7 Syntax

7.0 Introduction

Kopitar's much-noted insight concerning the Balkans (see §2.2.1) essentially defines what may be called, as Krishnamurti 2005: 485 terms it, a "translation area" (see also Choudhuri 2010: 113). In such a linguistic area, there appears to be a basic "template" for syntactic frames and one needs only to plug language-particular morphology and lexical items into the slots in the templates to produce appropriate sentences and utterances in the individual languages in the area. To some extent, this is evident in the phraseological calques discussed in §4.3.10, and it has been remarked upon for other sprachbunds, especially the sprachbund of Kupwar village (cf. §3.2.1.6, footnote 53, §3.2.2.10, and §3.4.2.1), as described by Gumperz & Wilson 1971, where the term "intertranslatability" (p. 154) is used; as they put it, "the codes used in code-switching situations in Kupwar have *a single surface syntactic structure*" (p. 155), a characterization reminiscent of Kopitar's "one grammar … with three lexicons" (cf. §2.2.1). Using the concept of templates for sentence structure means focusing attention on the actual realization of sentences, i.e., on the surface structure, as Gumperz and Wilson say explicitly. And, indeed, many syntactic convergences reported for other sprachbunds besides the Balkans and Kupwar involve superficial aspects of sentence patterning. For instance, regarding South Asia, Masica 1976 has mentioned SOV word order and dative subject constructions, while for Meso-America, Campbell et al. 1986 have noted the occurrence of nominal possession of the type *his-dog the-man* and of certain word order patterns for major sentential constituents. This surface orientation is consistent with the importance that must be placed on intimate and especially conversationally based lexical borrowing (ERIC loans – cf. §4.1 and especially §4.3) in understanding the conditions that give rise to a sprachbund, as emphasized in §3.2.1.7, since lexical borrowing is the quintessential surface-based phenomenon in language contact.[1]

Accordingly, in this chapter, we focus on these syntactic templates, organizing the presentation around the key elements of syntactic structure, as discussed below. The account here is necessarily selective. We attend mainly to phenomena in the

1 Cf., however, Aikhenvald 2012: 77–78, who discusses interaction between Tariana, an Arawakan language, and East Tucanoan languages in the Vaupés region of Amazonia. These groups practice linguistic exogamy. Children all grow up multilingual and identify with the father's language, and men must marry women whose fathers speak a different language. Among these groups, lexicon is emblematic and kept more or less strictly separate (with a tiny number of exceptions), while syntax is almost completely (albeit not entirely) convergent. One could argue that this is also the situation in Kupwar, without the exogamy but with different social restrictions.

realm of syntax that are Balkanologically interesting, and we incorporate observations made by others over the years that have syntactic relevance. Attention is thus given to those matters of detail where the languages are significantly similar or different, in keeping with our interest (see §3.3) in the divergences as well as the convergences in the Balkans. In this way, we are working towards a true comparative Balkan grammar (cf. Kazazis 1965, 1967; Friedman 2000d).

7.1 The Enterprise of Balkan Comparative Syntax

We begin with some observations about the enterprise of doing Balkan comparative syntax, i.e., cross-language syntactic comparison involving the Balkan languages.

There is a significant amount of work in this area that follows Sandfeld's general theoretical orientation and thus incorporates philological, structuralist, functional, and typological approaches. The early twenty-first century, however, saw a flurry of work in the framework of generative grammar aimed at comparing various Balkan languages in terms of their syntactic structures as modeled in that theoretical school. Such an approach is best exemplified in the work appearing in Rivero & Ralli 2001, and the references therein, as well as, from an earlier era in generative syntax, Joseph 1983a, Rudin 1986, 1988ab, and others. This work was driven in large part by an interest in so-called universal grammar, the goals of which can be succinctly stated as the determination of the ways in which all languages are alike and the ways in which they differ.[2] To that end, comparative syntax plays an important role: if only one language were looked at, it would be impossible to determine whether any property that was identified as putatively universal was in fact significant in the relevant way. Thus cross-linguistic comparison is at the heart of the enterprise of universal grammar when practiced in a principled, scientific fashion.[3]

It is worth asking, however, how such work advances the goals of Balkan linguistics as defined herein, given not only the areal basis for the discipline but also the paramount importance accorded to language contact in the formation of the object of study, the Balkan sprachbund. That is, do the results that have been obtained from generative comparisons of the syntax of various Balkan languages reveal anything beyond what might be found from a comparison of any arbitrary set of languages chosen on a basis other than geography (and the history associated with it), e.g., Quechua, Mandarin, and Dyirbal? Do they reveal anything that is significant from a Balkanological perspective? If the focus of attention is so-called

2 This is the formulation given by David Perlmutter in class lectures at MIT in the 1970s, and, as a formulation, it continues to be cited, although the assumptions underlying it are being found increasingly questionable (Evans & Levinson 2009; Levinson & Evans 2010).

3 We can add that without historical grounding, an examination of more than one language is equally uninformative. Thus, for example, a phenomenon shared by merely two languages, and moreover when those two languages are French and Spanish, which are known to be descended from a single ancestor, is hardly indicative of a universal in and of itself.

deep aspects of the syntax, such as underlying structures and parameter settings, involving the extension to these languages of analytic principles and constructs that are assumed to be part of a putative universal grammar, then, given the surface-oriented nature of contact phenomena, could any similarities that might emerge from such an investigation possibly have anything to do with language contact? An analysis assimilating Balkan syntactic structures into presumed universal principles is really a kind of typological comparative syntax (albeit in generative garb) of languages that happen to be in the Balkans, not comparative Balkan syntax in particular. Such approaches tell us nothing of the history of the phenomena in question, and, by attempting to explain the features under consideration in terms of so-called deep structures, do not take into account the effect of surface structures in one language on surface structures of another. Insights into the Balkan sprachbund as a contact-induced phenomenon are thus absent from such approaches.

Such investigations are not without value in the broader field of general linguistics, but their value does not lie in the area of Balkan linguistics, per se. For the purposes of relating any results of comparative syntax to the Balkan sprachbund, the most enlightening comparisons involve superficial phenomena that can be transferred in language contact situations. On the other hand, comparisons involving parametric variation or parallels at deeper levels of structure may shed light on universal grammar, but they do not further any understanding of the contact that created the Balkan sprachbund.

Of paramount importance here is a distinction discussed in §1.2 between *language of the Balkans*, the purely achronic geographic definition, and *Balkan language*, designating those languages of the Balkans that show convergences due to language contact. Using that dichotomy as a basis, one can distinguish, within the more general domain of Balkan comparative syntax, between working on the *comparative syntax of the Balkan languages*, i.e., examining the syntax of individual languages that happen to be spoken in the Balkans in comparison with other languages in that region and elsewhere, and engaging in the enterprise of *comparative Balkan syntax*, i.e., the investigation of the syntax of Balkan languages as participants in a sprachbund, thus keeping language contact – which is a diachronic phenomenon – in full view.

