

Andrea D. Sims and Brian D. Joseph

# Morphology versus Syntax in the Balkan Verbal Complex

AU: Since the note citation is not allowed in chapter title, we have set this note as unnumbered note in COP. Please confirm

**Abstract:** Various Balkan languages have a string of material called here the “verbal complex”, in which a verb occurs with various markers for tense, modality, negation, and argument structure. We examine here this verbal complex with regard to its status as a syntactic element or a morphological element. First, we carefully outline the theoretical basis for determining the status of a given entity and we then argue that the verbal complexes display different degrees of morphologization in the different languages. Albanian and Greek show the highest degree of morphologization of the verbal complex, with Macedonian close to them in this regard. Bulgarian shows a lesser degree of morphologization than Macedonian, making for an interesting split within East South Slavic, and ~~Bosnian-Croatian-Serbian~~ Serbian shows an even lesser degree. We argue further that certain aspects of the verbal complex, especially in the languages with the greatest morphologization, represent contact-related convergence, and draw from this a general claim about the role of surface structure in language contact.

**Keywords:** Albanian, Greek, Macedonian, Bulgarian, ~~Bosnian-Croatian-Serbian~~, verbal complex, particle, morphologization, language contact, surface structure

## 1 Introduction

Most of the languages in the Balkans have a string of material that can be called the “verbal complex” consisting of a verb and various associated elements – what might be termed “particles” for want of a better characterization at this point – marking tense, modality, negation, and argument structure. Examples from a few such languages are given in (1), with dialectal material given for Greek and Macedonian to maximize comparability with the other languages.<sup>1</sup>

---

<sup>1</sup> In this paper, Modern Greek is transcribed with broad IPA. Cyrillic for all relevant Slavic languages is transliterated according to the scientific system.

---

**Note:** We thank Iliyana Krapova for her comments on this paper, and Victor Friedman for consultation about Macedonian.

---

**Andrea D. Sims and Brian D. Joseph**, The Ohio State University

- (1) a. *S' do të ja jep.* (Albanian)  
 NEG FUT SBJV 3SG.DAT/3SG.ACC give.1SG  
 'I will not give it to him.'
- b. *Nu o să îl dau.* (Daco-Romanian)  
 NEG FUT SBJV 3SG.DAT/3SG.ACC give.1SG  
 'I will not give it to him.'
- c. *De ðe na tu to ðóso.* (dialectal Greek)  
 NEG FUT SBJV 3SG.GEN 3SG.ACC give.1SG  
 'I will not give it to him.'
- d. *Ne ќе da mu go davam.* (dialectal Macedonian)  
 NEG FUT SBJV 3SG.DAT 3SG.ACC give.1SG  
 'I will not give it to him.'
- e. *Neću da mu ga dam.* (Serbian)  
 NEG.FUT.1SG SBJV 3SG.DAT 3SG.ACC give.1SG  
 'I will not give it to him.'

Elements in these languages seem to line up in the same order, generalizable as in (2)<sup>2</sup>:

- (2) NEG – TNS – MOOD – IO – DO – V

This pattern gives the appearance of a template-like order for the relevant elements and suggests convergence as part of the broader Balkan Sprachbund, the well-known result of a long period of sustained and intense contact among the speakers of various languages leading to striking similarities in structure and form in these languages. The similarities in the verbal complex are schematized in (2), although not all elements necessarily occur in any given sentence.

While there are various descriptive issues to be resolved with the verbal complex in each language, for instance regarding the elements that can and cannot occur in it, we approach this construct from a theoretically oriented analytic perspective. In particular, we ask whether the elements that make up the verbal complex are morphological objects or syntactic objects, and accordingly whether the verbal complex itself is a word-level unit or instead is a phrase-level unit. It is important to question the status of the verbal complex in part because surface word order can obscure structural differences. To take a single example

<sup>2</sup> Abbreviations in this paper follow the Leipzig Glossing Rules: <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>; accessed August 9, 2016.

from a different phenomenon, but one that is well understood, Rudin (1988) shows that while multiple WH-fronted structures in Bulgarian (e.g. *Koj kakvo vižda?* lit. ‘Who what sees?’) and what was then called Serbo-Croatian (e.g. *Ko šta vidi?* lit. ‘Who what sees?’) seem superficially to be parallel, they nonetheless reflect a structural difference that is revealed by their syntactic behavior when they occur in embedded clauses. In Bulgarian, all of the WH-words in a multiple question must be moved out of an embedded clause. By contrast, for many Serbo-Croatian speakers, only one WH-word can be extracted from an embedded clause, and the other must remain in the lower clause.<sup>3</sup> In terms of Rudin’s theoretical framework, this means that all WH-words are in SpecCP in Bulgarian, but only one WH-word can inhabit SpecCP in Serbo-Croatian. A surface similarity between Bulgarian and Serbo-Croatian thus hides a deep structural difference. To the extent that situations of this sort are not uncommon, it is manifestly obvious that surface strings cannot be taken as having any deep reality.

Additionally, once we scratch the surface, it quickly becomes clear that the internal structuring of the Balkan verbal complex differs from one language to another, despite the surface-similar pattern identified in (1) and (2). As we show below, the verbal complex is parallel to multiple WH-fronting in the sense that surface-parallel ordering that is shared among languages nonetheless corresponds to different structures. At the same time, the verbal complex differs in that it reveals not so much different underlying syntactic structures as differences in the morphological versus syntactic status of the elements. As happens with other features of the Balkan Sprachbund, there are piecemeal commonalities and differences from one language to another, but the most interesting dimension of difference from a theoretical standpoint has to do with the extent of morphologization of elements. The question of whether the verbal complex is a morphological or syntactic object thus turns out to have significance, since there are differences among the verbal complexes in the various languages, and some of the differences point to a fundamental divide between those languages in which the verbal complex is substantially morphologized, and those in which it is morphologized less or not at all.

In examining the status of the verbal complex, one of our goals is to justify and explore this construct as a topic of relevance to the Balkan Sprachbund. The verbal complex highlights an important generalization about processes of grammatical convergence in language contact situations, namely that it is surface-oriented. We are not the first to draw attention to surface similarities as being significant to

---

<sup>3</sup> Rudin (1988: 453f.) notes that there is some speaker-by-speaker variation in this regard. Some Serbo-Croatian speakers do accept multiple WH-fronted structures, parallel to the Bulgarian pattern.

the Balkan Sprachbund. Most famously, Kopitar's (1829) description of Albanian, Balkan Romance ('Wallachian') and Bulgarian as being three lexicons with one grammar was based not on deep structural aspects of the grammatical structure, but on surface parallelisms. Our paper is thus in the vein of previous work arguing that the very fact of surface similarities across languages in contact, but with deep structural differences, is evidence of the surface-oriented nature of language contact. However, evidence of morphologization in the languages of the central Balkans additionally raises interesting questions about how language contact and syntactic borrowing intersect with diachronic processes of morphologization. Thus we explore here the theoretical issues raised by the fact that the verbal complex exists at the intersection of morphology and syntax.

## 2 The relationship between morphology and syntax

The question of whether the verbal complex in each of the various Balkan languages constitutes a word-level unit or a phrase-level unit can only be answered in the context of a model of the morphology-syntax interface. Of course, how morphology and syntax intersect is a major source of debate, with proposals ranging all the way from the claim that they are non-distinct (with morphology usually subsumed to syntax) to the claim that there are fully distinct morphological and syntactic components, each operating according to its own principles. We cannot here rehash the history of thinking on the issue and we simply adopt a position of the latter type. In particular, we assume the framework of inferential-realization morphology and a lexicalist approach to the morphology-syntax interface.<sup>4</sup>

In inferential-realizational models (e.g. Paradigm Function Morphology (Stump 2001) or Network Morphology (Brown and Hippisley 2012)), the combination of a lexeme and morphosyntactic values licenses rules that perform morphophonological operations on bases, such as affixation. This means that the classical notion of a morpheme as a lexically listed bundle of form and meaning has no status. Such models have the advantage that the meaning of a word need not be exactly the sum of the meanings of its parts. The parts may underspecify, overspecify, or even mismatch the meaning of the whole, and even radical violations of form-meaning isomorphism can be handled easily. Moreover, inferential-realizational models are paradigmatic in the sense that word-forms convey

---

<sup>4</sup> See Stump (2001) for a detailed justification of inferential-realizational morphology, also called Word-and-Paradigm morphology.

meaning by virtue of paradigmatic contrast with other forms of the same lexeme. Thus for example, the English noun form *cat* is interpreted as singular not because it has some zero morpheme that adds the morphosyntactic value SINGULAR, but because it contrasts with the plural form *cats*. This means that inferential-realizational models are not committed to a concatenative approach, and in fact can accommodate a wide range of morphophonological operations on bases.

Inferential-realizationalism is consistent with a modular grammar architecture that includes an autonomous component for (inflectional) morphology operating according to principles that are at least partly distinct from the principles governing syntax. Inferential-realizational models also tend to accept some version of the Lexical Integrity Principle (Bresnan and Mchombo 1995), also sometimes termed the Principle of Morphology-free Syntax (Zwicky 1990, 1992). In principle, such a model offers a clear definition of what it means for a construction to be ‘morphologized’. At the same time, various phenomena pose a challenge to a strict separation of morphology and syntax (and thus to the lexicalist position), requiring more careful thought about the nature of the morphology-syntax interface.

Our goal in this section is to consider how such models answer questions about what constitutes word-like or phrase-like behavior, and how they interpret the empirical properties of constructions. While it is beyond the scope of this article to give a fully formalized model, we summarize arguments that (mostly) operate from the perspective of inferential-realizational morphology. We focus on clitics and periphrasis as phenomena that have received significant theoretical attention exactly because they exist at the morphology-syntax interface and are thus informative about it. Both kinds of pattern are also central to the Balkan verbal complex.

## 2.1 Some problems with the morphology = synthesis equation

It is fairly uncontroversial that synthetic expression is a matter of morphology. In fact, when linguists talk about ‘morphologization’ as a diachronic process of language change, they tend to mean the development of a unit that was a free syntactic atom into one that is affixally bound to a stem. Morphology and morphologization are thus equated with synthesis and the development of synthetic expression, respectively. However, once we scratch the surface it becomes obvious that this equation is too simple. Despite a widespread assumption that morphological structure is coextensive with synthetic expression, such a characterization is problematic. Consider, for instance, Zwicky

and Pullum (1983)'s well-known criteria for distinguishing between clitics and affixes<sup>5</sup>:

- A. affixes show high selectivity with regard to their stem; clitics may show low selectivity with regard to their host;
- B. affixed words are relatively more likely to exhibit arbitrary gaps in their inflectional paradigms (i.e. arbitrary restrictions on the combination of a stem and set of morphosyntactic values); clitics are relatively less likely to exhibit arbitrary gaps in host-clitic combinations;
- C. affixed words are more likely to show morphophonological idiosyncrasies (e.g. affixes can trigger or undergo allomorphy that is not purely phonological in nature); clitics are less likely to exhibit such idiosyncrasies;
- D. affixed words are more likely to show semantic idiosyncrasies (non-compositionality); clitic-host combinations are more likely to be compositional;
- E. syntactic rules may take affixed words as their domain; they cannot take clitic groups as their domain;
- F. affixes cannot attach 'outside' of (further from the root than) clitics; clitics must attach outside of affixes.

Zwicky and Pullum (hereafter, Z&P), along with many researchers before and after them, equate clitics with independent morphosyntactic words and affixes with subparts of words: "... [W]ord-clitic combinability is largely governed by SYNTACTIC considerations. The conditions governing the combinability of stems with affixes are of quite a different sort: they are MORPHOLOGICAL and/or LEXICAL in character..." (Zwicky and Pullum 1983: 503, emphasis original). A logical and common interpretation is thus that Z&P's criteria are diagnostics of syntactic vs. morphological objects.

At the same time, grammatical particles can exhibit mixed properties. In Serbian the feminine accusative singular weak pronoun, normally *je*, and the 3rd person singular auxiliary, also unstressed *je*, are both clearly second position clitics. Their placement (as part of a clitic cluster) is syntactically and prosodically determined and exhibits promiscuous host selection (criterion A). Nonetheless, when these two clitics occur adjacently (accusative preceding the 3sg auxiliary in the clitic cluster), the accusative clitic surfaces as *ju*. The clitic combination thus displays morphophonological idiosyncrasy (criterion C) – allomorphy in the

---

<sup>5</sup> Some of the criteria are diagnostic only in one direction. For instance, the existence of paradigmatic gaps is indicative of affix status, but the lack of such gaps is not necessarily indicative of clitic status, since gaps are relatively rare also in inflectional affixation. Nonetheless, these are the most widely cited criteria when it comes to diagnosing whether a given grammatical element is an affix or a clitic.

accusative clitic occurs in the context of the 3sg auxiliary clitic. This example, along with numerous others, shows that we cannot escape the conclusion that there are formatives that are placed with respect to syntactic phrases that nonetheless exhibit some of the properties of affixes according to the criteria above.

This lack of consistent results when applying the diagnostic criteria has led to further descriptive subdivisions. For instance, Sadler (1997) distinguishes syntactic clitics from phrasal affixation and edge inflection. In her terminology, syntactic clitics are syntactic terminals and form semantically transparent and syntactically accessible constructions with their host.<sup>6</sup> By contrast, phrasal affixes and edge inflection exhibit the syntactic properties of promiscuous host selection and syntactic placement, but the morphophonological properties of an affix. Spencer and Luís (2012) divide clitics further, arguing for a distinction between phrasal affixation and edge inflection. In phrasal affixation, a morphologically generated clitic or clitic cluster is placed syntactically with respect to a phrasal host. (The Serbian allomorphy *je je* → *ju je* qualifies these formatives as phrasal affixes.) In edge inflection, an inflected word-form consisting of both a host and clitic is selected in a given syntactic context (i.e., when the host sits at the edge of a phrase). They summarize a convincing example from Samvelian (2007) of the morphological behavior of edge inflection in Persian *ezafe* constructions. The point here is that a binary distinction into clitics and affixes is insufficient because grammatical formatives may exhibit some traits of each and different combinations of empirical properties occur. (See Aikhenvald (2002) and Spencer and Luís (2012) for a survey.) This highlights that the question of whether something is an affix or a clitic may be descriptively useful, but is ultimately incoherent in theoretical terms. The important issue has to do not with classification, but with determining what the properties of a formative are, and how the elements of grammar architecture interact to produce those properties.

So returning to Z&P's criteria, and the division between syntactic and morphological objects that they are often taken to imply, one thing we can observe is that the criteria themselves are not uniform in what they diagnose. Property A tends to reflect whether a grammatical particle attaches at the lexical or phrasal level. Properties B through D essentially boil down to the claim that affixes show a higher degree of idiosyncratic behavior along a number of dimensions whereas clitics show a greater degree of freedom and regularity. This kind of idiosyncrasy has long been interpreted as indicating composition in the lexicon, rather than the syntax (Chomsky 1970). Properties E and F diagnose objects that are subject

---

<sup>6</sup> However, syntactic clitics do not necessarily motivate an analysis as syntactic functional heads. Sadler argues at length that Welsh pronominal object clitics are syntactic clitics, but nonetheless are best analyzed as morphologically generated objects.

to lexical integrity. Within the criteria there is thus a difference between those that diagnose synthetic objects (criteria A, E-F), and those that diagnose composition in the lexicon or morphological component (B-D). If synthetic objects were always morphological, *and vice versa*, then the formula that equates clitics with syntactic objects and affixes with morphological objects would be justified. However, phrasal affixation and edge inflection throw this equation into doubt, because they tend to exhibit the idiosyncratic properties of morphological objects but do not necessarily form synthetic objects with their hosts.

