The N.SG Dem forms *ekíno* ‘that’, *tuto* ‘this’, *auto* ‘this’ are likely candidates for morphological decomposition, especially considering the history of these pronouns (*ekíno* < minimally trimorphemic PIE **h₁e-ki-eno* and *tuto* < minimally bimorphemic PIE **h₂u-tod*; see Johansson and Carling 2015: §6.1). The crucial question, though, is whether or not Def is an integral morphological ingredient for Dem. We propose that the answer is yes, but that Def is not contained within the morphological structure of Dem, rather it is contained at the phrasal level of DemP. That is, it is a well known fact that Def is required in the presence of Dem in Greek, as seen in (34).

¹⁶ Note that BaseP for the Italian demonstratives is spelled out as *que(l)- /kwe/*; the smaller structures (Comp, Rvz, Wh) are all *che /ke/*, on the other hand, which we assume to be a portmanteau morpheme spelling out the constituent containing both PreP and BaseP (see the structure in (36) below).

- (34) a. *ekíno to pédi* (MG)
 Dem.DIST Def child
 ‘that child’
 b. *auto to spiti*
 Dem.PROX Def house
 ‘this house’
 (Holton et al. 2003: 93, 19–20)

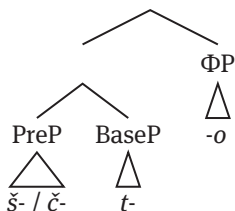
The ‘stacking’ of Dem and Def is also found in languages as diverse as Lakota (Ingham 2003: 90), Hungarian (Kenesei et al. 1997: 95), Koyra Chiini (Heath 1998: 61), Yucatec Mayan (Janssen 2004: 986), Welsh (Dryer 1992: 121), and D(r)ehu (Dryer 1992: 121).

In sum, the claim that Def is a building block in the construction of Dem has support from a diverse set of crosslinguistic morphological evidence.

5 The Slavic containment puzzle

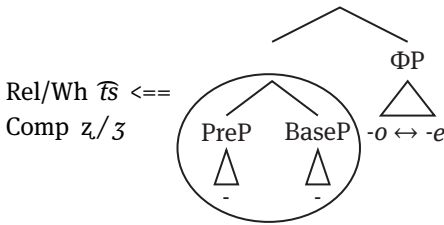
We now turn to Slavic, specifically Serbo-Croatian and Russian. It is clear that SC *što* and Ru. *čto* are easily decomposable into *š-t-o* and *č-t-o*. The first consonant derives historically from palatalization of the *wh-* morpheme *k-* before a front vowel (see the proto-forms below), the second consonant *t-* is the demonstrative root, and *-o* is the neuter singular inflection (i.e. Proto-Balto-Slavic **ki-to* > Proto-Slavic **čь-to* ‘what’) (Boban Arsenijević, Tomislav Sočanac, p.c.). Thus we can say that SC *š-* and Ru. *č-* (/ʃ/ and /ʧ/, respectively) correspond to PreP, *t-* to BaseP, and *-o* to Φ P.

- (35) Comp/Rel/Wh in Serbo-Croatian and Russian



This kind of decomposition can be extended to Polish and Czech as well, if we simply assume that *c-*, spelling the affricate /tʃ/ in both languages, is a portmanteau of PreP and BaseP (again making crucial use of phrasal spellout).

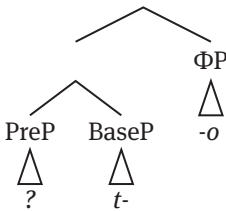
(36) Comp/Rel/Wh-structures in Polish and Czech



Interestingly, *-o* (in *t-o* and *c-o*) has the allophone *-e* after ‘soft’ consonants like Po. *ż-* /*z̥*/ and Cz. *ž-* /*ʒ̥*/, so also the complementizers Po. *ż-e* and Cz. *ž-e* have exactly the same basic structure, with the same neuter singular ending, i.e. ΦP .¹⁷

The demonstrative form in this languages, on the other hand, is simply *t-o*, which when compared to (35) above is clearly missing a realization for PreP.