This chapter is therefore devoted to *comparative Balkan syntax* in this diachronically and areally defined sense. We approach this task with a view of syntax that is aligned with traditional notions and informed by relevant insights from generative grammar, but we avoid comparisons that are theory-dependent. Instead we focus on key elements that linguists working on syntax in any theoretical framework would have to account for.[4]

4 Working with theory-dependent constructs would be the case, for instance, if we were to focus on sentence pairs involving changes in grammatical relations, something of key interest in a Relational Grammar approach to syntax (as presented, e.g., in the papers in Perlmutter 1983), or to compare which elements could appear in Specifier ("Spec") nodes, something of interest in a Government and Binding and related other approaches (as in Chomsky 1981, 1982, 1986). Achronic work in these frameworks begs the fundamental question of diachronic causation.

7.2 Balkan Syntax from a Traditional Standpoint

As the mention of Sandfeld in the previous section suggests, the more traditionally oriented handbooks have discussed comparative Balkan syntax. In fact, all or most of the numerous syntactic Balkanisms that have been noted in the literature to date are mentioned in one form or another in Sandfeld's work. The list that follows is intended to give an overview of Balkan syntactic phenomena that have received attention in traditional analyses of Balkan syntax. The list refers to both Sandfeld 1930 and the relevant sections in the present work. Selected other works treating these phenomena, often in considerably more detail than Sandfeld, are mentioned in the relevant sections.[5]

a. clitic order [KS 121; §7.4.1.2, also §§5.5.2, 5.5.3; cf. also §7.5]
b. absolute (invariant, universal) relative clause marker [KS107; §7.7.2.2.1]
c. factive versus nonfactive complementizer [KS 175; §7.7.2.1.3.1]
d. interrogative syncretic with relative [KS 107; §7.7.2.2]
e. sequence of tenses [KS 117–118; §7.7.2.1.3.2.1, also §6.2.1.2]
f. double direct objects (both with 'learn/teach' and 'ask' and with "small clauses" having an attribute with the direct object) [KS 201–204; §7.8.4]
g. NEG clause followed by positive (main clause) connected by 'and' [KS 196; §7.7.1.2]
h. parataxis over hypotaxis [KS 196–201; §7.7.1; for juxtaposed parataxis, cf. §4.3.4.1]
i. 'for' + subjunctive to indicate goal, intention, desire [KS 179; §7.7.2.3.3]
j. prepositional convergences [KS 120; §7.9]
k. convergences in nominal complementation [KS123; §7.7.2.1.2]
l. convergences in adverbial clauses [KS 108; §7.7.2.3]
m. object reduplication [KS 192–193; §7.5.1]

As noted, these, as well as numerous other syntactic characteristics, are taken up in the sections that follow.

7.3 Syntax as Approached Here

By way of defining terms and providing relevant historical preliminaries, we give here an overview of issues that pertain to the general areas that we take to delimit syntax as part of the comparative Balkan enterprise as defined in §7.1: ordering, complementation, clitics, and sentence-types. At the end of this section, too, we cover a key methodological point concerning parallelism versus convergence.

5 Chapter 4 references are for material that is more lexical in nature than syntactic proper, e.g., Asenova 2002: 276–290 treats the use of reduplication to indicate quantity, intensity, or distributivity as syntax, whereas we discuss it under the rubric of the lexicon, in §4.3.7.1. Similarly, prosodic factors intersect with some aspects of syntax, so there are some references to Chapter 5.

7.3.1 Ordering

The most fundamental facet of syntax is ordering, as seen in the origin of the term, Greek συντάσσω 'to arrange.' As for the "building blocks" whose order matters, there are independent elements, words, that come to be grouped together into higher-level constructs, phrases, and the phrases combine together to form clauses.[6] Among the phrasal combinations, of particular importance in the Balkans are the noun phrase, consisting of the noun plus adjectives and other modifiers, and that which we refer to here as the *verbal complex*, the combination of the verb with various markers that modify it along certain grammatical dimensions (tense, mood, negation, etc.), since there are pan-Balkan features involving the elements that compose each of these combinations.[7] These units in turn each have their own internal ordering conditions, i.e., their own internal syntax, so that each phrasal type merits discussion on its own. In addition, the ordering of phrases within clauses, what is usually called simply *word order* and often spoken of in terms of the relative ordering of subject, object, and verb, is relevant too. Finally, among the clause-internal and phrase-internal ordering phenomena that require attention are the ordering principles associated with clitics, the so-called little words that play an important role in signaling relationships among pieces within phrases and within and across sentences, as they can show special ordering relative to one another and relative to other elements in the clause or phrase they occur in.

7.3.2 Complementation

Beyond ordering, the expression of the relationship of clauses to one another is a major concern in syntax. Clause-combining includes subordination, especially complementation, i.e., that which certain words need to complete (complement) their lexical requirements and to make for higher-level syntactic units such as complex clauses. Thus complementation as understood here refers largely to elements, especially but not exclusively clauses and clause-like entities, that fill argument positions. In addition, though, there are adjunct subordinate clauses whose use overlaps somewhat with argument clauses and which thus deserve attention in a discussion of complementation in an extended sense. Relevant, too, are any special requirements that clausal combinations are subject to by virtue of the process of combining, such as a special form that a verb must take in a subordinate clause. On the other hand, clause-combining also subsumes coordination, including both overtly expressed linkages and those that are implicit via

6 We are assuming that "words" can be identified easily, though in reality this is hardly the case (see, e.g., Dixon & Aikhenvald 2002 on this issue), and similarly that the assignment of words into major syntactic categories, i.e., noun, verb, adjective, adverb, preposition, is also straightforward, which is likewise not a simple matter at the universal linguistic level.

7 This term, or its relative "verb complex," is used specifically with regard to Albanian by Hubbard 1985, and can be found in the analysis of many other languages showing similar conglomerations of verbs with markers; cf. Gómez & Sainz 1995 on the Basque "verb complex" and Monachesi 2005 on the "verbal complex" in Romance languages.

simple juxtaposition. These two key notions, complementation and coordination, converge in the form of parataxis, which is essentially complementation through coordination. All of these notions – complementation, subordination, coordination, and parataxis – have relevance in comparative Balkan syntax.

7.3.3 Clitics

The clitics mentioned in §7.3.1 figure prominently in syntactic arrangements in the Balkan languages (and many others) and require a separate treatment, parallel to those for ordering and complementation. Clitics are short and prosodically dependent "little" pieces that modify major syntactic categories in some instances and clauses in others. They are word-like in some respects though not fully so, typically showing some affixal properties as well. Nonetheless, they have some independence as far as their semantics and their contribution to phrasal and clausal syntax are concerned, thereby justifying treating them separately (thus here and also in §7.4.1.2 and §7.5).