To the extent that we define affixes as combining with stems to form synthetic objects, Z&P's criteria B through D are consistent with both affixation and morphologically generated clitic + host combinations. As Sadler (1997: 4) notes, these criteria are unable to distinguish between the two, and are thus insufficient by themselves to diagnose affixation in the traditional sense. However, we argue that these criteria are sufficient to diagnose morphological objects. Constructions may exhibit some morphological properties without being fully canonical affixes, highlighting that these notions cannot be conflated. In the following sections we consider a broader notion of what it means for an object of the grammar to be morphological, an understanding that extends beyond synthesis.

## 2.2 Lexicalism and the notion 'morphological object'

Ackerman et al. (2011) consider in detail the nature of the correspondence between synthetic objects and morphological objects in the context of a lexicalist model. The heart of lexicalism is what they call the Principle of Morphological Integrity: "Syntactic mechanisms neither make reference to a word form's proper subparts nor are able to create new word forms in constituent structure" (326). In other words, syntax has no access to the internal structure of synthetic objects.<sup>7</sup> However, other common tenets of lexicalism are to some degree independent issues, including the Principle of Unary Expression: "In syntax, a lexeme is uniformly expressed as a single morphophonologically integrated and syntactically atomic word form" (326). They argue that the Principle of Unary Expression is not inherent to lexicalism. In other words, there is no conceptual requirement that morphological realizations of lexemes be synthetic objects and syntactically atomic in the sense that is relevant for the Principle of Morphological Integrity. Modular architecture and a distinct morphological component are possible without assuming that synthetic objects are the only output. (See also Ackerman

---

7 This is essentially a (re)formulation of the Lexical Integrity Principle (see above, Section 2).



and LeSourd (1997) for a similar argument.) This raises the possibility of (partly) morphological analyses of phenomena like clitics and periphrasis.

## 2.3 Clitics as morphology

A number of arguments have been put forward against purely syntactic accounts of clitic cluster exponence and placement.<sup>8</sup> Here we highlight only a few but see Anderson (1992, 2005) and Spencer and Luís (2012) for details and additional arguments.

In an early account, Simpson and Withgott (1986) consider the problem of determining which word-formation happens in the syntax, and which in the lexicon, following the direction set by Chomsky (1970). Their paper is best known for criteria they propose as distinguishing templatic from layered (i.e. hierarchical) morphology, but their exemplification of templatic morphology centers on pronominal clitic clusters in Warumungu (a Pama-Nyungan language of Australia) and other languages. They argue that pronominal clitic clusters in Warumungu exhibit the properties of template morphology and that templates are fundamentally inconsistent with word-formation within the syntactic component because their internal structure is non-hierarchical and limited to the linearized ordering of elements.<sup>9</sup> They therefore take templatic structure as indicating that the cluster is generated in the lexicon (which is to say, in the morphological component), and conclude that the Warumungu clitic cluster is inserted into syntax as a single lexical entry.

An additional argument for this position comes from the fact that clitics do not always exhibit the expected properties of independent syntactic elements. Legendre (2001a) observes that clitic clusters in South Slavic (and other languages) are ‘syntactically inert’, which is to say, they are not available to be manipulated by syntax and in general do not behave as expected if clitics are

---

<sup>8</sup> At issue here are primarily special clitics. Following the terminology of Zwicky (1977), special clitics are clitics that do not have the same syntactic distribution as corresponding full lexical items or items belonging to the same syntactic category. Some aspects of their syntactic behavior thus require ‘special’ principles. These are contrasted with ‘simple’ clitics, which pose no special problems for the morphology-syntax interface. Phrasal affixes and edge inflection are both kinds of special clitics, although the notion of special clitic is broader than these.

<sup>9</sup> Even earlier, Perlmutter (1971) observes that in Spanish, the elements in a clitic cluster are ordered according to a template, leading him to propose a surface filter to account for clitic ordering.

syntactic heads. For example, Macedonian permits subject-verb inversion in questions, as shown in (3).<sup>10</sup>

- (3) a. *Katica ja čita knjiga.*  
           Katica 3SG.F.ACC read.3SG.PRS book.DEF  
           ‘Katica is reading the book.’
- b. *Ja čita li Katica knjiga?*  
       3SG.F.ACC read.3SG.PRS Q Katica book.DEF  
       ‘Is Katica reading the book?’

As in (3b), the subject normally inverts with the first verbal head (ignoring *li*, which is a clitic; its properties are discussed in Section 4.1 below.) However, when there is an auxiliary clitic, the subject follows the entire auxiliary + lexical verb complex (and any intervening pronominal clitics), not only the auxiliary (e.g. future *ќе*), as shown in (4) (example from Legendre 2001a: 247).

- (4) *Ќе ti ja dade li Penka knjiga?*  
       FUT 2SG.DAT 3SG.ACC give.3SG Q Penka book.DEF  
       ‘Will Penka give you the book?’  
       (\*Ќе Penka ti ja dade li knjiga?)

Legendre argues that “... head movement (to a projection higher than VP) operates as if the auxiliary were not present in the structure. The absence of subject-aux inversion does not make sense if... Macedonian *ќе* head[s] [a] functional project of [its] own...” (247). Analyzing *ќе* as a syntactic functional head thus makes incorrect predictions about syntactic behavior. The data instead suggest that it is part of a larger lexical structure and thereby not visible to the syntax.

In a somewhat similar vein, Anderson (2005: 113) identifies an issue related to the type of Wackernagel clitic sometimes termed ‘second word’ (2W) clitics, in which the clitics occur after the first word of the domain (as opposed to ‘second phrase’ clitics, which occur after the first phrasal constituent). In Bulgarian, for instance, definite determiners are prosodically enclitic to the first word in the NP, whether it is a noun or modifier (5).

<sup>10</sup> Sentence-initial *dali* without inversion is the more usual way of forming yes-no questions in Macedonian, though the use of *li* is possible; *li* is more characteristic of Bulgarian (for relevant discussion, see Englund 1977).

- (5) *mnogo=to interesni knigi*  
 many=DEF interesting books  
 ‘the many interesting books’  
 (cf. *knigi=te* ‘the books’)

Most syntactic accounts of 2W placement rely on displacement of a word from a phrase to a syntactic position above the clitic; see, for instance, Bošković (2001)’s analysis of 2W clitics in Serbo-Croatian. As Anderson (2005: 113) points out with reference to Bulgarian determiners, this kind of analysis creates two theory-internal problems:

On the syntactic view, the syntax generates the clitic in the head D position within a DP. A subsequent operation of displacement must then raise precisely the first word of the embedded NP to SpecDP... The syntactic account is motivated by a theory-internal assumption that the syntax must be the locus of description for such facts, so it is perhaps appropriate that it raises some general theory-internal problems... One such problem is the fact that the displacement... crucially involves a single word rather than a complete phrase. As such, it must be the kind of displacement known as ‘Head Movement’, rather than normal phrasal movement. But the target of the displacement in this case is not a head position, but phrasal: SpecDP. To the extent Head Movement is assumed to have properties distinct from those of ordinary phrasal movement, this presents a conceptual anomaly. Secondly, we can ask what the motivation is for the displacement... Apparently, this is driven only by the needs of the clitic determiner (its presumed prosodic requirements), and not by those of the word that moves. But within at least one version of the sort of theory that is at issue here, movement is only supposed to be driven by the needs of the item that moves, rather than by the resulting configuration.

Thus, Macedonian subject-auxiliary inversion and Bulgarian determiner clitics show that the syntactic properties of clitics can run afoul of theory-internal syntactic principles. Ultimately, we argue that these problems are largely a byproduct of the assumption that clitics are syntactic elements in their own right, rather than parts of larger lexical structures. While much of the focus of the (generative syntactic) literature has been on whether clitic clusters are syntactically or phonologically placed, this framing of the debate partly misses the point. Much of previous work has failed to notice that while clitic clusters are placed relative to syntactic phrases (possibly within prosodic constraints), they may also exhibit the properties of morphological objects (internally, or in combination with a host).

A morphological approach to clitic cluster exponence offers other benefits as well. Sadler (1997) discusses an interesting case of blocking between clitics and full noun phrases. In Welsh, a pronominal object clitic can optionally be doubled by a referentially identical pronoun. However, the clitic cannot double a full noun phrase in the same way. Working in Lexical Functional Grammar, Sadler argues that the clitic + host combination forms a small construction, which is to say, an  $X^0$  (lexical) category that contains adjoined  $X^0$  daughters. She argues that this

lexical/morphological status explains the blocking effects: the clitic + host, as a morphological construction, serves to block a full syntactic phrase for the object argument (along the lines of the proposal by Andrews (1990) that morphology blocks syntax). Since the blocking effect is independent of the basic motivation for proposing a small construction analysis, it comes ‘for free’ in her analysis. Sadler thus offers one way to formalize the idea that clitic + host constructions can form lexical/morphological units that consist of multiple syntactic atoms.

Bonami and Boyé (2007) approach French pronominal clitics from a different theoretical perspective (the inferential-realizational theory Paradigm Function Morphology), but likewise utilize morphological architecture to capture distributional facts surrounding clitics. They argue that French pronominal clitics should be handled as morphological objects, based in part on morphophonological fusion among clitics and conditioning on form (e.g. the object clitic is dropped in the context of dative). The clitics thus exhibit morphophonological idiosyncrasy (Z&P’s criterion C). Most interestingly, however, there are also restrictions on clitic combinability (e.g. reflexive and non-reflexive clitics cannot combine). Bonami and Boyé note that pre-existing mechanisms within inferential-realizational morphology for morphosyntactic feature licensing allow these feature co-occurrence restrictions to be defined in a natural way. To the extent that handling the co-occurrence restrictions in the syntax requires ad hoc principles, this argues for integrating the French pronominal clitics into the system of inflectional exponence.

In summary, clitic-host combinations often exhibit properties that make them anomalous if clitics are syntactic functional heads, but the same properties make clitics fundamentally similar to affixes. Moreover, a morphological analysis of at least some clitics can offer benefits in both general ways (blocking effects) and ways that are specific to the inferential-realizational framework (using independently needed feature-licensing mechanisms to capture clitic co-occurrence restrictions). These arguments and others lead us to the conclusion that at least some clitics are best analyzed as morphological objects.

## 2.4 Periphrasis as morphology


Finally, before moving on to the main data, we briefly turn to periphrasis. Periphrasis is the phenomenon in which multiple syntactic atoms collectively constitute a grammatical form of a lexeme. To take a widely discussed example, in Latin verbs the passive is sometimes expressed synthetically, as is the perfect tense. However, the intersection of perfect and passive is realized by a periphrastic construction: a past participle form of the lexical verb that inflects for gender and number (e.g. *laudatus* ‘praise.MSG’) plus a form of *ESSE* ‘be’ that inflects for

tense, person and number (e.g. *est* ‘be.3SG.PRES’). Each component part is an independent syntactic atom, as evidenced for example by the fact that they can be separated. However, only collectively do they express the perfect passive (*laudatus est*). Like clitics, periphrastic constructions have wide-ranging empirical properties (Bonami and Samvelian 2009).

There has been much debate about whether periphrastic constructions belong to the morphology or the syntax. Some approaches have sought to locate periphrastic constructions solely within the syntactic domain, with complementarity between synthetic and periphrastic forms treated as the result of blocking between the morphological and syntactic components (Ackema and Neeleman 2001; Andrews 1990; Bresnan 2001; Kiparsky 2005; Poser 1992). Inferential-realizational models, however, treat periphrastic constructions directly as exponents of inflectional paradigm cells, based primarily on the fact that periphrases are in complementary distribution with synthetic forms and convey morphosyntactic values that otherwise receive synthetic realization (Ackerman and Stump 2004; Sadler and Spencer 2000; Spencer 2001, 2003). Work within this framework has thus generally focused on the morphological aspects of periphrasis.

As with clitics, several arguments against purely syntactic accounts of periphrasis have been raised. Some relate to the internal logic of syntactic analyses (Ackerman and Stump 2004; Ackerman et al. 2011; Börjars et al. 1997). Others are rooted in the logic of inferential-realizational theories: a morphological account of periphrasis “... allows one to maintain a basic assumption of realizational morphology – that every well-formed morphosyntactic property set is available for morphological realization” (Ackerman et al. 2011: 335). Here we present just a sample.

One major issue has to do with the fact that periphrastic constructions are often not semantically compositional. Popova (2010: 174) gives the following example of non-compositionality in the so-called ‘inferential’ construction in Bulgarian. The construction can occur in different tenses, including with present time reference (6a) or with past time reference (6b).

- (6) a. *V*  *ta* *šte* (*da*) *e*, *kăde* *drugade*.  
 in house.DEF ŠTE (that) be.PRS.3SG where else  
 ‘He must be in the house, where else could he be?’
- b. *Šte* (*da*) *go* *e* *napisala* *pismoto*.  
 ŠTE (that) 3SG.M.ACC BE.PRS.3SG write.PTCP.SG.F letter.DEF  
 ‘She must have written the letter.’

The inferential construction is probably a reinterpretation of the future tense construction; the latter is formed with the future tense marker *šte*. “The inferential

construction has a modal meaning, expressing a supposition on behalf of the speaker... The future tense construction has inherent in it the meaning of predication or supposition and it is easy to see how this might be emphasized” (Popova 2010: 174–175). However, and importantly here, *šte* in the inferential construction is not inherently associated with futurity, as evidenced by its use with both present and past time reference in (6). Given that *šte* is undoubtedly a future marker in future tense constructions, we must either assume that there are two independent *šte* formatives, or we must assume that constructions need not be semantically compositional. Since, to the best of our knowledge, the two *šte* exhibit identical syntactic clitic properties – including different host selection/ placement than other clitics (see Section 4.2) – the latter analysis is preferable. Examples of this sort are far from rare.

Non-compositionality motivates an inferential-realizational morphological approach to periphrasis. Spencer (2001) notes that in Slavic languages, several kinds of form-meaning mismatch occur in periphrastic verb constructions, including cumulation, multiple exponence, empty morphemes, deponency and zero exponence. For example in Serbian (and some other Slavic languages), the periphrastic past tense construction (e.g. *napisala sam* ‘I wrote’) is built with a form of ‘be’ (e.g. *sam*) that is itself a *present* tense form. The observation that periphrases exhibit semantic non-compositionality and the same kinds of form-meaning mismatches as synthetic forms shows that it can be just as hard to assign grammatical functions to the individual components of a periphrastic construction as it is to assign grammatical functions to the individual morphemes of a word. In this respect, the arguments for treating periphrasis as morphological are the same as those that motivate an inferential-realizational model over an incremental and/or lexical one in general (see Stump 2001).