(37) Dem in Serbo-Croatian and Russian



That is to say, Dem in Serbo-Croatian and Russian is only bimorphemic while Comp/Rvz/Wh is trimorphemic; in other words, Dem is smaller than Comp/Rvz/Wh. This is unexpected since our sequence predicts that Dem should be the larger structure. Moreover, Dem *to* is contained within Comp/Rel/Wh [*š-*[*to*]] and [*č-*[*to*]], which is the opposite of what we expect. That is, if there is an overt containment to be observed, then Dem is expected to contain Comp, Rvz, and Wh.

On the other hand, this containment puzzle does not make an appearance in Macedonian (Dem *toa* vs. Comp/Rvz/Wh *što*¹⁸) or Bulgarian (Dem *tova* vs. Comp/Rvz *deto* vs. Wh *kāde*). In these languages the Dem form is not contained within the Comp, Rvz, or Wh structures. In fact, as expected, the (so-called ‘neutral’) definite marker *-to* in Macedonian is contained within Dem [[*to*]-*a*], as well as

¹⁷ Thanks to a reviewer for discussion of the data and suggestions.

¹⁸ *Deka* or *oti* is the default (broadly, non-factive) complementizer, but here we consider the factive complementizer *što* (Tomić 2006: 458, fn.86). *Deka*, interestingly, is also a relativizer in Macedonian.

within Rvz/Wh [š-*to*]. Bulgarian Def *-to* is similarly contained within Comp/Rvz [*de*-*to*]. Thus these languages do not pose a problem for the containment relation predicted by our fseq.

6 Solving the puzzle

We can account for the containment puzzle with a very simple observation: most Slavic languages do not have definite articles (see, among others, Bošković 2005, 2008, 2009, 2010 on the NP/DP parameter), Serbo-Croatian and Russian being perfect cases in point. In fact, the only Slavic languages with definite articles are Macedonian and Bulgarian (where the neuter singular definite marker is *-to* in both). Because Serbo-Croatian and Russian do not have definite articles, their demonstratives do not have the definite article (PreP) available to them. Thus Dem ends up being smaller than Comp/Rel/Wh, the latter forms having access to a PreP structure (š- and č-) since these PrePs are not dependent on being a DP language.

(38) ‘Containment puzzle’ in Serbo-Croatian and Russian

	PreP	BaseP	ΦP
Wh	š-/č- (<i>wh</i> -marker)	<i>t-</i>	<i>-o</i>
Rvz	š-/č- (<i>rvz</i> -marker)	<i>t-</i>	<i>-o</i>
Comp	š-/č- (EF-marker)	<i>t-</i>	<i>-o</i>
Dem	definite article (not available)	<i>t-</i>	<i>-o</i>

Indeed, the relevant Dem forms in Bulgarian and Macedonian contain Def (i.e. Bg. *to-va* ‘this’, Ma. *to-j* ‘that (close to hearer)’, both in the neuter singular), just like the data from Germanic, Romance, Klallam, Tahitian, and Koromfe above.

This accounts for the clear contrast between Serbo-Croatian and Russian on the one hand, where the containment is problematic, and Macedonian and Bulgarian on the other, where the containment is not problematic. As for the other languages at stake, namely Polish and Czech, the containment puzzle is still present, yet in a slightly less obvious guise. Above we mentioned that both Comp (Po. *ż-e*, Cz. *ž-e*) and Rvz/Wh *c-o* can be considered to be underlyingly tripartite structures (they are historically exactly equivalent to SC *što* and Ru. *čto*) as long as the initial consonant is analyzed as a portmanteau morpheme. However, the Dem form is again *to* in Polish and Czech (again historically exactly equivalent to

SC/Ru. *to*). Though we could always assume that *t-* is a portmanteau once again and in this way try to dissolve the containment puzzle, such an analysis might be considered suspect on the grounds that the voiceless stop *t-* is phonologically simpler than the palatalized consonants in the Comp, Rvz, and Wh forms (i.e. Po. /z/ and Cz. /z/, and Po./Cz. /ʃs/), which betray a more complex history; more important, though, is the crosslinguistic evidence from closely related languages like Serbo-Croatian and Russian, where it is clear that *t-o* is not trimorphemic when compared to *š-t-o* and *č-t-o*.