Although there are reasons to question whether clitics constitute a separate morphological type between words and affixes and thus whether they are to be treated as part of the morphological component of the grammar (in those theories that recognize one) or the syntactic component, we adopt a neutral position here and treat them as key surface elements that must figure in any description of Balkan phrasal and sentence structure. This is consistent with our theoretical stance concerning the importance of surface structure in any account of linguistic convergence and divergence in the Balkans (see §§7.0, 7.1).

There was a significant degree of unity with regard to clitics prior to the particular contact situations that gave rise to the Balkan sprachbund, so that even with similarities in clitic-related phenomena in the Balkans, their emergence entails numerous instances of divergence as well as convergence in the respective dialects of the languages in question. That is, clitics are to be reconstructed for Proto-Indo-European (PIE), based on the occurrence across various of the branches of the family of cognate prosodically weak forms of personal pronouns and of certain other discourse elements (e.g., the coordinative connective 'and,' possibly some negation markers, etc.) showing parallels in form of the sort that generally point to a common origin. As such, clitics clearly constitute part of the shared common inheritance from PIE that fed into the structures of the various Indo-European languages of the Balkans. And, it is not just their form, for they show parallel, and thus reconstructible, positional properties as well, discussed in §7.5.

For instance, Hittite, Ancient Greek, Indo-Iranian (especially Sanskrit and Avestan), Celtic (especially Old Irish), and Slavic (represented here by OCS) all show some clitic forms of personal pronouns, as shown in (7.1) for the first person singular dative pronoun.[8]

8 For most of the languages, similar facts can be marshaled for accusative clitics. However, illustrating with the dative allows Slavic to be brought into the picture since clitic accusative forms are not attested for OCS.

(7.1)

Hitt	*mu*
Grk	μοι
Skt	*me*
Aves	*mōi*
OIr	*m*
OCS	*mi*

These forms share the "nucleus" *-mV-*, found as well in the fuller ("strong") nonclitic forms of the first person singular dative pronoun (e.g., Grk ἐμοί, Skt *mahyam*, Hitt *ammuk*, OCS *mьně*), so that even though there are details of the vocalism of the forms in (7.1) that do not readily match up, a reconstructed "weak" pronoun form **mV-* is well justified. Similar considerations hold for the other personal pronouns.

From this inherited category, clitics developed into the ubiquitous element that they are in the Balkans. The Balkan-particular syntactic developments of these inheritances and of other, more recently emerging, clitics are discussed in §7.4.1.2 and §7.5.

7.3.4 Diverse Sentence Types

In our attempt to be as comprehensive as possible and to emphasize the most important features of comparative Balkan syntax, we also need to consider certain syntactic patterns that characterize the Balkan languages. Of concern here are the combinatory properties of classes of items and of individual lexical items relative to other lexical items, with regard to selection and grouping into phrases, and, importantly, function. Here we take a somewhat constructionist view of syntax, according to which the basic unit of grammar is taken to be the systematic pairing of phrases and clauses with a definable function.

7.3.5 On Parallels and Convergences as our Focus

As discussed in §7.1, our primary focus is on identifying those constructions in the various languages that are plausibly convergent and thus contact-related, and on making the necessary comparisons with the appropriate historical background to evaluate the role of contact. In several instances, we describe convergent aspects of Balkan sentence structure that cannot be definitively connected to language contact or are demonstrably the result of independent inheritance into or independent innovation within each of the languages involved. Nonetheless, even in such cases it is important to describe and examine the convergent elements. Such an exercise helps in developing a typological picture of what a "Balkan language" is like syntactically. More importantly, though, since speakers are not normally aware of the history of particular structures in their languages but can be sensitive to superficial structural similarities between their language and other languages, we would argue that any similarities that are noticeable, as sentence patterns would be, can give speakers a sense of commonality among languages that they know, i.e., of

cross-language parallelism, that they can then exploit in forming, and acting on, their own generalizations and ideologies about relationships among the languages in the region.[9]

7.4 Order

Two aspects of the ordering of elements are relevant here: constituent-internal ordering and the ordering of the major components of a clause. Under the rubric of the latter can be included some details about the order of weak object pronouns (clitics, cf. §7.3.3), which are also discussed in §7.5.[10]

7.4.1 Constituent-Internal Ordering

We treat first here the internal ordering of the major phrasal constituents of a clause, starting with the noun phrase and then moving to the verbal complex.

7.4.1.1 Noun Phrase

In the sections on nominal morphosyntax (§6.1), we discuss the internal structure of the noun phrase and to some extent, therefore, the ordering of elements within it. Although certain aspects of the order are relatively free, the orderings that involve markings for possession, especially by pronominal possessors, and for definiteness are noteworthy, since noun-phrase-internal weak (clitic) elements are involved and the languages show convergent syntax. These topics are treated in §6.1.1.2 and §6.1.2.2, respectively, from a morphosyntactic perspective, so additional elaboration covering the essentials of the syntactic convergence can be given here.

Albanian, Balkan Romance, and Balkan Slavic show marking for definiteness via an element that is often described as enclitic within the noun phrase, although these elements are also sometimes treated as suffixes.[11] The definite marker

9 That is, the situation for the individual speaker in the Balkans may be like speakers who form judgments about the relationship among dialects based on salient similarities rather than on an historical linguist's criterion of significant shared innovations. Here, Gal & Irvine's 2019: 18–19 et passim semiotic concept of *iconicity* is relevant. Speakers associate salient linguistic features with social categories, which, in turn, can both reflect and impact dialectology.

10 Clitics are also discussed §§5.5.2 and 5.5.3, since there are relevant prosodic considerations.

11 Thus, for example, Lunt 1952: 41 treats the Macedonian definite articles as clitic or affixal forms of the demonstrative pronouns, writing: "They are enclitics and can be termed suffixes." In support of this, he cites doubly determined forms of the type *ovie našive* 'these fellows of ours.' However, the occurrence of nonagreeing doubly determined forms of the type *ovaa celata rabota* 'this whole business' (Friedman 2019a and sources cited therein), where *ovaa* is proximal and *celata* is neutral, arguably demonstrate that the definite article is not simply a cliticized demonstrative. For standard Bulgarian and most of its dialects, the fact that there are both neutral and distal demonstratives (*t* and *n* respectively) but only a single definite article (*t*), and that the distal demonstrative can occur with the definite article, e.g., *onaja negovata koleška* 'that colleague of his,' also argues that the two categories are distinct. At the peripheries of the Balkan Slavic definite article area (e.g., Torlak

(article) is positioned after the first nominal element in the phrase that is capable of taking inflection, whether that element is the head noun or a modifying adjective (but not an adverb, compare Mac *mnogu golemite deca* 'the very big children' with *mnogute golemi deca* 'the many big children'); that is, the positioning of the definite article is relative to the noun phrase, not to the head noun alone. Thus one finds the equivalent of 'man-the good' and 'good-the man' but never *'good man-the.' Inasmuch as this marking of definiteness is a contact-induced phenomenon, in particular a possible substratum feature, as discussed in §1.2.1, the enclisis is an ordering feature that was propagated in the Balkans via language contact and is perhaps the most distinctively Balkan aspect of the definite article. Still, it is worth noting that in Early Modern Romanian, it was still possible for the article to occur on the second element (Cornilescu & Nicolae 2011), so at least for Romanian the current situation developed during the Ottoman period.