A different kind of argument comes from what Spencer (2001, 2003) calls the ‘underexhaustivity’ of the paradigm. Underexhaustivity describes the situation when a component element of a periphrastic construction lacks the full set of forms implied by the set of morphosyntactic values and their combinations. For example, the Serbian auxiliary derived from HTETI ‘want’ has only those forms that are used to compose the future tense ((*ho*)ću, (*ho*)ćeš, etc.), and lacks, for example, past tense forms, even though as a lexical verb HTETI ‘want’ has a full range of tense forms. The forms of the auxiliary are thus limited by virtue of it being a component part of a larger, periphrastic construction. The default assumption is that syntactically separate elements should not condition each other in this way, and all syntactically viable combinations should be generated. Underexhaustivity is a counterexample to this. However, if periphrasis is viewed as the realization of inflectional values, the underexhaustivity of the

auxiliary comes ‘for free’. The auxiliary is treated as part of the realizational of the lexical verb, so the only auxiliary forms that are licensed are the ones needed for future tense expression. Note that this argument is thus somewhat similar to the one made by Bonami and Boyé (2007) for French clitics, in that independently-needed mechanisms for morphosyntactic feature licensing within an inferential-realizational model are employed to account for distributional restrictions of grammatical particles.

Finally, patterns of blocking can also argue for a morphological approach to periphrasis, in a way that parallels and extends arguments in the clitic literature. In particular, Bonami (2015) argues that Pāṇinian splits – competition between periphrastic and synthetic realizations that is governed by Pāṇinian ordering, also called Elsewhere ordering or specificity-based ordering – offer strong evidence for the necessity of treating some examples of periphrasis as morphological in nature. Bonami observes that synthesis can function as the specific case that pre-empts the general periphrastic pattern (his example comes from Persian verbs), and crucially, the reverse relationship is also found, in which periphrasis functions as the specific case that pre-empts the general synthetic pattern. In Tundra Nenets, noun inflection for case and number is generally synthetic, including all singular and plural forms, and also the nominative, accusative and genitive dual forms. These three dual cells are realized by the same, syncretic form, indicating a case-underspecified default form. The interesting thing is that this synthetic form is preempted by morphosyntactically more specific – and periphrastic – forms in the remaining dual cases. In short, the periphrastic forms bleed the ‘elsewhere’ synthetic form. This interaction suggests direct competition between constructions in a way that is inherent and central to inflectional structure in inferential-realizational models.

These and other arguments have driven an approach that analyzes periphrasis as the realization of inflectional values, generated by principles of morphological structure (Ackerman and Stump 2004; Bonami 2015; Bonami and Samvelian 2009; Popova and Spencer 2013; Sadler and Spencer 2000). The logic lies in large part in the mapping between morphosyntactic values and morphophonological form: to the extent that periphrases are semantically non-compositional in the same manner as synthetic forms, have only those forms as are dictated by the set of licensed paradigm cells, and enter into Pāṇinian blocking with synthetic forms, the same logic that motivates inferential-realizational models in the first place serves to motivate a (partly) morphological approach to periphrasis. We must therefore consider periphrases that exhibit behavior of morphological objects to have been morphologized, at least in part, even if they consist of separate syntactic atoms. Some possible formal approaches to this issue are outlined as part of the discussion of Macedonian and Bulgarian in Section 4 below.

## 2.5 Summary

In this section we have developed an understanding of what it means for a construction to be ‘morphologized’ that is rooted in an inferential-realizational approach to morphological structure and a lexicalist interface to the syntax. The most important idea to emerge is that ‘morphological’ cannot be equated with ‘synthetic’. While periphrastic constructions by definition consist of at least two syntactic atoms, a degree of morphologicity is nonetheless possible, in terms of integration into paradigmatic structure and the system of realizational rules. The same applies to clitics. We reiterate that clitics and periphrasis are both wide-ranging phenomena, so a declaration that they are universally morphological in nature, or universally syntactic in nature, is not justified. How to handle constructions with mixed properties is an important question for formal theories, and while not all issues are yet solved, recent inferential-realizational analyses have offered possible ways to formalize the morphological properties of both clitics and periphrasis. So without further delay, we turn now to the more empirical side of our discussion, where the theorizing of this section is put to work on data from the verbal complex in three language groups represented in the Balkans: Greek, Albanian, and Slavic. We leave the analysis of Balkan Romance facts to another study (though see Bîlbîie and Mardale (This volume: Section 2.2.2) for some argumentation concerning the status of the Romanian subjunctive marker *să*).

## 3 The verbal complex in Greek and Albanian

We are not the first to be concerned about the nature of the verbal complex as a construct in a Balkan language. Newmark et al. (1982: 23), for instance, without any argumentation and without couching their remarks in any particular theoretical framework, have this to say about Albanian:

Verbs are typically thought of as single words, but in Albanian one or more proclitics and auxiliaries may precede the main verb and the whole sequence is then still referred to as ‘the verb’.

Similarly, ~~Schumacher and Matzinger (to appear~~ 2017: §2.3.2), basing their claim on the fact that weak pronouns can co-occur with full nominal objects, explicitly state, regarding the weak object pronouns, that “it is preferable to describe these pronominal elements as verbal affixes, i.e. as agreement markers belonging to the



verb. In other words, Albanian has a polypersonal verb with direct and indirect objects being optionally indicated on the verb.”<sup>11</sup>

Such concerns are not limited just to Albanian, as similar issues arise, and have been argued about, for the other languages. Philippaki-Warbuton and Spyropoulos (1999), for instance, claim that each piece of the Greek verbal complex is a word, each with its own node and projection in a syntactic tree, whereas Joseph (2002) counters that view with arguments that the whole complex is the word. And, one can ask if the ~~Schumacher and Matzinger~~ argument can be amplified upon and extended into the other languages.

### 3.1 The Greek verbal complex as a morphological object

We start with the Greek verbal complex, exemplified here in (7a), repeating (1c), and then given in its standard Modern Greek form (7b):

- (7) a. *Đe ðe na tu to đóso* (dialectal Greek)  
 NEG FUT SBJV 3SG.GEN 3SG.ACC give.1SG  
 ‘I will not give it to him.’
- b. *Đe ða tu to đóso* (Standard Modern Greek)  
 NEG FUT 3SG.GEN 3SG.ACC give.1SG  
 ‘I will not give it to him.’

We structure the presentation in this section in terms of the criteria of Z&P.

We begin the presentation with the combination of the verb with the weak object pronouns, the innermost elements that always occur immediately adjacent to the verb itself. Based on a verb-doubling construction found in Greek with the negator *ðe* ‘not’, giving the meaning ‘whether one VERBs or not’, there is an indication, consistent with Z&P’s criterion E (“syntactic rules may take affixed words as their domain; they cannot take clitic groups as their domain”), that the verb-plus-weak-object-pronoun as a unit is manipulated by a syntactic rule. The relevant construction is illustrated in (8):

- (8) *ðeli ðe ðeli*  
 want.3SG not want.3SG  
 ‘whether he/she/one wants (to) or not’

---

<sup>11</sup> See the postscript to this paper for more on the co-occurrence of weak pronouns with full nominal objects.

As (9) indicates, this construction appears to be restricted to doubling just one word.

- (9) a. *\*θeli bira ðe θeli bira, θa pji*  
 want.3SG beer.ACC not want.3SG beer.ACC FUT drink.3SG  
*jati jortazo.*  
 because celebrate.1SG  
 ‘Whether he wants a beer or not, he’ll drink because I am celebrating.’
- b. *\*θeli o Janis ðe θeli o Janis.*  
 want.3SG the John.NOM not want.3SG the John.NOM  
 ‘whether John wants (to) or not’

Importantly for understanding the morphological status of the verbal complex, weak pronouns can and in fact must be doubled along with the verb (10).<sup>12</sup>

- (10) *to θeli ðe to θeli*  
 it.ACC want.3SG not it.ACC want.3SG  
 ‘whether one wants it or not’  
 (NB: *\*to θeli ðe θeli* / *\*θeli ðe to θeli*)

Thus with respect to the doubling that constitutes this construction, the weak pronoun itself does not behave like a discrete unit, and the composite consisting of the verb plus weak pronominal elements, the only parts of the verbal complex suitable for use here,<sup>13</sup> itself behaves like a single word. The rule responsible for the doubling, therefore, treats this composite as a unit.

The remaining evidence for morphological status for the pieces of the verbal complex mostly concerns Z&P’s other criteria and in particular various behavioral idiosyncrasies which point towards the complex being morphological in nature.

<sup>12</sup> Iliyana Krapova pointed out to us that if weak pronouns also cannot be stranded under deletion of the verb (in a gapping structure), even when the weak pronouns have an appropriate prosodic host, this is further evidence that the pronoun and verb form an inseparable unit for purposes of syntactic operations. This is indeed true for Greek. We thank her for this comment.

<sup>13</sup> The markers *θa* and *na* are excluded on semantic grounds because the construction is future-like and modal-like as it is, and negation is already there as part of the construction itself. The insertion of other phonologically “minimal” material, like the 2SG nominative pronoun *si* ‘you’, that could in principle occur after the first verb, yields ungrammatical results: *\*θelis si ðe θelis* ‘whether you want or not’ (note too that *\*θelis si ðe si θelis*, is also hopeless, but there, *si* intervening between *ðe* and the verb dooms it from the start).

In particular, with regard to the Greek weak object pronouns, as argued in Joseph (2002), they:

- i. are selective: with some rare exceptions,<sup>14</sup> they occur only with verbs, not with other hosts (Z&P criterion A);
- ii. show idiosyncratic morphophonology: in the combination of 2SG.GEN /su/ + a 3rd person accusative pronoun, e.g. /to/ 'it.N' or /tus/ 'them.M', the /u/ of the 2SG form can be elided to give, e.g. [sto], [stus], etc.,<sup>15</sup> and in the combination of the future marker /θa/ with a 3rd person accusative pronoun, e.g. /θa + to/ 'FUT + it.ACC', etc., there can be otherwise unexpected voicing of /t/ to [d], giving [θa do] (Z&P criterion C);<sup>16</sup>
- iii. show idiosyncratic involvement in argument structure and consequently in semantics: in particular, in the expression in (11),

- (11) *pame na tin pesume s to krevati*  
 go.1PL SBJV her.ACC fall.1PL in the bed  
 'Let's go for some sleep in the bed'

which is literally "let's-go that we-fall her in the bed", there is a weak "object" pronoun *tin* occurring with a verb that outside of this expression is intransitive (meaning 'fall', as in 'X falls from a tree' or 'someone falls down'); here, however, anomalously the verb appears to be transitive, with an apparent direct object in the form of the accusative weak pronoun *tin*. In this case, then, idiosyncratically the weak pronoun is not contributing anything to the argument structure nor to the meaning, and yet it is there, a seeming object but actually not indicating an object at all.

Turning now to the other elements in the Greek verbal complex and applying the criteria to them, we find the following:

<sup>14</sup> For instance, there are a few particles and adverbs that have imperative-like semantics and take (postposed) weak object pronouns; an example is *kalos ton* 'welcome (to) him', with the accusative *ton* occurring after *kalos* 'welcome' (otherwise an adverb meaning 'well').

<sup>15</sup> As discussed in Joseph (2002), the issue here is the elision of /u/ without lip-rounding; in cases where an unstressed /u/ is elided in fast speech between /s/ and /t/, as in the verb *sutaro* 'shoot (a ball)', the /s/ is rounded: /sutáro/ 'shoot.1sg' → [sʷtáro].

<sup>16</sup> The voicing would be motivated if a nasal was involved synchronically, as nasals in Greek trigger voicing on voiceless stops. As it happens, diachronically, there was an historical stage in which *θa* ended in a nasal, having the form *θan*. This is the historical source of the synchronically unexpected voicing. However, there is no trace of the nasal anymore, e.g. it is not found before vowel-initial verbs where it might be expected to have remained, so there is no basis for setting up an underlying nasal for *θa* in present-day Greek.

- a. Future *θa*: it is selective in host attachment, allowing attachment only to verbs or weak pronouns that follow it in the verbal complex; it occurs in idiosyncratic combinatory morphophonology (cf. *θa do* above in (ii)); and it is involved in idiosyncratic semantics in the expression *ti θa pi* ‘what does it mean?’ (literally: “what FUT (it-)says”)
- b. Negative *ðen*: it is selective in host attachment, like *θa*, allowing attachment only to verbs or elements that can follow it in the verbal complex string (i.e., *θa* and weak pronouns); and it is involved in idiosyncratic semantics and morphosyntax in the expression *ðen mu les* ‘Tell me!’ (literally: “NEG me.GEN say.2SG.IND”, i.e. ‘you do not tell me’, thus seemingly an imperative though not imperatival from a formal standpoint and with no semantic negation despite a formal marker of negation).
- c. Subjunctive *na*: it provokes the occurrence of a special negation marker, *mi*, as opposed to the indicative *ðen*; and it shows idiosyncratic combinatory morphophonology – like /*θa to*/ giving [*θa do*] above in (a) (and (ii)), the combination of *na* with a 3rd person object pronoun allows for otherwise unmotivated voicing of the initial /t-/ of the pronoun, e.g. /*na + to*/ → [*na do*] (and see Bîlbîie and Mardale (This volume: Section 2.2.3) for additional argumentation concerning the status of *na*).

As argued in Section 2 above, these kinds of idiosyncrasies indicate that the elements of the verbal complex are generated as a unit within the morphological component (or lexicon, depending on perspective), with a high degree of host selectivity and invisibility to syntactic rules further suggesting synthesis. The conclusion to be drawn for Greek from this assemblage of facts is that the verbal complex represents a fully morphologized, even synthesized, construct.<sup>17</sup>

### 3.2 The Albanian verbal complex as a morphological object

Except for the Verb-NEG-Verb copying construction, which is particular to Greek and does not have a direct analogue in Albanian, the same sort of reasoning that

---

<sup>17</sup> And it is entirely appropriate to call it “morphologized” and not just “morphological”, because in earlier stages of Greek, some of the relevant pieces – or their historical sources – had greater integrity and independence; for instance, weak object pronouns could be positioned relative to a clause, or at least various sentence connectives – see Pappas 2004 – and the source of future marker *θa* was originally a fully inflected verb, with the same form as the verb ‘want’ (see Section 5 below on the relevant developments).

indicates that the Greek verbal complex is a morphological object can be given for the Albanian verbal complex, repeated here in (12) from (1a):

- (12) *S' do të ja jep.*  
 NEG FUT SBJV 3SG.DAT/ACC give.1SG  
 'I will not give it to him.'

We give the argumentation here in a more schematic form.

First, the elements in the Albanian verbal complex show a high degree of selectivity as to co-occurrence; for the most part, the verbal complex modifiers do not occur outside of the context of the complex. The future marker *do* has the same shape as its etymological source, the 3sg present tense form of the verb *dua* 'want', and there is an element *do* that occurs in indefinite pronouns, e.g. *kushdo* 'whoever' (cf. *kush* 'who?'), which is also best identified etymologically with the 2/3SG of the verb *dua* 'want',<sup>18</sup> but it is not clear that there is any reason to connect these forms synchronically. In fact, *do* 'future' behaves differently from *do* 'you/he want/s' in that it allows the elision of the subjunctive subordinator *të* whereas the verb 'want' does not; thus, *do të shkoj* can mean both 'I will go' (literally "will that I-go") and 'you want me to go' (literally, "you-want that I-go"), whereas *do shkoj*, with the *të* elided, can mean only 'I will go' and not 'you want me to go'.<sup>19</sup>

The strongest evidence for morphologized status comes with the weak pronouns and the fact that they show special portmanteau realizations – tantamount to special morphophonology – in certain combinations with one another, e.g. DAT.3SG *i* + ACC.3SG *e* → *ia*, DAT.3SG *i* + ACC.3PL *i* → *ia*, DAT.1SG *më* + DAT.3SG *e* → *ma*, inter alia, and with the subjunctive marker, e.g. *të* + *e* → *ta*; the realizations here are unexpected, in that based on other aspects of Albanian phonology, one might expect [ie] to remain, *i* + *i* to end up as [i], and *të* + *e* to yield simply *të* with the *ë* elided.