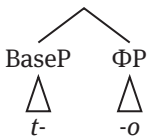
(39) ‘Containment puzzle’ in Serbo-Croatian and Russian

	PreP	BaseP	ΦP
Wh		<i>c-</i>	<i>-o</i>
Rvz		<i>c-</i>	<i>-o</i>
Comp		<i>ž- / ž-</i>	<i>-e</i>
Dem	definite article (not available)	<i>t-</i>	<i>-o</i>

In other words the containment puzzle is relevant for Polish and Czech as well, but the puzzle can be solved in the same way as it was for Serbo-Croatian and Russian, since Polish and Czech are also languages lacking definite articles.

The basic structure for Slavic NP languages is given in (40), where PreP is not part of the structure.

(40) Dem in SC/Ru (as well as Po/Cz)



Even though Serbo-Croatian and Russian do not have definite articles, we also do not want to assert that the semantic features responsible for definiteness are completely absent in these languages. Indeed, Slavic demonstratives must still be “disguised definite descriptions” (Bennett 1978: 22), as in other languages. Rather, we believe that these definiteness features are packaged with items elsewhere in the grammatical system (that is, not in the form of distinct, overt Def morphemes found in DP languages), but we leave the specifics of this hypothesis for future research.

7 Concluding remarks

Our paper is based on Baunaz and Lander's (2017) nanosyntactic analysis of complementizers in Romance and Germanic. In nanosyntax, syncretism reflects structural adjacency, that is, forms that are syncretic are taken to reflect a specific ordering of syntactic heads merged in a functional sequence. Here we have shown that (nominal) complementizers in the Balkan and Slavic languages participate in systematic syncretisms with Dem, Rel, and Wh pronouns, verifying previous findings about these patterns in Germanic and Romance and empirically strengthening the proposal that these items are internally complex and built using a single functional sequence. Moreover, there is evidence for a structurally deficient 'light' noun at the bottom of this fseq, which overtly appears as a bound morpheme in the internal morphological structure of quantifiers in these languages.

Following the discussion of the syncretism data, we have proposed to further decompose the forms under observation into at least a tripartite structure made up of a prefix (PreP), a base (BaseP), and an inflectional ending (ΦP). Most interesting is the generalization that demonstratives (in DP languages) have a definite article acting as their PreP, which is backed up by evidence from a typologically diverse range of languages.

We have also looked at a potential problem for our approach, namely the 'containment puzzle' exhibited by languages like Serbo-Croatian and Russian. In these languages it is clear that Comp, Rvz, and Wh *što/čto* show morphological containment of Dem (*to*), even though we expect to find the opposite (Dem containing Comp, Rvz and Wh). We account for this puzzle by referring to the well known fact that, with the exception of Bulgarian and Macedonian, Slavic languages lack definite articles. Thus, the fact that PreP is missing in languages like Serbo-Croatian and Russian can be attributed to independent reasons about the availability of a grammaticalized definite article; indeed, in Bulgarian and Macedonian, which are the only Slavic languages to have definite articles, there is no containment puzzle to speak of.

Acknowledgements: We would like to thank Boban Arsenijevic, Anamaria Bentea, Joanna Blochowiak, Pavel Caha, Goljihana Kachaeva, Iliyana Krapova, Teodora Radeva-Bork, Gabriela Soare, Tanja Samardzic, Tomislav Sočanac, Branimir Stankovic and Bartosz Wiland for their help with the data. We would like to specially thank Iliyana Krapova for showing interest in our work, as well as an anonymous reviewer for interesting suggestions and discussions. Lena Baunaz's research is supported by FWO project 2009–Odysseus-Haegeman-G091409. Eric Lander's research was funded by BOF grant 01D30311 awarded by Ghent University and FWO project 2009–Odysseus-Haegeman-G091409.