As for ordering involving possession, there are similarities and differences across the languages, some of which may have been inspired or affected by language contact. Of some interest, regardless of whether it is linked to contact, is the situation with pronominal possessives. In Balkan Slavic, when weak dative pronominal forms are used for possession,[12] these possessors are enclitic within the noun phrase, occurring in second position, as shown in (7.2) where the same facts hold for Bulgarian and Macedonian in (7.2a) but not in (7.2b):[13]

(7.2) a. brat mi / *mi brat (Blg and Mac)
brother me.DAT
'my brother'

b. po-golemijat mi brat (Blg only) / *po-golemijat brat
CMPV-big.DEF me.DAT brother
'my older brother'

The facts from Romanian resemble those for Bulgarian in terms of the use of dative clitics for possession, e.g., *inima-i* 'his heart' (lit., 'heart.the-him.DAT'), but, unlike Balkan Slavic, strong form datives can also be possessive, and, moreover, while they normally precede, they can also follow, e.g., *cumnat nouă* 'our brother-in-law' (lit., 'brother-in-law us. DAT.STR').

In Greek, one can observe a positioning of a pronominal possessive within the noun phrase similar to that in (7.2b), as in (7.3):

(7.3) ο μεγαλύτερός μου γιος
the bigger my son
'my older son' (NOM)

BCMS, Korča in Albania), the transition from clitic demonstrative to article is still visible (see §6.1.2.2.1).

12 These forms are used under different circumstances in Bulgarian and Macedonian (only with kin terms in Macedonian but more broadly instantiated in Bulgarian); see §6.1.1.2.4.2.

13 In Standard Macedonian and the western dialects on which it is based, possessive dative clitics cannot occur in the same constituents with a modifier.

However, in Greek, the nonfronted version, with the possessive following the noun even when there is an adjective in the noun phrase, is perfectly grammatical and indeed represents the unmarked position:

(7.4) o μεγαλύτερος γιος μου
the bigger son my
'my older son' (NOM)

Thus, for Greek, "second position" within the noun phrase is not an issue, and movement of the possessive pronoun is not prosodically driven as it is in Balkan Slavic. Instead, the position in (7.3) represents a stylistic movement, offering a slightly marked nuance of emphasis or affect.

In Romani, as seen in examples (7.5) below, possessive pronouns have distinctive forms in the first two persons (e.g., *moro, mlo, mo,* etc. 'my') but are transparent genitives in the third person (cf. Matras 2002: 99–100). While possessives normally precede the noun possessed (Matras 2002: 166), possessives can follow, especially in dialects in contact with Greek, as in (7.5ab). This type of order is also found in Romani in the Balkans not in contact with Greek, as in (7.5c), based on Arli from Skopje; Igla 1996: 166 has a similar example with *tumaro* 'your.PL.' Both Greek and Balkan Slavic could have played a role in these developments.[14]

(7.5) a. lakoro phuro dad *or* o phuro dad lakoro
her.GEN.M old.M father the old.M father her.GEN.M
'her father'
b. i bari čhej *or* i čhej i bari[15] (SVlax; Igla 1996: 166)
the big girl' the girl the big
c. (o) lakoro dad *or* o dad lakoro
the her father the father her

In Albanian, the normal position for a pronominal possessive is postnominal, as in (7.6a), but a small class of kinship and kinship-like nouns ('father,' 'son,' 'sister,' 'aunt,' etc.) can have prenominal positioning when the possessor is first or second person singular:[16]

(7.6) a. fshat-i im / fshat-i yt
village-the my village-the your
'my village' 'your village'
b. im bir / yt bir
my son your son
'my son' 'your son'

14 Note that in (7.5a) and (7.5c), the definite article *o* can also precede the possessive when it precedes the possessum, but it is not required.

15 Igla compares this with the Grk polydefinite construction (see §6.1.2.2.1.4), e.g., η κοπέλα η μεγάλη ('the girl the big'). See also Tirard 2019, who gives a detailed discussion of these kinds of constructions in various Romani dialects of Albania, and Adamou & Matras 2020, who do so for those of Romania.

16 Note that there is no definiteness marking in the marked phrases (7.6b), and the gender of the pronoun is determined by the kinship term rather than the possessor. In the third person, the particle of concord, agreeing with the gender of the modificand, is used with a definite kinship term, e.g., *i ati* 'his/her/its father.' The dialectal distribution of preposed versus postposed possessive pronouns with kinship terms is complex, and most dialects have only one or the other (Gjinari 2007: Map 212).

Thus even with seemingly convergent elements to the ordering, namely variation between prenominal and postnominal placement of the pronominal possessor, there are differences of detail that make it hard to see the surface convergences as anything but parallel independent developments.

There are two languages in the Balkans, however, that do seem to have been influenced by the possessive facts (ordering and type of pronoun) of a contact language. These are the cases of Meglenoromanian and Aromanian in contact with Macedonian. For Meglenoromanian, as Narumov 2001: 680 observes, a possessive pronoun comes before the possessum, as in (7.7a), diverging from both Romanian and Aromanian, where the possessive follows the head noun, as in (7.7b), and converging with Macedonian, as in (7.7c).

(7.7) a. méu ficiór (Megl)
my son
'my son'

b. ficioru ameu (Aro)
son.DEF my
'my son'

c. mojot sin (Mac)
my.DEF son
'my son'

In the case of Aromanian, the convergence with Macedonian comes in the form of the limitation of dative possessive clitics to kinship terms, e.g., *fratu nju* 'my brother,' as shown by Markovikj 2007: 73; that is, for Aromanian of the Ohrid-Struga region, (postposed) possessive pronouns are normal but dative possessives occur with some kinship terms, e.g., *ficioru ameu* = Mac *moeto dete* 'my son/child' but *fratu nju* = Mac *brat mi* 'my brother.'

As for possession more generally, not just pronominal possession, the Indo-European Balkan languages converge on possessed-preceding-possessor as the generally preferred order, with some of the languages requiring definiteness markers or linking elements, e.g., Mac *kukjata na Ali* 'the.house of Ali; Ali's house,' Grk το σπίτι της Μαρίας 'the house of Mary; Mary's house,' Alb *shtëpia e Agimit* 'the.house of Agim; Agim's house,' Rmn *cărţile professorului* 'the books of the professor; the professor's books,' Aro *dintsãli a lúchilor* 'the teeth of the wolves; the wolves' teeth.'