Moreover, as noted above, the future marker *do* allows complete elision (deletion) of the subjunctive marker *të* with no change of meaning or grammaticality, only the stylistic difference of *do shkoj* being more colloquial than (future) *do të shkoj*. This elision is thus a special feature of the combination of future *do*

<sup>18</sup> For a typological parallel, cf. Spanish *cualquier* 'any', where the second element is based (historically at least) on the verb *quiero* 'want'.

<sup>19</sup> What makes this situation somewhat complicated is that the future marker itself derives historically from the 3sg present form of the verb 'want', but via a different route. And there are formal differences synchronically in that in some dialects, the 3sg present form (2sg too) has been inflectionally regularized to *don* 'wants' (with 2/3SG – *n*) but the future *do* remains as *do*.

with the subjunctive marker, and is as much an idiosyncrasy of future *do* as it is of subjunctive *të*.

The conclusion to be drawn from this idiosyncratic behavior together with the selectivity shown by the various elements of the Albanian verbal complex is that the complex is a synthetic object. This is much as Newmark et al. (1982) say (see in Section 3 above), though not in so many words, and it accords with the conclusion about the parallel entity in Greek.

### 3.3 Summary

It must be admitted that the evidence cited in Sections 3.1–3.2 from Greek and Albanian is compelling to different degrees. While all the evidence points in the same direction, towards recognizing the morphologization of the respective verbal complexes, it is not clear, for instance, that one lone morphophonological idiosyncrasy with the second person singular genitive pronoun *su* is enough in itself to require a categorization of all the weak pronouns as affixes. Still, all in all, it seems fair to say that the accumulation of the evidence from Greek and Albanian points to a high degree of morphologization, even synthesis. We believe that the data can be accounted for fairly straightforwardly in theoretical terms as affixation, for which a garden-variety inferential-realizational theory will do. This result becomes especially interesting in the light of the evidence from South Slavic.

## 4 The verbal complex in Macedonian, Bulgarian, and Serbian

We now turn to the verbal complexes in the South Slavic languages of the Balkans – Macedonian, Bulgarian and Serbian. The languages differ substantially in the extent to which elements in the verbal complex exhibit synthesis and attachment to the verb so the picture that emerges is a continuum of morphologization. While Greek and Albanian, as just discussed, exhibit a high degree of synthesis within the verbal complex, Serbian falls on the other end of the continuum, with no attachment of elements to the verb. Macedonian and Bulgarian represent intermediate points, with the verbal complex more morphologized in the former than in the latter. As in Greek and Albanian, the elements of the verbal complex in all three languages are prosodically dependent, with the exception of the lexical verb. The facts surrounding the pronominal and auxiliary clitics in the

complex are already well established, and we focus here only on those issues that are most directly relevant to the question of morphologization. However, readers are referred to Franks and King (2000) and Friedman and Joseph (in prep) for a fuller description of the facts.

#### 4.1 The Macedonian verbal complex as a (mostly) morphological object

As already observed, dialectal Macedonian has a surface word order that parallels the order of elements seen above in Greek and Albanian. Example (13a) repeats (1d); compare this to the verbal complex template, repeated as (13b).

- (13) a. *Ne ke da mu go davam.*  
 NEG FUT SBJV 3SG.M.DAT 3SG.M.ACC give.1SG  
 ‘I will not give it to him.’  
 b. NEG – TNS – MOOD – IO – DO – V

By contrast, in Standard Macedonian, *ke* combines with the verb (plus any object clitics) without modal *da* (14).<sup>20</sup> *Da* can follow *ke* in suppositional clauses, as in (15) (example from Kramer 1986: 76), although Friedman (1993: 285) marks this construction as ‘colloquial’.

- (14) *Ke mu ja dadam knjigata.*  
 FUT 3SG.M.DAT 3SG.F.ACC give.1SG book(F).DEF  
 ‘I will give him the book.’  
 (15) *Ke da imaše edno osumnaest godini.*  
 ke DA have.IMPF.3SG about eighteen years  
 ‘He must have been some eighteen years old.’

In the standard language, *da* occurs at the beginning of the clitic group.<sup>21</sup> The order of elements for the standard language can thus be distilled to the template

<sup>20</sup> The negated future is also typically formed with *nema da* (with the negated form of ‘have’) rather than *ne ke*.

<sup>21</sup> A reviewer commented that modal *da* and complementizer *da* may need to be distinguished, along the lines of Franks and King (2000: 81, table 2.16). They say that “... *da* as complementizer

in (16), (adapted from Franks and King 2000: 81), and exemplified in (17) (from Friedman 1993: 285).

(16) *da* – *ne*<sup>22</sup> – *ke/bi*<sup>23</sup> – AUX – DAT OBJ – ACC OBJ – *e/se*<sup>24</sup> – V

(17) ... *da ne ke sum si mu go dal?*  
 ... that NEG FUT AUX.1SG REFL.DAT 3SG.M.DAT 3SG.M.ACC gave  
 ‘(They didn’t say) ...that I won’t have given it to him (did they)?’

The order of elements in the verbal complex thus differs somewhat within Macedonian, with the pattern in (13a) more closely matching the surface order of elements found in Greek and Albanian. From a contact perspective, it is unsurprising that a regional variety, rather than the standard one, should be more closely similar to languages with which Macedonian has been in historical contact, since Balkan contact has been characterized by intense multilingualism at the local – thus, dialectal – level. However, other than this difference, the relevant facts are largely the same, and in the following discussion we give examples from the standard language.

---

is initial; modal *da* follows hypothetical *ke*” (81). We admit that this is an area in which the facts are not entirely clear to us. Mišeska Tomić (2006: 243) lists a ‘subjunctive’ marker in two different places within the clitic cluster. These two placements are the same as those identified by Franks and King, but Mišeska Tomić identifies *da* before *ke* as modal rather than as a complementizer. Friedman (1993: 285) likewise labels *da* in examples like (17) as subjunctive, but he lists only one position in the cluster (initial). He does not mention the possibility of *da* after *ke* except in examples like (15), and thus appears to treat the difference as a function of construction type and register or dialect. Kramer (1986) argues that there is no evidence for two *da* in Macedonian, in contrast to Serbian, although her discussion in general focuses more on semantic properties than formal syntactic ones. Here we base our discussion on Friedman, but these sources do not lead to a clear answer on whether two *da* need to be distinguished with regard to the verbal complex. <sup>22</sup> The position of the negative marker *ne* depends on the scope of negation. In particular, *ne* precedes *da* if it takes scope over the entire clause, and occurs after *da* if it has narrower scope, over the VP. This affects the meaning of the sentence (Friedman 1993: 290). Here we include the latter position, since *ne* in this position is clearly part of the verbal complex.

<sup>23</sup> *Bi* is a conditional marker. Like *ke*, its form is invariant.

<sup>24</sup> These are the third person forms of ‘be’ – singular and plural, respectively. Since the third person auxiliary has a null form, these forms occur only in the copula function, but in that function they are positioned within the clitic cluster.



In Macedonian the clitic cluster is strictly adjacent to the verb, appearing to the left of the finite verb (see (14) above) and to the right of a non-finite verb, including the imperative, (18).<sup>25</sup>

- (18) *Donesete mu ja knjigata.*  
 bring.IMP 3SG.M.DAT 3SG.F.ACC book(F).DEF  
 ‘Bring him the book.’  
 (\**Mu ja donesete knjigata.*)

Paralleling Greek and Albanian, nothing can intervene between the elements of the verbal complex.<sup>26</sup> In fact, the only things that can intervene between *da* and the verb are members of the cluster (Kramer 1986: 8), making *da* functionally part of the verbal complex. We note that analyses of *da* as a preverbal morpheme go back at least to Gołąb (1954) and Maslov (1956) (as cited in Kramer 1986: 54–55). The clitic cluster (including *da* and with limited exceptions *ne*) is thus inseparable from the verb, even though it sometimes appears obligatorily to the left of the verb and sometimes equally obligatorily to the right of it. Additionally, as shown in (14) above, the cluster can also appear in absolute sentence-initial position when proclitic to the verb. This again makes Macedonian similar to Greek and Albanian (but different than Bulgarian, as we will see), suggesting that the clitic-first pattern is a Balkanism (Alexander 1994: 4).

The extent of cohesion of the verbal complex can be observed in the behavior of the question particle *li*, which is phonologically enclitic but not a part of the verbal complex in Macedonian (Englund 1977: 116). *Li* strictly follows the first prosodic word in the clause and is strongly associated with focus. When anything other than the verb is focused, it is fronted and serves as the prosodic host for *li* ((19), from Friedman 1993: 287). Otherwise, *li* appears immediately after the verb (20).

- (19) *Vo Bitola li će odiš?*  
 to Bitola Q FUT go.2SG  
 ‘Is it BITOLA you will be going to?’

<sup>25</sup> The placement of clitics in non-verbal predicates (e.g. *Tatko mi e.* / ‘He is my father.’) is less rigid. See Mišeska Tomić (1996) and Franks and King (2000: 85–88) for some discussion. Also, Friedman observes that in the *beše* pluperfect, clitic pronouns may either precede or follow the auxiliary, and that “[t]he sense of past resultativity is stronger when the auxiliary is closer to the verb” (Friedman (1993: 286), citing Koneski (1967)).

<sup>26</sup> Friedman (1993) notes that some ‘old-fashioned’ phrases (curses and blessings) are exceptions to this generalization. This is what we might expect from the fact that there used to be greater syntactic freedom of movement; word orders that are no longer licensed in the language became frozen in set phrases.

- (20) a. *Ķe odiš li vo Bitola?*  
 FUT go.2SG Q to Bitola  
 ‘Will you go to Bitola?’  
 (\**Ķe li odiš vo Bitola?*)
- b. *Ti go dade li?*  
 2SG.DAT 3SG.M.ACC gave.3SG Q  
 ‘Did s/he give it to you?’  
 (\**Ti li go dade?* \**Ti go li dade?*)

With two possible exceptions, *li* cannot be placed internally to the cluster or between the cluster and the verb. The first exception has to do with emphatic negation. Rudin et al. (1999) observe that *li* can immediately follow the negator *ne*, but only when *ne* receives independent lexical stress as a result of emphatic negation (21a). In neutral negation, *ne* does not receive independent stress and is phonologically proclitic to the verb, like the rest of the verbal complex. *Li* then follows the verb (21b).<sup>27</sup> (Stress is indicated with capital letters and the prosodic word with square brackets.)

- (21) a. [NE] *li* [*ti* GO dade]?  
 NEG Q 2SG.DAT 3SG.M.ACC gave.3SG  
 ‘Did s/he really not give it to you?’
- b. [*ne* *ti* GO dade] *li*?  
 NEG 2SG.DAT 3SG.M.ACC gave.3SG Q  
 ‘Didn’t s/he give it to you?’  
 (\**ne li ti GO dade?* / \**ne ti GO li dade?*)

This variable placement of *li* relative to *ne* suggests that the negator is sometimes a part of the verbal complex, and sometimes not.<sup>28</sup> Crucially, *li* can directly follow *ne* only when the latter has its own lexical stress, which offers independent evidence of it not being part of the verbal complex when emphatically stressed. We therefore consider it to be a pseudo-exception.

<sup>27</sup> We can see that it is lexical stress on *ne* that matters for placement of *li*, not just any stress, by the fact that when *ne* is stressed by virtue of being the antepenultimate syllable within the verbal complex, it does not host *li*: [*NE sakaš*] *li da odiš*? ‘Don’t you want to go?’ (Rudin et al. 1999: 556). Stress is always antepenultimate within the prosodic word in Macedonian.

<sup>28</sup> This behavior is paralleled in Greek, where the negator *den* may, but need not, receive its own stress; stressed *den* carries emphasis; however, in Greek there is no other indicator like Macedonian *li* that could provide independent confirmation of the stressed negator being outside of the verbal complex.

The other exception has to do with non-finite verbs – the condition in which the clitic cluster follows the verb. In all of the examples above, the clitic cluster precedes the verb. The question is: When the cluster follows the verb, where is *li* placed? The relevant type of sentence seems to be quite rare for several reasons: there are restrictions on fronting of verbal participles in Macedonian (Rudin et al. (1999: 576), citing Embick and Izvorski (1997)); when there is a finite auxiliary in the clause the clitic cluster attaches to it, reducing the occurrence of clitic clusters following non-finite lexical verbs; and it is hard to form a question phrase with *li* that also contains an imperative (which, recall, behaves as a non-finite verb). However, Victor Friedman (personal communication) suggests the example in (22). The relevant part is B's response, which includes the question particle *li*.

- (22) A: *Davajki mu go stapot, sliznal i padnal.*  
 giving 3SG.M.DAT 3SG.M.ACC baton.DEF, slipped.M.SG and fell.M.SG  
 'While giving him the baton, he slipped and fell.'

B: *Davajki li mu go? A jas mislev deka*  
 giving Q 3SG.M.DAT 3SG.M.ACC But I thought that  
*sliznal porano!*  
 slipped.M.SG earlier  
 'While giving it to him? I thought he slipped earlier!'

Notice that *li* is enclitic to the non-finite verb and precedes the dative and accusative object clitics. It is unclear whether this ordering is consistent throughout Macedonian, but for at least some speakers it appears to be an exception to the generalization that the clitic cluster is strictly adjacent to the verb.<sup>29</sup>

What does the placement of *li* tell us about the cohesiveness of the Macedonian verbal complex? As part of an argument that the syntactic placement of *li* is the same in Macedonian and Bulgarian, Rudin et al. (1999) analyze *li* as attaching to a prosodic word domain in Macedonian that consists (potentially) of several syntactically separate elements, i.e. the grammatical elements of the verbal complex. This is shown by the bracketing in (21). This analysis allows them to posit that the only difference between *li* in Macedonian and Bulgarian has to do with the size of the domain to which *li* is prosodically enclitic. (The Bulgarian facts are discussed

<sup>29</sup> Rudin et al. (1999: 576) cite a similar example: *Predupreden li si bil za toa?* 'Were you WARNED (really) about that?' Here, *li* separates the adjectival participle *predupreden* 'warned' from the auxiliary verb clitic *si*. However, as noted in footnote 25 above, the placement of clitics in non-verbal predicates is not as rigid as with verbs. The importance of this example for the placement of *li* is thus not fully clear.

below.) However, as far as we can see, there is an equally viable analysis positing that *li* attaches to a lexical level ( $X^0$ ) unit in both languages. In this approach, differences in *li* placement in the two languages relate to what constitutes a lexical unit. Analyses along these lines have been sketched by Spencer (2000) for Macedonian (separately from the question of *li* placement) and Sadler (1997) for Welsh pronominal clitics (see Section 2.3 above), which bear partial resemblance to Macedonian clitics. Here we try to at least give the flavor of the accounts.