References

- Baunaz, Lena. 2016. Deconstructing complementizers in Serbo-Croatian, Modern Greek and Bulgarian. *Proceedings of NELS* 46, 1: 69–77.
- Baunaz, Lena and Eric Lander. 2017. Syncretisms with the nominal complementizer. *Studia Linguistica* 71: 1–34.
- Baunaz, Lena. 2018. Decomposing complementizers: The fseq of Serbo-Croatian, Modern Greek, Bulgarian and French complementizers. In Lena Baunaz, Karen De Clercq, Liliane Haegeman and Eric Lander (eds), *Exploring Nanosyntax*. Oxford Studies in Comparative syntax, New York: Oxford University Press. 149–179.
- Baunaz, Lena, Karen De Clercq, Liliane Haegeman, and Eric Lander (eds). 2018. *Exploring Nanosyntax*. Oxford Studies in Comparative Syntax. New York: Oxford University Press.
- Baerman, Matthew, Dunstan Brown, Greville G. Corbett. 2005. *The syntax-morphology interface: A study of syncretism*. Cambridge: Cambridge University Press.
- Bennett, Michael. 1978. Demonstratives and indexicals in Montague Grammar. *Synthese* 39: 1–80.
- Bențea, Anamaria. 2010. On restrictive relatives in Romanian: Towards a head-raising analysis. *GG@G* 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.
- Bošković, Željko. 2005. On the locality of left branch extraction and the structure of NP. *Studia Linguistica* 59: 1–45.
- Bošković, Željko. 2008. What will you have, DP or NP? *Proceedings of NELS* 37: 101–114.
- Bošković, Željko. 2009. More on the no-DP analysis of article-less languages. *Studia Linguistica* 2: 187–204.
- Bošković, Željko. 2010. On NPs and clauses. Ms., University of Connecticut.
- Caha, Pavel. 2009. *The nanosyntax of case*. PhD dissertation, University of Tromsø.
- Caha, Pavel and Marina Pantcheva. 2014. Locatives in Shona and Luganda. Ms., University of Tromsø.
- Cardinaletti, Anna and Michal Starke. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In Henk van Riemsdijk (ed.), *Clitics in the Languages of Europe*: 145–233. Berlin: Mouton De Gruyter.
- Chappell, Hilary. 2008. Variation in the grammaticalization of complementizers from *verba dicendi* in Sinitic languages. *Linguistic Typology* 12, 1: 45–98.
- Cinque, Guglielmo. 2008. More on the indefinite character of the head of restrictive relatives. *Rivista di Grammatica Generativa* 33: 3–24.
- Cinque, Guglielmo and Luigi Rizzi. 2008. The cartography of syntactic structures. *CISCL Working Papers on Language and Cognition* 2: 43–59.
- Déchaine, Rose-Marie and Martina Wiltschko. 2002. Decomposing pronouns. *Linguistic Inquiry* 33: 409–443.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68: 81–138.
- Giannakidou, Anastasia. 2009. The dependency of the subjunctive revisited: Temporal semantics and polarity. [Special issue on Mood] *Lingua* 119: 1883–1908.
- Grimshaw, Jane. 1991. Extended projection. Ms., Brandeis University.
- Grosu, Alexander. 1994. *Three studies in Locality and Case*. London: Routledge.
- Hansen, Björn, Alexander Letuchiy, and Izabela Błaszczuk. 2016. Complementizers in Slavonic (Russian, Polish, and Bulgarian). In Kasper Boye and Petar Kehayov (eds.), *Semantic*