It can be noted that as a marked order, Greek allows the possessor to precede, e.g., της Μαρίας το σπίτι 'Mary's house' (lit., 'of Mary the house'). Likewise Balkan Slavic can have the order exemplified by Mac *na Ali kukjata* 'Ali's house' (Blg *na Ali kăštata mu*).[17] In Albanian, such orders occur in the Eastern Diaspora dialects in Bulgaria and Ukraine, e.g., *Jani Janakit plaka mëmë* 'Jani Janaki's old

17 In Balkan Slavic, the order is attributed to Turkish influence (that order being canonical in Turkish). Slavic preposed adnominal genitives also occur colloquially in North Slavic (Houle 2013), where they are attributed to the influence of Germanic, Baltic, etc. Thus they occur throughout Slavic, although the causality may still be related to the respective contact situations.

mother' (Friedman 2004b: 59–155). In the case of Greek, the variation may be internal, but there, as for Balkan Slavic, it is possible to make an argument for the influence of Turkish (see (7.8) below). In the case of Eastern Diaspora Albanian dialects, the effect of Bulgarian on adjective order has extended to genitive order (cf. also Hamp 1965; Sokolova 1983; Liosis 2021; but cf. Friedman 1994c).

There are dialects of West Rumelian Turkish (cf. Friedman 2003a: 61–65), e.g., in Kosovo (Ibrahimi 1982), that clearly show contact-induced ordering influence in this domain. As a comparison with standard Turkish indicates, WRT converges with co-territorial Indo-European languages in allowing the possessor to precede the possessed, as in (7.8a), quite unlike standard Turkish, in (7.8b), and like, for instance, local varieties of Slavic, as in (7.8c):

(7.8) a. baba-si Ali-nın (WRT)
father-3SG.POSS Ali-GEN
'(the) father of-Ali'
b. Ali-nin baba-sı (StTrk)
Ali-GEN father-3SG.POSS
'Ali's father' (lit., 'of-Ali his-father')
c. tatko mu na Ali (Mac)
father to.him of Ali
'father of Ali' (lit., 'father-his of-Ali')
d. ati i Aliut (Alb)
father.DEF PC.M.N Ali.GEN.DEF

Thus in (7.8a), *baba* 'father' is initial in the noun phrase, matching the position of Mac *tatko* and Alb *ati*. A similar development is seen in Balkan Romani: all Romani dialects have genitive-head order as in (7.9a), but Romani dialects in the Balkans also have head-genitive constructions, as shown in (7.9b) and (7.9c), as well as a calqued construction using the preposition *tar-* 'from' (7.9d) based on the use of Mac *od/na*, Blg *na*, in (7.9e):

(7.9) a. o/e thagar-es-koro čhavo (Rmi, Cech et al. 2009: 12)
DEF.M.NOM/OBL king-OBL-GEN.M son
'the king's son/the son of the king (NOM)'[18]
b. e padišaj-es-kere o šeja (Rmi, Cech et al. 2009: 33)
DEF.OBL king-OBL-GEN DEF.NOM clothes
'the king's clothes' (ACC)
c. o kher e Drag-es-koro (Rmi, Cech et al. 2009: 200)
DEF.M.NOM house DEF. Drago-OBL-GEN.M
OBL
'Drago's house/The house of Drago' (NOM)
d. i čhib tar-i aždaja (Rmi, Cech et al. 2009: 200)
DEF.F.NOM tongue from-DEF.F dragon
'the dragon's tongue/the tongue of the dragon'

18 There is variation in Arli Romani with regard to the case of the definite article in preposed genitive constructions. The final vowel (and sometimes the penultimate vowel) agree in gender/number/case with the head, which, in (7.9a) is nominative. The definite article, however, can be either nominative to reflect the role of the head, or oblique to reflect the genitive. The variation appears to be free, as in these examples from a single narrator in a single paragraph.

e. jazik-ot od aždaja-ta (Mac)
tongue-DEF.M of dragon-DEF.F
'the dragon's tongue/the tongue of the dragon'

The head-genitive order is the norm in the Balkans and Balkan Romani comes to match that prevailing order in ambient languages.

Within the noun phrase, there is also the question of ordering of adjectives vis-à-vis the head nouns they modify; e.g., in Macedonian, ADJ-NOUN is canonical and NOUN-ADJ is expressive as in *rasipana metla* 'ruined broom' versus *metla rasipana* 'idem' but also with the metaphorical sense of 'scraggly beard.'[19] Similarly in Romani (Arli) *jek bari aždaja* 'a big dragon' is unmarked but *jek aždaja bari* occurs to emphasize the largeness of the dragon (Cech et al. 2009; see also Tirard 2019 and Adamou & Matras 2021 on variations in Romani adjective order in the NP). In Albanian, however, the opposite orders are the canonical and marked ones. In some dialects, however, these relationships are reversed. Thus, for example, in Eastern Diaspora Albanian, ADJ-NOUN is the normal order, whereas in Golobrdo Macedonian NOUN-ADJ is normal (Steinke & Ylli 2008: 83; Friedman 2004b: 59–155). Given that the Eastern Diaspora Albanian dialects of Bulgaria (Mandrica) and Ukraine (and elsewhere) show reversal from the canonical Albanian order in both adjective phrases and genitive phrases, the evidence from what is today Turkish (Eastern) Thrace, which shows only the head-adjective reversal but not head-genitive, suggests that the former preceded the latter (Friedman 2004b: 59–155). Similar phenomena are found in some of the Albanian dialects of Greek Thrace (Liosis 2021).

Similarly, under Macedonian influence (Narumov 2001: 680; Atanasov 2002: 263–265), Meglenoromanian shows a Macedonian-like order with the adjective preceding the head noun; thus *noáua casă* 'new house' in (7.10a) is just like Macedonian *nova kukja* in (7.10b) and the possessive constructions in (7.10c) are calques on the Macedonian in (7.10d):

(7.10) a. dărắm únă noáua casă (Megl)
built.1SG a new house
'I built a new house'
b. sum izgradil nova kukja (Mac)
am build.LF new house
'I have built a new house'
c. lu ąmpirátu íl'a = ąmpirátului feátă (Megl)
to king daughter.DEF = king.DEF.GEN girl
'the king's daughter'[20]

19 Another aspect to adjective order is the relative ordering of different classes of adjectives, e.g., Macedonian *golemata crvena kukja* 'the.big red house' versus *crvenata golema kukja* 'the.red big house.' This phenomenon has semantic repercussions, involving scopal differences (as in the English translations of the Macedonian in this note), so we leave it aside here, as it is not clear how language contact could play a role in a semantically driven phenomenon.