Spencer (2000: 379–381) argues that since only verbs can host the clitic cluster in Macedonian, this motivates an analysis of the clitics as exponents of morphosyntactic properties of the verb. In other words, in Macedonian the paradigm of the verb includes auxiliary and object clitics. Working within Paradigm Function Morphology, Spencer shows that this can be formalized in terms of a(n Extended) Paradigm Function that defines realizational rules realizing the combination of an inflected verb and clitic-realized morphosyntactic properties. In essence, the clitics are generated as affixes, albeit ones that attach to already-inflected words. Whether the cluster is positioned before or after the cluster is treated as morphological conditioning based on the properties of the verb (whether it is finite or not). Since the combination of verb and pronominal and auxiliary clitics is output by the morphology as a single lexical unit, this naturally captures the fact that they are syntactically inseparable.

Sadler (1997)'s analysis of Welsh pronominal clitics captures some of the same insights about clitics and verbs forming lexical units. However, she takes a somewhat different approach that has some advantages when applied to Macedonian. Remember that the essence of her analysis is that the clitic cluster and verb are each generated as separate morphological objects, and each forms a lexical level ( $X^0$ ) construction in syntax. However, the clitics are functional categories that do not project to a maximal projection, and instead attach as a lexical sister to the verb, forming a(nother) lexical level (small/ $X^0$ ) construction with it. This is congruent with the essence of Spencer's proposal that the clitic cluster is morphologically generated as a unit and then affixed to the already inflected verb. And like in Spencer's analysis, the fixed, templatic order of elements in the clitic cluster are a direct result of the fact that it is morphologically generated as a unit. While we haven't worked out the details, we expect an analysis of Macedonian along the lines of either Spencer's or Sadler's would be able to handle the syntactic 'inertness' noted by Legendre (2001a). Also, note that the Macedonian clitics do not exhibit allomorphy depending on the verb that it combines with and there are no lexical exceptions to cluster-verb combinability (Z&P's criteria C and B). Sadler's account as extended to Macedonian precludes morphophonological interactions of this sort on principle, since the clitic cluster and the verb are generated as independent morphological objects. Spencer's account does not preclude allomorphy, but does not require it either.

Admittedly, neither Spencer's nor Sadler's analysis has an obvious way to account for the placement of *li* between a non-finite verb and following clitic cluster (22), given that it is clear that *li* is not generated as part of the verbal complex/clitic cluster. Both analyses would somehow need to assume that *li* can be inserted into the middle of a lexical unit. At the same time, we note that Rudin et al. (1999)'s analysis faces equal problems, since when the verb is non-finite, it must posit either that *li* gets inserted into the middle of a prosodic word, or that *li* is enclitic to a stressed syntactically-minimal word. Both options contradict the analysis made of finite verbs. Of these three, Sadler's approach, as extrapolated to Macedonian, seems to offer the greatest possibility for a viable solution, since the boundary between the verb and clitic cluster is visible to the syntax (unlike Spencer's account, in our understanding), and the account is not oriented to the boundaries of the prosodic word (unlike Rudin et al.'s). Moreover, an analysis that locates the difference between Bulgarian and Macedonian in the extent of morphologization has the advantage of also being able to explain why no other elements can intervene between the clitics and the verb. This is something that Rudin et al.'s analysis offers no direct account of.

Finally, it is worth briefly considering two other arguments, from Kramer (1986: 7), for considering the verbal complex in Macedonian to be syntactic. First, she posits that since the cluster appears both before and after the verb, placement of the cluster must be according to syntactic rules. Second, she observes that the future marker *ќе* has been 'deparadigmaticized' in the sense that it has a frozen form and no longer inflects for person and number. Although she does not elaborate on the argument, the idea seems to be that its invariant form allows it to be treated as a purely syntactic particle. And indeed, the loss of person-number marking on *ќе* resulted in loss of multiple exponence of person-number; multiple exponence is a diagnostic of morphological objects (Spencer 2012). In the construction *ќе (da) V*, it localizes person-number marking to the finite verb (either lexical verb or auxiliary).<sup>30</sup> It is thus not necessary to assume that the future has constructional status.

However, Kramer's observations are not actually in conflict with a morphological account of the kind sketched above. First, Anderson (2005: 85) observes that in Macedonian and similar languages, "...the order which is strictly required under one set of circumstances [finite lexical verb] is replaced by another under a complementary set of conditions [non-finite verb]... Parallel to these cases are

---

<sup>30</sup> Also, since the meaning of the construction can be parceled out to the individual components, the construction does not exhibit distributed exponence either. Distributed exponence was proposed by Ackerman and Stump (2004) as a criterion for identifying that a periphrastic construction is morphological, although this criterion has been criticized (Brown et al. 2012; Spencer 2012).

examples in word-level morphology where the same affix may show up either as a prefix or as a suffix, depending on specific factors... What is notable is that in each case, the position of the affix is not at all free,” making the placement of the clitic cluster rather unlike syntactic phenomena like scrambling. So while variability in placement of an affix relative to the root is unusual in morphology, it is not unheard of (see Nevis and Joseph 1993, for instance, on word-internal Wackernagel-like second positioning of the reflexive marker in Lithuanian), and in fact its lack of freedom of movement is unusual from a syntactic perspective. As we have already seen, there are at least a couple of different models for how to account for this fact in a fundamentally morphological way. Second, while the frozen form of *ke* does not force us to recognize a constructional status for the future in our analysis, neither does it preclude it. And when we take into account the cohesion of the clitic cluster with the verb (high selectivity, non-separability), the evidence tips in the direction of recognizing at least some degree of morphologization.

To review, it is clear that the grammatical elements in the verbal complex are not canonical affixes. They occur on either side of the verb,<sup>31</sup> depending on its finiteness, do not induce allomorphy within the cluster or in the verb, and there are no lexical exceptions that we are aware of. At the same time, the grammatical particles always take verbal hosts (i.e. have high host selectivity) and with the possible exceptions of the negative *ne* when independently lexically stressed, and the placement of *li* when the clitic group follows the verb, they are strictly adjacent to the verb. They also exhibit a fixed, templatic order within the clitic cluster. The clitic cluster is thus partly synthesized. These properties are surprising if the clitics each form their own maximal projection in the syntax, and instead indicate some amount of morphologization. An analysis of the clitic cluster as a morphologically generated object that combines into a lexical unit with the verb (whether along the lines of Spencer’s analysis or Sadler’s) seems likely to offer a good account of the intermediate status of the Macedonian verbal complex between syntax and morphology.

## 4.2 The Bulgarian verbal complex as a morphological object?

Like the other languages, Bulgarian has a series of grammatical elements that can occur before the verb and form a (surface-descriptive) verbal complex with it; see

---

<sup>31</sup> Interestingly, variable ordering of objects in Greek in postposed contexts (after nonfinite forms) can be observed, i.e. *dos mu to ~ dos to mu* ‘give.IPV.SG me.GEN it.ACC’ ~ ‘give it me’, suggesting that the postposed pronouns are less synthetic with the verb (though still adjacent).

(23a) (examples from Franks and King 2000: 59). However, no subjunctive marker occurs between the future marker and object clitics. The future formed with *ne šte* is actually formal and archaic and *njama da* is the more prevalent way to form the negative future (23b). Here a subjunctive marker does occur, but negative and future are synthetically expressed, *njama* being historically *ne + ima* ‘have’.

- (23) a. *Ne šte ni ja četeš.*  
 NEG FUT 1PL.DAT 3SG.F.ACC read.2SG  
 ‘You won’t read it to us.’
- b. *Njama da ni ja četeš.*  
 NEG.FUT SBJV 1PL.DAT 3SG.F.ACC read.2SG  
 ‘You won’t read it to us.’

As for the other elements in the verbal complex, the dative and accusative objects and verbal auxiliaries are clitics. The order of clitics can be extrapolated as in (24), illustrated in (25) (Franks and King 2000: 61–62). These elements appear in the same surface order as in the other languages.

(24) *šte* – AUX – DAT OBJ – ACC OBJ – *e*<sup>32</sup>

- (25) *Šte ste mu go kazali li?*  
 FUT AUX.2PL 3SG.M.DAT 3SG.M.ACC told.PL Q  
 ‘Will you have told him it?’

While the negator *ne* and the subjunctive marker *da* do not bear lexical stress, neither is generally considered to be part of the clitic cluster in Bulgarian. However, even the elements that are superficially similar to the verbal complexes of Macedonian, Albanian, and Greek turn out in Bulgarian to be less cohesive than these examples suggest.

The first issue has to do with the placement of the clitic cluster, which is prosodically conditioned in Bulgarian. Bulgarian object and most auxiliary clitics (but not *šte*, as we see below) are always verb-adjacent. The clitic cluster is syntactically proclitic to the verb by default, but prosodically enclitic. When there is a suitable leftward prosodic host the clitic cluster appears to the left of the verb (26a), but when there is not, the cluster appears after it, with the verb becoming the prosodic host (26b) (Franks and King 2000: 63).

<sup>32</sup> This is the third person singular auxiliary verb. As in other Slavic languages, it occurs at the end of the cluster, rather than in the position of other auxiliaries.

- (26) a. *Vera mi go dade včera.*  
 Vera 1SG.DAT 3SG.M.ACC gave.3SG yesterday  
 ‘Vera gave it to me yesterday.’
- b. *Dade mi go Vera včera.*  
 gave.3SG 1SG.DAT 3SG.M.ACC Vera yesterday  
 ‘Vera gave it to me yesterday.’

This means that Macedonian and Bulgarian are similar in both having verb-adjacent clitic clusters. However, placement of the cluster before or after the verb is prosodically conditioned in Bulgarian, rather than morphosyntactically conditioned as in Macedonian. Prosodic conditioning is suggestive that placement in Bulgarian is as a second position (Wackernagel) phrasal phenomenon. We assume that clitic positioning in Bulgarian is thus best treated via a combination of attachment to the head of the VP (i.e. above the lexical *leve*) and a prosodic constraint.

Second, some elements of the verbal complex are discrete units structurally. In addition to *da* and *ne*, *šte* behaves differently than other auxiliary clitics. For instance, the full-form auxiliary *BĀDA* can intervene between *šte* and other clitics in interrogatives (27) (example adapted from Spencer 2000: 362). (*Li* also here intervenes; we return to its placement below.)

- (27) *Šte bādeš li se vārnal do 5 časāt?*  
 FUT AUX.2SG Q REFL returned.M.SG by 5 hour.DEF  
 ‘Will you have returned by 5 o’clock?’  
 (\**Šte se bādeš li vārnal do 5 časāt?*)

Like *da* and *ne*, *šte* can also serve as a leftward prosodic host for the cluster (see (25) above), an unexpected property if it is itself a member of the cluster. Instead, its appearance in the verbal complex reflects the interaction of a phrasal domain of attachment and prosodic requirements that happen to sometimes place *šte* at the head of the cluster, rather than true unity with the other elements in the verbal complex (Hauge 1999[1976]).

Third and finally in this vein, the placement of the question particle *li* in Bulgarian demonstrates that elements can intervene into the cluster itself, or between the cluster and the verb, in contrast with Macedonian. This is particularly clear in negative contexts. The negator *ne* always throws stress onto the following morphosyntactic word, even if it is a clitic. *Li* follows the leftmost stressed element,<sup>33</sup>

<sup>33</sup> Actually, it is ambiguous between taking the leftmost stressed element as host, or the leftmost prosodic word. Rudin et al. (1999) make the latter analysis. This issue has no bearing



meaning that it can intervene between a stressed clitic and the host verb (28a), or between two clitics when the leftward clitic is stressed (28b,c) (adapted from Franks and King 2000: 60).

- (28) a. *Ne mu li izpratix kniga?*  
 NEG 3SG.M.DAT Q sent.1SG book  
 ‘Didn’t I send a book to him?’
- b. *Ne mu li ja izpratix?*  
 NEG 3SG.M.DAT Q 3SG.F.ACC sent.1SG  
 ‘Didn’t I send it to him?’
- c. *Ne si li mu go kazal?*  
 NEG AUX.2SG Q 3SG.M.DAT 3SG.M.ACC said  
 ‘Had you not said it to him?’

In principle, *li* can separate any two members of the verbal complex, assuming the necessary stress conditions. This is thus another way in which the verbal complex is not as cohesive a unit in Bulgarian as in Macedonian, and much less so than in Albanian and Greek.

The ‘core’ cluster has at least two properties in common with affixation: parallel to inflectional morphology, the cluster selects a verbal lexical head (Z&P criterion A), and clitic ordering within the cluster is invariant. Despite this, the Bulgarian verbal complex is not a coherent unit of analysis structurally. The clitic facts show that it is not synthetically bound and does not exhibit lexical integrity. In comparison to Macedonian – and, e.g., a Sadler-esque analysis in which clitics are adjoined as sisters to the lexical verb – a crucial difference in Bulgarian is that the clitic cluster must be assumed to attach at a higher (above lexical) level, in order to account for the separability of the cluster.

Spencer (2000) formally captures the ordering of both *šte* and the cluster elements (other auxiliary and object clitics) in Optimality Theory. He proposes a series of constraints, including INITIAL(*ŠTE*), stating that *šte* must be initial in its syntactic/prosodic domain, that is ranked above both ALIGN(CL) (have a leftward prosodic host) and INITIAL(CL) (align the left edge of the cluster with the left edge of the domain). This captures the generalization that while *šte* often appears on the surface as a member of the verbal complex, it is subject to different rules of placement and ordering than other members of the verbal complex. Legendre

---

here, however, with the only relevant thing being that *li* can (and often must) be inserted after a stressed pre-verbal clitic.

(1996, 1999, 2001b) also offers a formal analysis along similar lines, in the spirit of Anderson (1992, 1995, 2005)'s analysis of clitics as phrasal affixes.

This is not the end of the story, however. The complex structure of Bulgarian verbs offers additional evidence of morphologization in a different sense. The issues can be illustrated with the negated future perfect, for which Popova and Spencer (2013) argue that morphological constituency and syntactic constituency mismatch.<sup>34</sup> They analyze the future perfect form, e.g. *šte sām mislila* (29a), as being composed of a perfect form *sām mislila* (29b) nested formally and semantically inside the future construction *šte + V* (29c), as shown schematically in (29d).

- (29) a. *Do utre šte sām mislila veče za statijata.*  
by morning FUT AUX.1SG.PRS think.F.PTCP already about article.DEF  
'By tomorrow I will have thought about the article already.'
- b. *Mnogo pāti sām mislila za statijata.*  
many times AUX.1SG.PRS think.F.PTCP about article.DEF  
'I have thought about the article many times.'
- c. *Mnogo pāti šte mislja za statijata.*  
many times FUT think.1SG.PRS about article.DEF  
'I will think about the article many times.'
- d. [*šte [sām mislila]<sub>V</sub>*]

While (29a) is semantically compositional in a way that reflects the formal constituency, crucially, the nesting need not reflect syntactic constituency. Popov

<sup>34</sup> Semantic non-compositionality is also relevant here. For example, the emphatic renarrated mood is constructionally related to the renarrated mood (i).