- functions of complementizers in European languages*. Berlin and Boston: De Gruyter Mouton. 175–223.
- Heath, Jeffrey. 1998. *A grammar of Koyra Chiini*. Berlin: De Gruyter Mouton.
- Hill, Virginia. 2002. Complementizer Phrases (CP) in Romanian. *Rivista di Linguistica* 14,2: 223–248.
- Holton, David, Peter Mackridge, Irene Philippaki-Warbuton. 2004. *Greek: An essential grammar of the modern language*. London: Routledge.
- Ingham, Bruce. 2003. *Lakota*. München: LINCOM.
- Janssen, Theo A. J. M. 2004. Deixis and reference. In Geert Booij, Christian Lehmann, Joachim Mudgan, Stavros Skopeteas (eds), *Morphology: An international handbook on inflection and word-formation*, Vol. 2: 983–997. Berlin and New York: De Gruyter.
- Johansson, Niklas and Gerd Carling. 2015. The de-iconization and rebuilding of iconicity in spatial deixis. *Acta Linguistica hafniensia: International Journal of Linguistics* 47, 1: 4–32.
- Kayne, Richard. 2008. Why isn't this a complementizer? Ms., New York University.
- Kayne, Richard and Jean-Yves Pollock. 2010. Notes on French and English demonstratives. In Jan-Wouter Zwart and Mark de Vries (eds), *Structure preserved: Studies in syntax for Jan Koster*: 215–228. Amsterdam: John Benjamins.
- Kenesei, István, Robert Michael Vago, Anna Fenyvesi. 1997. *Hungarian*. London: Routledge.
- Klamer, Marian. 2000. How report verbs become quote markers and complementisers. *Lingua* 110: 69–98.
- Krapova, Iliyana. 2010. Bulgarian relative and factive clauses with an invariant complementizer. *Lingua* 120: 1240–1272.
- Leu, Tom. 2015. *The architecture of determiners*. Oxford and New York: Oxford University Press.
- Lord, Carol. 1993. *Historical change in serial verb constructions*. Amsterdam and Philadelphia: John Benjamins.
- Manzini, Maria Rita and Leonardo Maria Savoia. 2003. The nature of complementizers. *Rivista di Grammatica Generativa* 28: 87–110.
- Manzini, Maria Rita. and Leonardo Maria Savoia. 2011. *Grammatical categories*. Cambridge: Cambridge University Press.
- Mitrović, Ivana. 2012. Relative clauses in Serbian. Handout for paper presented at Syntax Lab, University of Maryland, 11 September.
- Montler, Timothy. 2007. Klallam demonstratives. *University of Columbia Working Papers in Linguistics* 20: 409–425.
- Murelli, Adriano. 2011. *Relative constructions in European non-standard varieties*. Berlin and Boston: De Gruyter Mouton.
- Rennison, John. R. 1997. *Koromfe*. New York: Routledge.
- Roberts, Ian and Anna Roussou. 2003. *Syntactic change: A Minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.
- Roehrs, Dorian. 2010. Demonstrative-reinforcer constructions. *The Journal of Comparative Germanic Linguistics* 13, 3: 225–268.
- Roussou, Anna. 2000. On the left periphery: Modal particles and complementisers. *Journal of Greek Linguistics* 1: 65–94.
- Roussou, Anna. 2006. *Simpliromatiki diktes [Complementizers]*. Athens: Patakis.
- Roussou, Anna. 2010. Selecting complementizers. [Special issue *Exploring the left periphery*] *Lingua* 120: 582–603.

- Roussou, Anna. 2012. Complements, relatives and nominal properties. Paper presented at *GIST 5: Generalizing relative strategies*. Ghent University, 22–23 March.
- Sanfelici, Emanuela and Cecilia Poletto. 2014. On the nature of complementizers : Insights for Italian subject relative clauses. Paper presented at *Going Romance 28*. Lisbon, 4–6 December.
- Savu, Carmen and 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*. PhD dissertation, University of Geneva.
- Sonnenhauser, Barbara. 2015. Functionalising syntactic variance: Declarative complementation with *kako* and *če* in 17th to 19th century Balkan Slavic. *Wiener Slavistisches Jahrbuch* 3: 41–72.
- Sportiche, Dominique. 2011. French relative *qui*. *Linguistic Inquiry* 42: 83–124.
- Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. [Special issue on Nanosyntax] *Nordlyd, Tromsø University working papers on language linguistics* 36: 1–6. <http://septentrio.uit.no/index.php/nordlyd/>.
- Starke, Michal. 2011. Towards an elegant solution to language variation: Variation reduces to the size of lexically stored trees. *LingBuzz*/001183.
- Taraldsen, Tarald. 2018. Spanning vs. constituent lexicalization: The case of portmanteau prefixes. In Lena Baunaz, Karen De Clercq, Lilliane Haegeman and Eric Lander (eds), *Exploring Nanosyntax*. Oxford Studies in Comparative syntax, New York: Oxford University Press. 88–107.
- Taube, Moshe. 1994. On factivity, emotivity and choice of conjunction in Yiddish. *Studies in Language* 18, 1: 113–125.
- Todorović, Nataša. 2012. *The subjunctive and indicative da-complements in Serbian: A syntactic-semantic approach*. PhD dissertation, University of Illinois at Chicago.
- Tomić, Olga Mišeska. 2006. *The Balkan Sprachbund morpho-syntactic features*. Dordrecht: Springer.
- Tryon, Darell Trevor. 1970. *Conversational Tahitian*. Canberra: Australian National University Press.