20 Narumov uses the etymologically related 'emperor,' but 'king' renders the basic folkloric meaning of the context.

d. na carot kjerkata (Mac)
to king.DEF daughter.DEF
'the king's daughter'

7.4.1.2 Verbal Complex

At the center of the Balkan sentence is the verbal complex, described in §7.3.1. In all the Balkan languages, overt subjects are pragmatically determined, so that the combination of the verb and its modifiers can constitute in itself a well-formed sentence; that is, the various verbal modifiers in the complex provide all the relevant information needed for full semantic interpretation.[21] Much of the morphosyntax of the verbal complex, involving the particular categories and the markings and the cross-language interactions and developments they show, is discussed in the subsections of §6.2, and the matter of the marking of argument structure is treated in §7.5.2, inasmuch as it is not concerned with order per se. Nonetheless, there are matters pertaining to the ordering of elements internal to the verbal complex that are substantive and that show parallels across the languages, so a treatment of this aspect of the verbal complex is appropriate.

Discussing the order of the pieces that go into making the verbal complex here, in a chapter on syntax, would seem to be implying that these pieces are word-like entities, each associated with its own node in a syntactic tree and each manipulated by rules of syntax.[22] And indeed, traditionally, since these pieces are written as separate words in standard orthographies, with spaces on either side of each of them, they have generally been treated as words and the verbal complex thus as a phrase-level construct. This is the position that Sandfeld 1930: 177 takes when he discusses the parallelism in "*l'ordre des mots dans les propositions introduites par vá, të, etc.*" ('the order of words in propositions introduced by *vá, të*, etc.'). Moreover, there are theoretical frameworks that treat all such verbal operators – markers for tense, mood, negation, etc. – as words that are thus addressed in the syntactic component of the grammar. However, it is also possible to view the pieces of the verbal complex as morphosyntactic markers that are a part of a morphological component and are introduced by morphological rules spelling out bundles of semantico-syntactic features. Thus we want to emphasize that the placement of our discussion in this chapter is not meant as a statement of a theoretical stance as to the division of labor within a grammar between morphology and syntax. Rather, this positioning of our discussion is a reflection of our view that syntax at its most elemental level is about ordering. Since the surface manifestation of the verbal

21 See §7.8.2.1 for more on this "null subject" property.

22 Such a statement of course assumes a particular type of representation of syntactic relationships, one associated with the arboreal structures of transformational generative grammar; we refer to "trees" here as a way of highlighting a particular approach in which the material that spells out a given morphosyntactic category is treated as a word, even though we are fully aware that there are numerous other ways in which syntactic relationships and morphosyntactic material can be schematized.

complex necessarily has phonetic material in a particular ordering, for the purposes of description, we have to pay attention to order, whether the pieces get into their position by rules of morphology or by rules of syntax. Indeed, some of the pieces of the verbal complex in at least some of the languages seem to be best viewed as affixal (morphological) in nature whereas in other languages they appear to be more word-like (syntactic) in nature.[23] As interesting as such questions might be for general linguistic theory, they have less relevance for Balkan linguistics.[24] Nonetheless, the facts of the ordering of elements in the verbal complex contribute to the impression of a convergence of different languages on a common structure and thus they merit serious attention.

7.4.1.2.1 The Verbal Complex: Canonical Orders

The point of departure for discussing the ordering of elements in the verbal complex can be called the canonical order, by which we mean the most frequent and revealing patterns utilizing a maximal expansion in which each modifying category is represented by an overt expression. Example (7.11) comes from material composed in dialectal Greek and dialectal Macedonian as well as Standard Albanian and also Aromanian and illustrates the ordering of material for negation, TAM (future), indirect object, and direct object, all present with a verbal form inflected for the person and number of its (understood) subject;[25] the meaning is the same for all examples in (7.11), 'I will not give it to him':

(7.11) a. δε θε να του το δώσ-ω (dialectal Grk)
NEG FUT SBJV 3SG.GEN 3SG.ACC give-1SG

b. ne kje da mu go dada-m (dialectal or archaic Mac)
NEG FUT SBJV 3SG.DAT 3SG.ACC give-1SG

c. nuk do të i-a jap-Ø (standard Alb)
NEG FUT SBJV 3SG.DAT-3SG.ACC give-1SG

d. nu va sã l-u da-u (Gorna Belica Frasheriote Aro)
NEG FUT SBJV 3SG.DAT-3SG.ACC give-1SG

Another canonical example showing the pieces in a subjunctive modality is (7.12), with standard varieties of Greek and Macedonian as well as Albanian and

23 See, for instance, Joseph 1988, 1989, 1990, 2002ade for argumentation in favor of taking the pieces of the verbal complex in Greek as affixes and not separate words; Sims & Joseph 2019 extends this argumentation to Albanian, but also considers Croatian as well, though with different results.

24 See §1.2 on the difference between Balkan languages and languages of the Balkans and §7.1 for a discussion of this distinction between Balkan syntax and syntax of the Balkans.

25 In Macedonian the negated 'will'-based future has a volitional nuance vis-à-vis the more neutral negated 'have'-based future. Also, *da* here is either dialectal or, in the standard language, has a meaning of attenuation. In Greek, θε να as the future marking is found in regional dialects; the standard language has θα, which actually derives from θε να (see §6.2.4.1.1). See §6.2.4.1 on future forms more generally. Albanian and Greek have distinct nominal versus verbal negators, respectively *jo / nuk ~ mos,* όχι / δεν ~ μη(ν)*.* In Balkan Slavic and Balkan Romance, the negator is *ne* and *nu*, respectively in both types of functions, although the languages also possess many other lexical means of negation.

Aromanian; again, all examples have the same meaning, 'I should not give it to him'/'May I not give it to him':

(7.12)

a. να	μη	του	το	δώσω	(Grk)
DMS	MNEG	3SG.GEN	3SG.ACC	give.1SG	
b. da	ne	mu	go	dadam	(Mac)
DMS	NEG	3SG.DAT	3SG.ACC	give.1SG	
c. të	mos	ia		jap	(Alb)
DMS	MNEG	3SG.DAT+3SG.ACC		give.1SG	
d. sã	nu	l-u		dau	(GBFA)
DMS	NEG	3SG.DAT-3SG.ACC		give.1SG	

In presenting these canonical cases, we give only synthetic verb forms; the analytic paradigms utilizing 'have' or 'be' raise additional issues for some of the languages, some of which are discussed in the sections that follow. Romani, however, does permit its modal negator *ma* after the DMS *te*. While *ma te holjane, mi daj* 'don't be angry mother' (lit., 'MNEG DMS be.angry.2SG.PRS my mother,' Topaanli Arli) is more of a negative imperative, *te ma holjane ...* has an optative reading. The latter thus has a modal reading, but is not, strictly speaking, a negative imperative (see 7.31b'). Cf. here the Macedonian threat of the type *da ne sum te videl* 'don't [you dare] let me see you' (lit., 'DMS NEG you.ACC am.1SG.PRS be.LF'), cited in §4.3.4.1.1. Also, the Romani equivalent of (7.12), *te na dav* 'may I not give,' would have pronominals on one side or the other of this verbal complex. The examples in (7.11) and (7.12) can be generalized into the following templatic schemata (see §6.2.4.2.1 and §§7.4.1.2.2.2.2, 7.7.2.1.3.1):

(7.13)

a. INDICATIVE~MODAL:	NEG - FUT - (DMS)	- IO - DO - VERB	
b. MARKED MODAL:	DMS - NEG	- IO - DO - VERB	

All elements are optional except for the (inflected) verb itself, which can stand alone as a complete utterance, although it is usually accompanied by one or more modifying elements.