(i)	Renarrated	Emphatic Renarrated
Present/Imperfect	<i>sām pišel</i>	<i>bil sām pišel</i>
Aorist	<i>sām pīsal</i>	<i>bil sām pīsal</i>
Present Perfect	<i>bil pīsal</i>	--

Spencer (2003: 264) argues that the emphatic renarrated is "... a non-compositional extension of a construction [i.e. renarrated] which is already pretty non-compositional...", and that this reflects the paradigmatic morphological structure of the constructions. His reasoning seems to be that while the form is identical to the pluperfect in some other Slavic languages, in Bulgarian the pluperfect has a different form. The forms in the right column are thus available to be assigned a different grammatical function. Emphatic renarrated happens to be the paradigmatic contrast that gets expressed by the available form. Notice that this description rests fundamentally on treating the construction *as a whole* as a realization of a combination of lexeme and morphosemantic values and in paradigmatic contrast to other forms of the verb.

and Spencer argue that the perfect (e.g. *săm mislila*) is similarly morphologically nested inside the negated future (e.g. *njama da mislja*) to form the negated future perfect (e.g. *njama da săm mislila*), based in part on the fact that an alternate form for the perfect (*băda mislila*) is also inherited by the negated future perfect (alternate form: *njama da băda mislila*). The posited morphological structure is thus a composition of the two constructions, with *njama da* as a unit. (Notice that the nesting relationship is the same as in the future and future perfect constructions – the negated future perfect form replaces *mislja* in the negated future construction with *săm mislila*.) Syntactically, however, *da* introduces a subordinate clause and serves as prosodic host for *săm* and any other verbal clitics in the clause. The syntactic constituency must therefore be [*njama* [*da săm mislila*]], and *njama* can be separated from *da* (30).

- (30) *Utre po tova vreme njama v nikakäv slučaj da*  
 morning at this time NEG.FUT.3SG in no case DA  
*săm dal statijata na redaktora.*  
 AUX.1SG give.M.PTCP article.F.DEF to editor.M.DEF  
 ‘There is no way I will have given the article to the editor by this time tomorrow.’  
 (\**Utre po tova vreme njama da v nikakäv slučaj săm dal statijata na redaktora.*)

The emphatic renarrated construction is thus interesting because morphophonologically, its parts (auxiliary and lexical verbs) are discrete syntactic atoms and do not cohere with other elements of the Bulgarian verbal complex, as we saw above. At the same time, the construction behaves like a morphological object – it is an exponent of morphosyntactic properties and the realization of a paradigm cell of the verb. This sort of data shows that elements of the (surface-oriented) verbal complex, like *da* and *săm*, form part of paradigmatically structured, morphological constructions, even though they do not form synthetic units.

Although they do not look at Bulgarian, Bonami and Webelhuth (2013) propose an analysis of verb constructions that is designed to handle this kind of data, and specifically, to make periphrases maximally parasitic on existing syntactic and morphological mechanisms, while also optimizing the ability of morphological and syntactic structures to function independently. They build an interface between the lexicalist syntactic framework Head-driven Phrase Structure Grammar (HPSG) and the inferential-realizational morphological framework Paradigm Function Morphology. In periphrases consisting, e.g., of an auxiliary and participle, the auxiliary subcategorizes for the participle, as part of the lexical representation for the periphrastic construction as a whole. On the morphological

side, the component elements of periphrastic constructions are generated by realizational rules – the same mechanism used to generate synthetic forms – including referrals to independently existing forms where relevant. For instance, an auxiliary might be referred to the copula for its form, to the extent that the auxiliary and the copula exhibit the same morphological behavior. On the syntactic side, relating the component parts of the periphrasis via valence allows the model to piggyback on independently existing mechanisms for syntactic functional relations, to the extent that the syntax of the periphrastic construction is the same as ‘normal’ syntax in the language. Crucially, however, and differently from purely syntactic accounts, the periphrases are treated as lexical representations and constructional units, with morphosyntactic values associated to the construction as a whole and inflectionally realized as such.<sup>35</sup> The analysis is thus fundamentally morphological in nature. Popova and Spencer (2013) offer an analysis of Bulgarian verb constructions that applies Bonami and Webelhuth’s approach.

To sum up, the verbal complex exhibits less synthesis in Bulgarian than in Macedonian. Elements like *da* and *ne* are generally not considered part of the clitic cluster because they exhibit signs of forming distinct syntactic and prosodic units, and while *šte* often appears at the head of the cluster, it is likewise positioned there by distinct principles. *Da*, *ne*, and *šte* can also serve as prosodic hosts for the cluster, which is surprising if they are part of the cluster themselves, and *li* can in principle intervene between any elements in the verbal complex (subject to restrictions created by stress requirements). All of these facts suggest that the verbal complex is not a coherent unit structurally. At the same time, the auxiliaries and lexical verb (sometimes also including *da*) form constructional units that cannot be reduced to the individual syntactic elements. This is a different notion of morphologization – one that extends beyond synthesis – but as argued in Section 2, it is one deeply embedded in a lexicalist, inferential-realizational model of inflection. We suspect that morphologization of the Bulgarian type is prerequisite to synthesis – constructional status precedes the development of a fixed order of elements. If the Balkan verbal complex is at least in part a Sprachbund phenomenon, it is unsurprising that Bulgarian should show some morphologization, but not synthesis, since it lies on the periphery of the core contact zone.

---

<sup>35</sup> Bonami (2015) builds on this work, adding the observation that periphrasis has a lot in common with idioms. Following work within HPSG on idioms, he proposes to capture periphrasis through ‘reverse selection’ – i.e., the elements of the periphrasis mutually select each other through bidirectional valence. He also solves a problem of competition between synthesis and periphrasis that the earlier paper faced by having periphrastic realizational rules stated at the level of the paradigm function, rather than at the level of rule blocks. This is able to capture the fact that periphrasis competes with synthesis as a whole, not with individual affixes.


### 4.3 The Serbian verbal complex as a (mostly non-) morphological object

Finally, we look very briefly at some relevant facts of Serbian, as a point of contrast to both Macedonian and Bulgarian. Serbian is well known for having second position clitics, with the order of elements within the clitic cluster as in (31), illustrated in (32).

(31) *li* – AUX – DAT – ACC/GEN – *se* – *je*

(32) *Da li si mu ga dala?*  
 DA Q AUX.2SG 3SG.M.DAT 3SG.M.ACC gave.F.SG  
 ‘Did you give it to him?’

As in the other Slavic languages, *li* is the yes/no question particle, *se* is a particle used for various agent-backgrounding functions (e.g. reflexive, passive, impersonal), and *je* is the third person singular ‘present’ auxiliary clitic. Only the third person singular auxiliary occurs at the end of the cluster; all other present auxiliaries, as well as future and conditional/irrealis auxiliaries, occur in the AUX slot. The negator must occur immediately before the finite verb, whether this is the lexical verb (present tense) or auxiliary (past, future, conditional). Negation thus shows high selectivity with regard to its stem, behaving as a prefix for much the same reasons that Zwicky and Pullum (1983) argue that *n’t* is affixal in English.<sup>36</sup> This means that while its position sometimes lines up with the surface order of the Balkan verbal complex ((33), repeated from (1e)), *ne* is not part of the Serbian clitic cluster, and its positioning is governed by different principles from the other grammatical elements of the verbal complex. This is evident in sentences like (34), where the accusative object clitic *ga* must follow the negated verb.

(33)  *da mu ga dam.*  
 NEG.FUT.1SG SBJV 3SG.DAT 3SG.ACC give.1SG  
 ‘I will not give it to him.’

(34) *Ne vidim ga ovde.*  
 NEG see.1SG 3SG.ACC here  
 ‘I don’t see him/it here.’

<sup>36</sup> This means that unlike in Bulgarian, *ne* cannot itself host the clitic cluster, although *ne* + a finite verb can host the cluster.

Notice that in (33), *da* is also not a part of the cluster. It introduces a subordinate clause and can serve as host for the cluster (here *mu ga*).

In contrast with Macedonian and Bulgarian, the clitic cluster in Serbian is not required to be verb-adjacent. It instead occurs in second position within the clause, with a leftward prosodic host. While the definition of what constitutes ‘second position’ has been the subject of much investigation and exhibits dialectal (perhaps idiolectal) variation, what matters here is that the clitic cluster exhibits promiscuous host selection. This can be seen in (35), where the verbal auxiliary clitic *sam* can attach to whatever stressed constituent happens to be at the beginning of the sentence (marked with brackets), and cannot occur in any other position in the sentence.

- (35) a. [*Kupio*]        *sam*        *Vesni*        *zanimljivu*        *knjigu*.  
           bought.M.SG AUX.1SG Vesna.DAT interesting.ACC book.ACC  
           ‘I bought Vesna an interesting book.’  
       b. [*Vesni*] *sam kupio zanimljivu knjigu*.  
       c. [*Zanimljivu knjigu*] *sam Vesni kupio*.  
       d. \**Sam kupio Vesni zanimljivu knjigu*.  
       e. \*[*Vesni*] [*kupio*] *sam zanimljivu knjigu*.

Moreover, although Serbian has a general preference for *da*-clauses over functionally equivalent infinitival constructions (36a), in infinitival structures clitic climbing is possible (36b).<sup>37</sup> Here *joj ga* climbs out of the lower clause. This underlines the fact that the positioning of elements of the verbal complex does not depend on the verb.

- (36) a. *Marija će da joj ga predstavi*.  
           Marija FUT SBJV 3SG.F.DAT 3SG.M.ACC introduce.3SG  
           ‘Marija will introduce him to her.’  
       b. *Marija joj ga neće predstaviti*.  
           Marija 3SG.F.DAT 3SG.M.ACC NEG.FUT.3SG introduce.INF  
           ‘Marija will not introduce him to her.’

In short, *ne* and *da* do not cohere with the other elements of the (surface-descriptive) verbal complex and while the auxiliary and object clitics (+ *li*) do

<sup>37</sup> For discussion of clitic climbing in the Balkan languages, see Krapova and Cinque (this volume).

form a templatically ordered clitic cluster, they do not form a unit with the verb. In Serbian the verbal complex is thus not a coherent unit of analysis.

It is worth pointing out that there is some allomorphy within the clitic cluster. As noted in Section 2.1, the accusative clitic, which is normally *je*, is realized as *ju* when preceding the third person singular auxiliary clitic *je*. The third person singular auxiliary clitic *je* also drops when preceded by *se*.<sup>38</sup> This is the kind of morphophonological idiosyncrasy that Z&P's criteria diagnose as indicative of morphological objects (their criterion C). However, there is no reason to think that the cluster forms any kind of morphological construction with the verb, certainly not that it realizes a verbal paradigm cell. In the system of Anderson (1992, 2005), the entire cluster (but not together with the verb or any prosodic host) is thus treated as a single morphologically generated object that is positioned with respect to a phrasal domain ('phrasal affixes'). Importantly, positioning of the cluster in Serbian is purely prosodic and syntactic, unlike in the verb-adjacent languages Bulgarian and Macedonian.

## 5 The verbal complex and the Sprachbund – and contact – (re)considered

The evidence presented in Section 3 from Greek, Albanian, and Section 4 from Balkan Slavic shows that as far as the verbal complex is concerned, for all the interesting similarities, there are real and quite significant differences evident across these languages. Particularly noteworthy is the fact that Bulgarian and Macedonian differ with respect to the conditions for preverbal vs. postverbal positioning of weak pronouns, and with regard to interruptability of the pronouns by the question marker *li*. Moreover, they both differ from Serbian in requiring verb-adjacency for the weak pronouns. Greek and Albanian show no possibility for interrupting the weak pronouns.


Thus the geographically adjacent Albanian, Greek, and Macedonian all converge on a set of characteristics for the verbal complex that point towards a high degree of synthesis. Since these are strings that are safely assumed to have once been syntactic in nature (see footnotes 17 and 26, and discussion below regarding Greek), this is evidence of morphologization. Moreover, Albanian and Greek are especially convergent here, showing the highest degree of morphologization in the verbal complex, with Macedonian diverging from them, analytically

---

<sup>38</sup> In usage there is some variation in this regard, but here we describe normative standard practice.

speaking, due to the fact that *li* can intervene between post-verbal weak pronouns and the verb and due to an absence of any telling morphophonological idiosyncrasies involving elements in the Macedonian verbal complex. Furthermore, and quite importantly, the geographically more remote Bulgarian and Serbian diverge from this core clustering within the Balkans that the other languages constitute here. This geographic dimension is suggestive, and would seem to indicate that contact among the speakers of the converging languages is what is responsible for the convergence.

If contact is indeed involved in this convergence – and it is hard to argue otherwise, though we offer some evidence below suggesting what the limits are of the contact-related influence – then it would *appear* that one has to reckon here with contact effects of a nonsuperficial nature and especially with the borrowing of processes, as opposed to specific forms. We examine this claim in some detail in what follows, but first, it is important to see why it is an interesting claim in the first place.

Heine and Kuteva (2005), in arguing for “contact grammaticalization” say that in a situation in which there is “grammaticalization” in a contact zone – as with the parallel morphologization seen with the verbal complexes in the central Balkans – what is borrowed, what is replicated across languages, is the set of processes of grammaticalization evident in the donor language, so that the chain of events that led from, say, a periphrastic construction in one stage of a language to a “grammaticalized” structure in a later stage is replicated in a contact-affected language. Since processes are by nature abstractions, such a view is at odds with the characterization given in Section 1 of borrowing and  contact effects more generally as an inherently surface-oriented phenomenon. In what follows, we attempt to reconcile this surface-oriented view of language contact with the facts of the verbal complex in the Balkans.

First, it is fair to ask whether the processes were indeed borrowed. Among the facts that make the Greek verbal complex appear to be a word-level unit is morphophonology, in particular the behavior of the combination of the second person singular genitive pronoun *su* together with third person pronouns (*su t* > *st*) and the synchronically unexpected voicing of the initial *t* of third person object pronouns after *θa* and *na*. And for Albanian, the portmanteau realizations of the pronouns are particularly telling. These effects, however, are language-specific developments and could not in themselves have been borrowed.

Further, there are empirical chronological considerations arguing against a borrowing (replication) of the processes. One of the pieces of the verbal complex is the invariant marker for futurity, Greek *θα*, Bulgarian *šte*, etc., and this marker in each case derives from a present tense verb that at an earlier stage was fully inflected for person and number of the subject, e.g. 2/3SG Greek *θeleis/θelei* (phonetically, [θelis/θeli]). In Greek, for instance, skipping a few of the intermediate



steps, the progression was from 3SG *θelei na grapsei* for ‘(s)he will write’ (literally ‘(s)he will that (s)he writes’) to *θel na grapsei* to *θe na grapsei* to *θα na grapsei* and ultimately to *θα grapsei* ([θa ɣrapsi]), by a series of regular sound changes and well-motivated analogies, and similar sorts of changes define the emergence of the invariant marker in the other languages as well.<sup>39</sup> Asenova (2002[1989]) gives dates for the appearance of distinct phases in the reduction of this future auxiliary that shows significant differences in the period at which each step in the reduction is to be found in each language. Specifically, she states that the *θe na grapsei* reduction is found from the 14th century in Greek, whereas the corresponding Balkan Slavic *šte/ke da piša(m)* is found from the 16th century, Albanian *do të shkrojë* from the early 18th century, and Balkan Romance *o să scriu* from the mid-18th century.<sup>40</sup> And the fully reduced form without the subordinating element *na/da/të/să* is attested from the early 16th century for Greek, the early 15th century for Slavic, the mid-18th century for Albanian, and the early 19th century for Romanian. These dates do not line up as we might expect if the reductive processes were borrowed (replicated) across the languages. To take the earliest fully reduced form – Slavic – as the starting point, just for the sake of argument,<sup>41</sup> one has to admit that once a process of reduction has done its work, so to speak, what remains is not the process itself but rather the result of the process. Thus after the 15th century in Slavic territory, what would have been available to be passed to another language was not the process of reduction but the outcome of the reductive process. Thus it is hard to see how the process of reduction could be borrowed as only the results are evident, not the process itself. We thus conclude that we must reckon with a certain degree of historical independence to the specific developments leading to morphologized verbal complexes in the different languages here.

In defense of a contact-based approach, one might argue that it was not the process of reduction that was borrowed but the idea of reduction. While it is hard to see how an abstraction like such an “idea” could be borrowed (as if it

<sup>39</sup> See Joseph and Pappas (2002) for detailed discussion of this progression.

<sup>40</sup> These last two dates may be affected, of course, by the rather late attestation of these languages and the scarcity of older materials in general; substantial documents from Albanian and Romanian are known from roughly only the mid-16<sup>th</sup> century (1555 and 1521, respectively) and then of a somewhat limited nature.

<sup>41</sup> There is admittedly an anomaly in Asenova’s chronology for Slavic in that the reduced form occurs earlier in her reckoning than the fuller form; this is most likely a matter of attestation as far as the fuller form is concerned. For what it is worth, the fuller form in Greek, while occurring early enough, is not all that robustly attested, in part, perhaps, because it was ambiguous between a future reading and a volitional reading (‘he will write’ versus ‘he wants to write’), as also in Slavic.

were “in the air” in a speech community, so to speak), we do give a reasonable scenario below as to how that might happen, working from the evident results of the reductive process(es).

Still, one might ask whether it is perhaps the case that contact is simply irrelevant here and whether it might not be more reasonable to say that the parallel synthesis in the verbal complexes of the different languages is just a coincidence. To this, we say no. As we see it, such a claim amounts to a denial of the geography and moreover it strains credulity when one considers the significant number of convergent features in the Balkan languages in general and especially among Greek, Albanian, and Macedonian, features which taken together are what motivate the notion of a Sprachbund involving these languages in the first place.<sup>42</sup>

So it must be asked how the convergence in the verbal complex came to be. Vital to our account is a recognition that what can be borrowed is, to reiterate, present in surface structure, in the output of the grammar that speakers produce. By this we mean not the actual surface forms, but rather patterns that are derivable from surface forms, patterns that can be “read” off of the surface. That is, in a social context of fairly intense contact, where there is mutual multilingualism so that speakers of one language have a reasonable command of the language(s) of their neighbors, and have a sense of what the pieces of the other language do, cross-language matching up of elements at the templatic level can easily occur. In this way, patterns can be transferred between languages, being read off of output essentially by speakers imposing an analysis on the elements in the other language that correspond to known entities in their own language. This matching and pattern imposition process does involve some level of abstraction, to be sure, but crucially what is borrowed is not a process per se but rather an analysis. As we envision this process, it is abstract, but it is not deeply syntactic; speakers are not borrowing strictly surface form but rather patterns – surface *structure*, that is – evident from the surface forms. The surface here gives a target structure for speakers of a language to aim at, in the case of the verbal complex yielding a particular surface ordering of elements.

At this point, it is useful to recall Kopitar’s (1829) dictum from Section 1, regarding there being one grammar but three lexicons for the Balkan languages

---

<sup>42</sup> See Sandfeld (1930) for what is still the best collection of these convergent features; Friedman (2006) is perhaps the best concise statement of the evidence for the Balkan Sprachbund, though see also Joseph (2010), and Friedman and Joseph (in prep). These features are not just structural in nature but also lexical, and among the lexical items are numerous discourse markers, indicating that the speakers of the different languages were indeed speaking to one another; see Friedman and Joseph (2014) for a discussion of these conversationally based loanwords (“ERIC loans”, in their formulation, for those “Essentially Rooted In Conversation”).

he was examining. The “one grammar” here can be a template, an ideal surface string,<sup>43</sup> for the order of elements, and that gets realized with different lexical material. An informal characterization of what was going on, as we see it, is that essentially speakers were saying “OK, I recognize that you treat, for instance, your weak object pronouns as verb-adjacent; well, I can do that too in my language”. And by doing so, such a speaker was either accommodating to (= adopting) the particular feature of the other, or was selecting from among variants in his/her own language that match the other’s pattern; in either case, for either motivation, there is a facilitating of communication between the speakers through the use of a shared structure.

Then, once the ordering of elements is fixed in this way, ordinary processes of language change can operate, much as Joseph (2001) argues with respect to the Greek future and as suggested earlier in this section. The workings of such processes can lead independently to synthetic structure to greater or lesser degrees, for reasons that are particular to specific speech communities and their social (etc.) circumstances. Nonetheless, the original impetus will have been language contact and the borrowing of structure will have been achieved through the cross-language matching of elements of structure in comparable entities in the respective languages.

There is also the possibility for cross-language analogies to play a role. Such a mechanism may be behind the possible absence of the subjunctive marker *të* in the future tense in Albanian, where the pattern of two independent but related variants for the future in Greek could have been the model for its allowable absence in Albanian, if Albanian speakers recognized a relation between the two variants and were in a position to equate them with their own future tense<sup>44</sup>:

Greek:    ~~the~~ na grapso : ~~the~~ grapso  
                        ::

Albanian: do të shkruaj : X,  $X \rightarrow$  do shkruaj

A similar mechanism, perhaps with Albanian as the model, though Greek would have been possible too, may have been at work in Macedonian, where the standard language now has simply the invariant future marker *ke* with a finite verb to

43 Note that for Kopitar, working at the time he did, there was not a coherent notion of deep structure and surface structure.

**44** These future variants are of independent origin within Greek; the type with subjunctive marker *na* shows the regular replacement of the earlier infinitive by a finite clause introduced by *na*, while the type without *na* resulted from a reanalysis of an infinitival form as finite, due to the merger of the infinitive ending with 3SG ending. This is a Greek-particular development, not found in other languages.

form the future, e.g. *ķe dojdām* ‘I will go’ whereas dialectally the modal marker *da* can occur as well, e.g. *ķe da dojdām*. Similarly, the emergence of evidentiality dialectally in Aromanian under heavy Macedonian influence, as described by Friedman (1994), is based on a recognition by Aromanian speakers of parallel structures in Macedonian that could be used in a novel way in Aromanian.

In the general view of Balkan language contact being advocated here, the convergence becomes a mix of contact-induced impetus and some borrowing with language-particular developments. For a complex contact situation like the Balkans, such a scenario is probably closer to the truth than any single-cause account would be, and it mirrors the multiple-causation account given in Joseph (1983: Ch. 7) for the parallel infinitive-loss and replacement developments in the various Balkan languages.

In this regard too, one can also think of contact scenarios, such as that proposed by Friedman (1999) for the development of evidentiality in the Balkans, in which what occurs might be termed “contact-induced ‘enhancement’ of native possibilities”. Friedman argues that Turkish influence did play a role but not through the importation of completely new material and new categories from Turkish into Balkan Slavic languages, but rather by native tendencies already present in Slavic being enhanced and given a chance to develop through contact with Turkish, a language with grammatical expression of evidentiality. By contact, outright borrowing (actually, calquing) seems to have been involved in the further spread of evidentiality to Aromanian from a Macedonian model.

All of this suggests that the areal pattern of morphologization and synthesis within the verbal complex follows in parallel to other contact patterns that we find in the same region. So in some sense, there is nothing particularly special about the areal distribution of the verbal complex. It does, however, highlight the way in which speakers in contact situations can be sensitive to the ordering and other surface-structural properties of grammatical elements. When speakers recognize and import structural properties that are related to morphologization (e.g. verb-adjacency of weak object pronouns), the result may be convergent grammaticalization. We hypothesize that it is exactly this kind of observational power that has led to the areal distribution of morphologization in the verbal complex.

## 6 Conclusions

In this paper we have explored theoretical issues raised by the Balkan verbal complex by virtue of its positioning at the intersection of morphology and syntax. Parallelisms in the content and order of functional elements formed the basis for

the cross-linguistic comparison and motivated the idea of the verbal complex as a Balkan contact phenomenon in the first place. However, the differences across languages are ultimately at least as interesting as the similarities. We have argued for a broad notion of morphologization, as well as the idea that morphologization processes proceed in piecemeal fashion, resulting in the frequent occurrence of both morphological and syntactic properties in constructions synchronically. This conceptualization has allowed us to explore the question of morphologization in the Balkan verbal complex in a relatively fine-grained (if necessarily non-comprehensive) way. While there are idiosyncratic differences in the verbal complexes of the individual languages, the overarching pattern that emerged was one of decreasing morphologization as we move outwards from the core Sprachbund contact zone. We find a high degree of synthesis in Albanian to Greek, with progressively less morphologization in Macedonian, Bulgarian and Serbian. This offers an interesting window into the questions of contact that have to do with morphologization processes. We argued here that while contact is very likely involved in the pattern of convergent morphologization, it is not the *process* of morphologization that is borrowed from one language to another (contra Heine and Kuteva 2005). Rather, we need assume nothing more than that the *outcomes* of morphologization in the source language serve as the basis for structural calquing in the borrowing language. So while ‘extent of morphologization’ is a rather abstract kind of contact effect, we ultimately find that this parallels and is no more exotic than other contact outcomes that are firmly established as part of the Balkan Sprachbund. We offer the verbal complex as a member of that canon.

## Postscript: Object doubling and the verbal complex

By way of a postscript, we look briefly at object doubling, a phenomenon that touches on the verbal complex, and the issue of morphologization, but in a different light from the discussion above. Our observations here are by necessity speculative and in need of further investigation. But we offer them as consideration of what implications morphologization of the verbal complex may have for broader issues and debates.

As noted above in Section 3, ~~Schumacher and Matzinger (to appear 2017)~~ observe that in Albanian, the weak object pronouns can co-occur with full nominals as objects, either nouns or strong forms of pronouns, as in (37):

- (37) a. *E pashë Gjonin.*  
           him.ACC.WK saw.1SG John.ACC  
           ‘I saw John.’

- b. *Më pa mua Gjoni.*  
 me saw.3SG me.ACC.STR John.NOM  
 ‘John saw me.’

This construction is actually found in other Balkan languages; examples are given in (38) from Greek, in (39) from Macedonian, and in (40) from Bulgarian (though see below for clarification of the status of such examples from Bulgarian):

- (38) a. *To pino efxaristos ena uzaki.*  
 3SG.WK.ACC drink.1SG gladly an.N.ACC ouzo.DIM  
 ‘I would gladly have an ouzo.’ (literally: “it I-drink gladly an ouzo”)
- b. *To pino afto efxaristos*  
 3SG.ACC drink.1SG it.STR.ACC gladly  
 ‘I would drink it gladly’ (literally: “it I-drink this gladly”)
- (39) *Mu go davam molivot na momčeto.*  
 3SG.M.DAT 3SG.M.ACC give.1SG pencil.DEF to boy.DEF  
 ‘I give the pencil to the boy.’ (literally: “to-him it I-give pencil-the to boy-the”)
- (40) *Kučeto ja goni edna kotka.*  
 dog.DEF it.ACC.F chases one cat.F  
 ‘It’s the dog that is chasing a cat’- (Guentchéva 1994:111)

This phenomenon, known in the literature as “Clitic Doubling” or “Object Reduplication” or “Object Doubling”, has attracted much attention over the years; the vast literature is summarized admirably and insightfully in Anagnostopoulou (2006); see also Kallulli (this volume). One reason for the considerable interest is that object doubling presents analytic challenges to assumptions and claims made within generative syntax; in particular, sentences like those in (37) – (40) appear to show two potential accusative case-marked entities – the weak object pronoun (e.g. *e* in (37)) and the full nominal object (e.g. *Gjonin* in (37)) – but only one accusative case-licensing entity, namely the verb. The Balkan data thus offer a typological point of contrast to Spanish and other Romance languages where object doubling is possible only when there are two distinct case-licensors (e.g. a verb that licenses the clitic object and a preposition that licenses the full nominal). The Balkan facts also contrast with the pattern in Welsh, discussed by Sadler (1997), where pronominal clitics block full noun phrases, and object doubling does not occur in this context.

There turn out to be a number of ways in which the relevant object doubling facts can be handled, in part depending on other sorts of assumptions that

AU: Please check the punctuation marks behind the sentences here and possibly unify (fullstop yes/no, order of fullstop and closing quotation mark).

are made within a given theoretical framework for syntax. However, assessing these requires a full evaluation of these other assumptions and the constructs they require and depend on. Such an evaluation is beyond the scope of this paper, and in any case, it is covered well in Anagnostopoulou (2006). However, among the possibilities is one that bears on the morphological versus syntactic status of the weak object pronouns. In particular, if case licensing is a syntactic requirement and the weak object pronouns are morphological in nature, occurring say as agreement markers,<sup>45</sup> then one could in principle exclude the weak object pronouns from a case-licensing requirement. So we cannot help but wonder whether the theoretical issues raised by Balkan object doubling are in fact obviated by the independent evidence in this paper for the reanalysis of object pronouns as affixes to the verb. In other words, is there object doubling at all, or instead an emergent agreement system? We note that object doubling is most grammaticalized exactly in those languages where there is also the most synthesis among elements of the verbal complex, including object pronouns.

Any full exploration of this idea would need to account for object doubling in Bulgarian. As noted in Section 4 above, we find little evidence in support of an affixal analysis of weak object pronominals in Bulgarian, in part because they can be separated from the verb by *li*, which is not structurally part of the clitic cluster. However, we also note that object doubling is not required in Bulgarian except in the impersonal existential use of *ima/njama* ‘there is / there is not’, literally ‘has/not.has’. Thus in the Bulgarian sentence in (41), no weak pronouns doubling the objects are required.

- (41) *Davam moliva na momčeto.*  
 give.1SG pencil to boy.DEF  
 ‘I give the pencil to the boy.’

However, object doubling can, but need not, be used to disambiguate case relations (Stojanov 1983: 192–193). More important here is the claim in Leafgren (2002: 197) that object doubling in Bulgarian serves a pragmatic function, marking *aboutness*, most typically in a contrastive setting (i.e., ‘about X as opposed to Y’).

---

<sup>45</sup> They could be marking agreement in that their properties (gender, number, case) match those of the full nominal, but they would be atypical agreement markers – different for instance from the agreement that verb endings show with the person and number of the subject or the agreement in gender, number, and case between articles or adjectives and nouns – in being sometimes optional.

- (42) *Banan ne običam da go jam.*  
 banana NEG like.1SG that 3SG.M.ACC eat.1SG  
 ‘I don’t like to eat bananas.’ (Leafgren (2002: 176); context = discussion of markets)

In this way, the presence of a doubled pronoun serves as an overt marker of topicality.<sup>46</sup> Object doubling in Bulgarian is thus fundamentally discourse-oriented, rather than grammatical. This is exactly the pattern that we would expect if ‘object doubling’ and morphologization were related.

Ultimately, we cannot explore this idea in any detail here and leave it for future work. However, we raise the question of the proper analysis of the object pronominals as a way to demonstrate that the question of whether the verbal complex is morphologized bears on larger theoretical issues. We think that the relevance of verbal complex morphologization for licensing of object doubling is an issue that merits further work. Moreover, it underlines the importance of considering the structural properties of the relevant elements, since these may not be obvious from the surface string of elements.