The canonical examples in (7.13ab) show the usual order in the verbal complex; there is, however, some cross-linguistic variability in the Balkans as to the content of the verbal complex, to be presented in §7.4.1.2.2. Despite this variability, the order in which these elements occur in the canonical examples is fixed. Sandfeld 1930: 177 was the first to draw attention to this convergence in the order of DMS and subordinate verb and other potential intervening elements: *"c'est une règle sans exceptions que le verbe suit immédiatement la conjunction et ne peut en être séparé que par une negation, un pronom conjoint ou une particule*" ('It is an exceptionless rule that the verb immediately follows the conjunction [= DMS – VAF/BDJ] and cannot be separated from it except by negation, an atonic pronoun, or a particle.') That is, the order of the pieces of a full verbal complex indicated in example (7.13) above is the only order that is possible, with the negation outside (i.e., to the left) of the future marker, and all of the other elements as preposed markers to the left of a finite verb; any other

order yields unacceptable strings (see also §5.5.3).[26] Similar facts are found in all the languages, with such full expansions of the verbal complex; even where some deviations in ordering occur (see §7.4.1.2.2.2), no language allows a wholesale reordering of elements any which way. In the case of Balkan Slavic the lack of intervention is specific to it and not found in BCMS.

7.4.1.2.2 Deviations from the Canonical Patterns

There are deviations from canonical patterns, both in the content of the pieces and in their order. A consideration of these deviations is essential for understanding the full range of ways in which the verbal complex is realized in the Balkans. Some of the elements considered here are also discussed in the relevant subsections of §6.2, but in the sections that follow, they are considered specifically in the context of how they converge and diverge in the constitution of the verbal complex.

7.4.1.2.2.1 Content-based Deviations

There are several parameters of divergence from the canonical that pertain to the content of the verbal complex. While these are not a matter of order per se, they are involved as elements in being part of an order. Their presence, especially when they are innovative elements that need to be integrated in some way into the verbal complex, constitutes a situation in which an order must be imposed by speakers accepting the innovation. Discussion of the history of these pieces and how they came to fit into the verbal complex is mostly deferred until §7.4.1.2.3.

Thus, for example, Albanian is unique among the Balkan standard languages in having a progressive aspect marker, *po*, that occurs with present and imperfect tenses (including 'be' when functioning as AUX).[27] It is preverbal, separable from the verb only by weak object pronouns, and occurs to the right of the negation marker, which puts it in the same slot as the DMS (with which it is incompatible) and

26 See §7.4.1.2.2.2.1 for details on ordering difference. Note that for some verbs in Greek, the ordering of indirect and direct weak object pronouns, when postverbal, can vary; thus, in addition to the more usual order δος του το, with an imperative, one can also find δος το του for 'give it to-him!'(cf. Holton et al. 1997: 304, also mentioned in Householder et al. 1964). For Balkan Slavic, there are different orders with 'be' depending on person, e.g., *jas sum/ti si mu prijatel* 'I am/you are his friend (lit., 'I am/you are him.DAT friend')' but *Toj mi e prijatel* 'he is my friend' (lit., 'he me.DAT is friend'). Macedonian and Bulgarian differ, however, with regard to the co-occurrence of personal clitics. Thus, *toj mi go dade* 'he gave it to me' is acceptable in both languages, but *Toj mi te dade* 'he gave you to me' (e.g., in connection with an arranged marriage), while acceptable in Bulgarian, is not permissible in Macedonian. Moreover, there are ordering differences for these elements between Macedonian and Bulgarian that are prosody-based, as discussed in §5.5.3, with a ban in Bulgarian (unlike Macedonian) on sentence-initial positioning for the prosodically weak object pronouns and some other weak elements being paramount (cf. Alexander 2000b: §1.8). In some Bulgarian dialects of Greek and Turkish Thrace, the weak object pronoun can come before the negator, e.g., *az gi ne znam* vs. *az ne gi znam* 'I don't know them' (lit., 'I them NEG know' versus 'I NEG them know'); cf. Bojadžiev 1991: 274.

27 See §6.2.2.4 on Macedonian dialects in Albania. Makartsev 2020 reports the use of *gje* 'where' + present as a progressive in Macedonian dialects in Albania. Cf. the use of *toko* 'only' as a progressive marker with finite verbs in the Macedonian of Boboshtica, where apparently the phonological similarity to the Albanian (Tosk) gerund marker *duke* is relevant.

the TAM marker (i.e., *po* is not compatible with *do*). Example (7.14) is illustrative:[28]

(7.14) (nuk) po e këndoj / këndoja / (nuk) do ta këndoj/këndoja
NEG PROG it.ACC sing.1SG.PRS sing.1SG.IMPF
'I am (not) singing it' / 'I was (not) singing it'/ 'I will/would (not) sing it'

It seems reasonable to treat the aspectual *po* as part of the Albanian verbal complex, especially since it occurs after (i.e., to the right of) the negation marker and replaces the DMS and FUT.[29] The collocation *po* + DMS (± *mos*), however, is not aspectual but rather modal and can be followed by clitic pronouns before the verb, which can even be in the aorist tense. That construction marks the protasis of a conditional clause, and the type of conditional depends on the tense of the verb (see §6.2.2.3, §6.2.4).

Albanian has an additional expression for progressivity consisting of 'be' in any appropriate tense plus the gerund composed of *duke* (*tue* in Standard Geg; see Gjinari 2007: Map 314 for dialectal variants) and the participle, as in (7.15):[30]

(7.15) jam duke mos e kënduar
am GRDM NEG it. ACC sing.PTCP
'I am not singing it'

The aspectual marker *duke* is outside of (to the left of) negation, unlike its aspectual counterpart *po*, but the ordering of the weak object pronoun(s) relative to negation and to the verb is as with *po*.

Albanian also has a yes/no question marker *a* (see also §7.8.3.1), as in (7.16), that is positioned sentence-initially and preverbally, and can occur in such a way that it appears to be connected to the verb, with just verbal complex material, e.g., a negation marker or a weak object pronoun, intervening between it and the verb, as in (7.16ab), though under suitable conditions of emphasis a full NP can intervene, as in (7.16c):

(7.16) a. A e ke kërkuar në shtëpi?
Q him.ACC AUX.2SG look.for.PTCP at home
'Have you looked for him at home?'

b. A mos e ke kërkuar në shtëpi?
Q NEG him.ACC AUX.2SG look.for.PTCP at home
'Haven't you looked for him at home?'

c. A Rexhepin ke kërkuar në shtëpi?
Q Rexhep.ACC AUX.2SG look.for.PTCP at home
'You have looked at home for Rexhep?'