## References

- Ackema, Peter, and Ad Neeleman. 2001. “Competition between syntax and morphology.” In *Optimality-theoretic syntax*, edited by Géraldine Legendre, Jane Grimshaw and Sten Vikner, 29–60. Cambridge, MA: The MIT Press.
- Ackerman, Farrell, and Philip LeSourd. 1997. “Toward a lexical representation of phrasal predicates.” In *Complex predicates*, edited by Alex Alsina, Joan Bresnan and Peter Sells, 67–106. Stanford, CA: CSLI.
- Ackerman, Farrell, and Gregory T. Stump. 2004. “Paradigms and periphrastic expression: A study in realization-based lexicalism.” In *Projecting morphology*, edited by Louisa Sadler and Andrew Spencer, 111–158. Stanford, CA: CSLI.
- Ackerman, Farrell, Gregory T. Stump, and Gert Webelhuth. 2011. “Lexicalism, periphrasis and implicative morphology.” In *Non-transformational syntax: Formal and explicit models of grammar*, edited by Robert D. Borsley and Kersti Börjars, 325–358. Malden, MA: Wiley-Blackwell.
- Aikhenvald, Alexandra Y. 2002. “Typological parameters for the study of clitics, with special reference to Tariana.” In *Word: A cross-linguistic typology*, edited by R.M.W. Dixon and Alexandra Y. Aikhenvald, 42–78. Cambridge: Cambridge University Press.
- Alexander, Ronelle. 1994. “The Balkanization of Wackernagel’s law.” *Indiana Slavic Studies* 7: 1–8.

---

<sup>46</sup> The geographic distribution is interesting too, in that there is a west-to-east trend, with a decrease in frequency as one moves east within Bulgarian-speaking territory. The end result is that object doubling is relatively infrequent in eastern Bulgarian (Keremedčieva 1993: 297–299).



- Anagnostopoulou, Elena. 2006. "Clitic doubling." In *Blackwell companion to syntax*, vol. 1, edited by Martin Everaert, Henk van Riemsdijk, Rob Goedemans and Bart Hollebrandse, 519–581. Oxford: Blackwell.
- Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.
- . 1995. "Rules and constraints in describing the morphology of phrases." In *CLS 31: Papers from the 31st Regional Meeting of the Chicago Linguistic Society*, vol. 2, edited by Audra Dainora, Rachel Hemphill, Barbara Luka, Barbara Need and Sheri Pargman, 15–31. Chicago: Chicago Linguistic Society.
- . 2005. *Aspects of the theory of clitics*. Oxford: Oxford University Press.
- Andrews, Avery D. 1990. "Unification and morphological blocking." *Natural Language and Linguistic Theory* 8: 507–557.
- Asenova, Petja. 2002[1989]. *Balkansko ezikoznanie: Osnovni problemi na balkanskija ezikov sąjuz*. Sofia: Faber.
-  e, Gabriela and Alexandru Mardale. This volume. The Romanian subjunctive from a Balkan perspective.
- Bonami, Olivier. 2015. "Periphrasis as collocation." *Morphology* 25(1): 63–110.
- Bonami, Olivier, and Gilles Boyé. 2007. "French pronominal clitics and the design of Paradigm Function Morphology." In *On-line proceedings of the Fifth Mediterranean Morphology Meeting (MMMM5), Fréjus, 15–18 September 2005*, edited by Geert Booij, Luca Ducceschi, Bernard Fradin, Emiliano Guevara, Angela Ralli and Sergio Scalise, 291–322. Bologna: Università degli Studi di Bologna.
- Bonami, Olivier, and Pollet Samvelian. 2009. "Inflectional periphrasis in Persian." In *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar*, edited by Stefan Müller, 26–46. Stanford, CA: CSLI.
- Bonami, Olivier, and Gert Webelhuth. 2013. "The phrase-structural diversity of periphrasis: A lexicalist account." In *Periphrasis: The role of syntax and morphology in paradigms*, edited by Marina Chumakina and Greville G. Corbett. Oxford: Oxford University Press and British Academy.
- Börjars, Kersti, Nigel Vincent, and Carol Chapman. 1997. "Paradigms, periphrases and pronominal inflection: A feature-based account." In *Yearbook of morphology 1996*, edited by Geert Booij and Jaap van Marle, 155–180. Dordrecht: Kluwer.
- Bošković, Željko. 2001. *On the nature of the syntax-phonology interface: Cliticization and related phenomena*. Amsterdam: Elsevier.
- Bresnan, Joan. 2001. "Explaining morphosyntactic competition." In *The handbook of contemporary syntactic theory*, edited by Mark Baltin and Chris Collins, 11–44. Malden, MA: Blackwell.
- Bresnan, Joan, and Sam Mchombo. 1995. "The Lexical Integrity Principle: evidence from Bantu." *Natural Language and Linguistic Theory* 13(2): 181–254.
- Brown, Dunstan, Marina Chumakina, Greville Corbett, Gergana Popova, and Andrew Spencer. 2012. "Defining 'periphrasis': Key notions." *Morphology* 22(2): 233–275.
- Brown, Dunstan, and Andrew Hippisley. 2012. *Network morphology: A defaults-based theory of word structure*. Cambridge: Cambridge University Press.
- Chomsky, Noam. 1970. "Remarks on nominalization." In *Readings in English transformational grammar*, edited by R. Jacobs and P. Rosenbaum, 184–221. Waltham, MA: Ginn.
- Embic, David, and Roumyana Izvorski. 1997. "Participle-auxiliary word orders in Slavic." In *Annual workshop on formal approaches to Slavic linguistics*, vol. 4: *The Cornell meeting, 1995*, edited by Wayles Browne, Ewa Dornisch, Natasha Kondrashova and Draga Zec, 210–239. Ann Arbor: Michigan Slavic.

- Englund, Birgitta. 1977. *Yes/no-questions in Bulgarian and Macedonian*. Stockholm: Almqvist & Wiksell International.
- Franks, Steven, and Tracy Holloway King. 2000. *A handbook of Slavic clitics*. Oxford: Oxford University Press.
- Friedman, Victor A. 1993. "Macedonian." In *The Slavonic languages*, edited by Bernard Comrie and Greville Corbett, 249–305. London: Routledge.
- . 1994. "Surprise! Surprise! Arumanian has had an admirative!" *Indiana Slavic Studies* 7: 79–89.
- . 1999. "Evidentiality in the Balkans." In *Handbuch der Südosteuropa-Linguistik*, edited by Uwe Hinrichs, 519–544. Wiesbaden: Harrassowitz.
- . 2006. "The Balkans as a linguistic area." In *Encyclopedia of language and linguistics*, edited by Keith Brown, 657–672. Oxford: Elsevier.
- Friedman, Victor A., and Brian D. Joseph. 2014. "Lessons from Judezmo about the Balkan Sprachbund and contact linguistics." *International Journal of the Sociology of Language* 226: 3–23.
- . in prep. *The Balkan languages*. Cambridge: Cambridge University Press.
- Gołąb, Zbigniew. 1954. "Funkcja syntaktyczna partykuly *da* w językach południowosłowiańskich (bułgarskim, macedońskim i serbo-chorwackim)." *Biuletyn Polskiego Towarzystwa Językoznawczego* 13: 67–92.
- Hauge, Kjetil Rå. 1999[1976]. "The word order of predicate clitics in Bulgarian." *Journal of Slavic Linguistics* 7(1): 89–137.
- Heine, Bernd, and Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Joseph, Brian D. 1983. *The synchrony and diachrony of the Balkan infinitive: A study in areal, general, and historical linguistics*. Cambridge: Cambridge University Press.
- . 2001. "Is there such a thing as 'grammaticalization'?" *Language Sciences* 23(2–3): 163–186.
- . 2002. "Defining 'word' in Modern Greek: a response to Philippaki-Warbuton & Spyropoulos 1999." In *Yearbook of morphology 2001*, edited by Geert Booij and Jaap Van Marle, 87–114. Dordrecht: Kluwer.
- . 2010. "Language contact in the Balkans." In *Handbook of language contact*, edited by Ray Hickey, 618–633. Oxford: Blackwell.
- Joseph, Brian D., and Panayiotis A. Pappas. 2002. "On some recent views concerning the development of the Greek future system." *Byzantine and Modern Greek Studies* 26: 247–273.
- Kallulli, Dalina. This volume. Balkan Clitic Doubling Revisited: Micro-Variation, Typological Generalizations, and a Universal.
- Keramić, Iliana. 1991. *Porăt na Ropkata*. Sofia: MicroPrint.
- Kiparsky, Paul. 2005. "Blocking and periphrasis in inflectional paradigms." In *Yearbook of morphology 2004*, edited by Geert Booij and Jaap van Marle, 113–135. Dordrecht: Springer.
- Koneski, Blaže. 1967. *Gramatika na makedonskiot literaturni jazik*. Skopje: Kultura.
- Kopitar, Jernej. 1829. "Albanische, walachische und bulgarische Sprache." *Jahrbücher der Literatur* 46: 59–106.
- Kramer, Christina. 1986. *Analytic modality in Macedonian*. Munich: Otto Sagner.
- Krapova, Iliana, and Guglielmo Cinque. This volume. "Universal constraints on Balkanisms. A case study: The absence of clitic climbing." In *Balkan syntax and (universal) principles of grammar*, edited by Iliana Krapova and Brian D. Joseph. Berlin: Mouton de Gruyter.

- Leafgren, John. 2002. *Degrees of explicitness: Information structure and the packaging of Bulgarian subjects and objects*. Amsterdam: John Benjamins.
- Legendre, Géraldine. 1996. Clitics, verb (non-)movement and optimality in Bulgarian. In *Technical Report JHU-CogSci-96-5*: Johns Hopkins University, Department of Cognitive Science.
- . 1999. "Morphological and prosodic alignment at work: The case of South Slavic clitics." In *WCCFL 17: The proceedings of Seventeenth West Coast Conference on Formal Linguistics*, edited by Kimary N. Shahin, Susan Blake and Eun-Sook Kim, 436–450. Stanford, CA: CSLI.
- . 2001a. "Masked second-position effects and the linearization of functional features." In *Optimality-theoretic syntax*, edited by Géraldine Legendre, Jane Grimshaw and Sten Vikner, 241–278. Cambridge, MA: The MIT Press.
- . 2001b. "Morphological and prosodic alignment of Bulgarian clitics." In *Optimality Theory: Syntax, phonology and acquisition*, edited by Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer, 423–462. Oxford: Oxford University Press.
- Maslov, Jurij. 1956. *Očerk bolgarskoj grammatiki*. Moskva: Izdatel'stvo Literatury na Inostrannyjazykax.
- Mišeska Tomić, Olga. 1996. "The Balkan Slavic clausal clitics." *Natural Language and Linguistic Theory* 14(4): 811–872.
- . 2006. *Balkan sprachbund morpho-syntactic features*. Dordrecht: Springer.
- Newmark, Leonard, Philip Hubbard, and Peter R. Prifti. 1982. *Standard Albanian: A reference grammar for students*. Stanford, CA: Stanford University Press.
- Perlmutter, David. 1971. *Deep and surface structure constraints in syntax*. New York: Holt, Reinhart and Winston.
- Philippaki-Warbuton, Irene, and Vassilios Spyropoulos. 1999. "On the boundaries of inflection and syntax: Greek pronominal clitics and particles." In *Yearbook of morphology 1998*, edited by Geert Booij and Jaap Van Marle, 45–72. Dordrecht: Kluwer.
- Popova, Gergana. 2010. "Features in periphrastic constructions." In *Features: Perspectives on a key notion in linguistics*, edited by Anna Kibort and Greville G. Corbett, 166–184. Oxford: Oxford University Press.
- Popova, Gergana, and Andrew Spencer. 2013. "Relatedness in periphrasis: A paradigm-based perspective." In *Periphrasis: The role of syntax and morphology in paradigms*, edited by Marina Chumakina and Greville G. Corbett, 191–225. Oxford: Oxford University Press and British Academy.
- Poser, William J. 1992. "Blocking of phrasal constructions by lexical items." In *Lexical matters*, edited by Ivan A. Sag and Anna Szabolcsi, 111–130. Stanford, CA: CSLI.
- Rudin, Catherine. 1988. "On multiple questions and multiple WH fronting." *Natural Language and Linguistic Theory* 6: 445–501.
- Rudin, Catherine, Christina Kramer, Loren Billings, and Matthew Baerman. 1999. "Macedonian and Bulgarian *li* questions: Beyond syntax." *Natural Language and Linguistic Theory* 17(3): 541–585.
- Sadler, Louisa. 1997. Clitics and the structure-function mapping. In *Proceedings of the LFG '97 conference*, edited by Miriam Butt and Tracy Holloway King: CSLI Publications Online.
- Sadler, Louisa, and Andrew Spencer. 2000. "Syntax as an exponent of morphological features." In *Yearbook of morphology 2000*, edited by Geert Booij and Jaap van Marle, 71–96. Dordrecht: Kluwer.
- Samvelian, Pollet. 2007. "A (phrasal) affix analysis of the Persian ezafe." *Journal of Linguistics* 43(3): 605–645.

- Sandfeld, Kristian. 1930. *Linguistique balkanique: Problèmes et résultats*. Paris: E. Champion.
- ~~Schumacher, Stefan, and Joachim Matzinger. to appear 2017. "The morphology of Albanian." In *Handbook of comparative and historical Indo-European linguistics*, edited by Jared Klein, Brian D. Joseph and Matthias Fritz. Berlin: de Gruyter Mouton.~~
- Simpson, Jane, and Margaret Withgott. 1986. "Pronominal clitic clusters and templates." In *Syntax and semantics 19: The syntax of pronominal clitics*, edited by Hagit Borer, 149–174. New York: Academic.
- Spencer, Andrew. 2000. "Verbal clitics in Bulgarian: A Paradigm Function approach." In *Clitics in phonology, morphology and syntax*, edited by Birgit Gerlach and Janet Grijzenhout, 355–386. Amsterdam: John Benjamins.
- . 2001. "The paradigm-based model of morphosyntax." *Transactions of the Philological Society* 99(2): 279–313.
- . 2003. "Periphrastic paradigms in Bulgarian." In *Syntactic structures and morphological information*, edited by Uwe Junghanns and Luka Szucsich, 249–282. New York: Mouton de Gruyter.
- . 2012. "Sentence negation and periphrasis." In *Periphrasis: The role of syntax and morphology in paradigms*, edited by Marina Chumakina and Greville G. Corbett, 227–266. Oxford: Oxford University Press and British Academy.
- Spencer, Andrew, and Ana R. Luís. 2012. *Clitics*. Cambridge: Cambridge University Press.
- Stojanov, Stojan. 1983. *Gramatika na sǎvremenija bǎlgarski knižoven ezik, tom II: morfologija*. Sofia: BAN.
- Stump, Gregory T. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press.
- Zwicky, Arnold M. 1977. *On clitics*. Bloomington, IN: Indiana University Linguistics Club.
- . 1990. "Inflectional morphology as a (sub)component of grammar." In *Contemporary morphology*, edited by Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer and John R. Rennison, 217–236. New York: Mouton de Gruyter.
- . 1992. "Some choices in the theory of morphology." In *Formal grammar: Theory and implementation*, edited by Robert Levine, 327–371. Oxford: Oxford University Press.
- Zwicky, Arnold M., and Geoffrey K. Pullum. 1983. "Cliticization vs. inflection: English n't." *Language* 59(3): 502–513.