28 See also §6.2.2.4.

29 Progressive *po* may be tied etymologically to the independent word *po* meaning 'yes; exactly; just.' Diachronically, Joseph 2011b argues, the connection comes through the use of an element meaning 'just' in aspectual contexts denoting momentarily on-going activity, as in *I am just (now) stepping off the plane*. Note that in Indian English, *just now* is used as an aspectual marker for progressivity, as in *I am just now going to school*, meaning 'I am going to school' (whereas *I am going to school* has the generic or simple present meaning 'I go to school').

30 See also §6.2.2.4.

As for *a*, as (7.16c) shows, it has some independence and in fact can even stand alone sentence-finally as a scornful or ironic confirmatory tag question; Newmark et al. 1982: 320 give (7.17a) (with *a* capitalized to suggest emphasis); note also (7.17b) with sentence-initial *a*, where the scornful/ironic meaning is not present:

(7.17) a. Kështu janë burrat, A?
thus are.3PL men.NOM Q
'Men are like this, are they?!'
b. A kështu janë burrat? / A janë kështu burrat?
Q thus are.3PL men.NOM Q are.3PL thus men.NOM
'Are men like this?'

Here the semantic and functional connection between sentence-initial question *a* and sentence-final questioning *a* seems synchronically clear, so such independence for *a* should most likely be taken to mean that it ought not to be considered a part of the verbal complex, even if it is closely connected to the verb when it occurs sentence-initially.

Balkan Slavic also has a yes-no question marker, *li* (see also §7.8.3.1), that has somewhat complex ordering properties, at least superficially, interacting with more general prosodic patterns in each language and with the expression of focus. It can be argued to be tied to the verb, since it is generally adjacent to the verb; compare the Macedonian examples (7.18a–d), where *li* is either preverbal or postverbal, as in (7.18a–c), but never separated from the verb, as in (7.18d):

(7.18) a. Ivan li dojde?
Ivan Q come.3SG
'Is IVAN coming?'
b. Ivan dojde li?
c. Dojde li Ivan?
d. *Dojde Ivan li?

Nonetheless, *li* can put question focus on a word contained within a noun phrase, and thus be quite separate from the verb, as in (7.19) from Bulgarian:

(7.19) Novata li kola prodade (ili starata)?
new.DEF Q car sold.3SG or old.DEF
'Did she sell the new car (or the old one)?'

Thus Balkan Slavic *li* would seem to be best treated not as a part of the verbal complex per se but rather as a separate element that can nonetheless interact with the pieces of the verbal complex under appropriate conditions. In Englund's 1977: 137–143 corpus of yes/no questions, 60.4 percent of the Bulgarian questions used *li* as opposed to 30 percent in Macedonian, and 44.1 percent of her Macedonian questions had no lexical interrogative marker but only 19.9 percent did in Bulgarian.

A second parameter of divergence from the pattern as far as the future tense is concerned is that, as discussed in §6.2.4.1, there is considerable variability in the

realization of the future along two dimensions. First, the subjunctive marker in the future tense is optional for many of the languages, being absent regularly in Balkan Slavic (Mac *kje dojdam* 'I will come,' Blg *šte dojda*, Trlk *ču/če dođem*), colloquially in Albanian (*do shkruaj* 'I will write'), and dialectally, or perhaps more accurately, historically for some dialects, in Greek (θε γράψω 'I will write'). In terms of the verbal complex template, this means one slot fewer (or one unfilled slot) but does not affect ordering at all. Second, a different verbal basis for the future marker is found in some of the languages: Balkan Romance has the possibility of using a 'have'-based form as a variant on the volitionally based marker, and in Balkan Slavic, and dialects of Aromanian in contact with Balkan Slavic as well, 'have' is usual as the future marker in negated forms, while in much of Geg Albanian, a 'have' auxiliary is the norm in the future, followed by an infinitive as the main verb rather than a (modally marked) finite verb.[31] The Geg formation with 'have' and an infinitive means that another element – the infinitival marker *me* – must be reckoned into the verbal complex template, but it can be taken to occupy the same slot as the DMS(SBJV) in (7.13) in §7.4.1.2.1. Since weak object pronouns occur inside of, i.e., to the right of, the infinitive marker (cf. Buchholz & Fiedler 1987: 153), as in (7.20a),[32] the earlier template works, perhaps with a relabeling of the subjunctive/infinitive slot as "MOD" (for "modal"), as in (7.20b), and thus is exactly parallel to the "ideal" full expansions in §7.4.1.2.1:[33]

(7.20) a. nuk kam me e bâ
NEG have.1SG INFM it.ACC PTCP
'I will not do it'
b. NEG – FUT(HAVE) – MOD(INFM) – IO – DO- VERB

Also, as shown in §6.2.4.1.5, some Balkan dialects of Romani use a negative possessive construction for the negative future, with the negated form of 'be' and an accusative pronoun for the possessor:

(7.21) nae man te hav
NEG.be 1SG.ACC SBJV eat.1SG
'I will not eat' (lit., '(There) is-not (to-)me that I-eat')

but this does not affect the template, as it is still NEG – FUT(HAVE) – DMS(SBJV) – VERB.

As discussed in §7.6 and especially §7.6.2, the negation slot in Albanian, Greek, and Romani must be divided between indicative and nonindicative negation (in Romani, mostly nonimperative and imperative). As a comparison of (7.11) and (7.12) in §7.4.1.2.1 shows (and cf. the summation in (7.13)), there are different orderings associated with this different content in these three languages, whereas Balkan Romance and Balkan Slavic in general require just one slot for negation; the

31 See §6.2.4.1.4 on the distribution of Albanian future formations.

32 A Google search does turn up instances of the pronoun occurring outside of, i.e., to the left of, *me*, thus *e kam me bâ* 'I will do it' (lit., 'it I.have to do'); this variation clearly needs further study.

33 Cf. the view espoused by some ancient grammarians, e.g., Apollonios Dyskolos, by which the infinitive was seen as a kind of mood.

Aromanian and Macedonian dialects that have borrowed modal *mi* from Greek for use in prohibitions (see §4.3.3.3 and §7.6.2) replaced, respectively, native *nu* and *nemoj* (or the like) with *mi*, so presumably the prohibitive template was not altered except for having a different surface realization.

A final parameter for content-based divergences from the ideal form has to do with instances where adjacent elements fuse, or historically have fused, into single portmanteau realizations. For instance, the θε να of Greek, consisting of the future marker and the subjunctive marker, has, through a few sound changes and analogies, ended up as the widely occurring θα, and Macedonian and Bulgarian dialects show *kja, ža*, *šča*, *šta,* etc. from earlier **tje da* (cf. Mazon 1936), *ža* presumably via *šte da* > *še da* > *š da* > *žda* > *ža*, as suggested by Motoki Nomachi (p.c., March 2022). In such cases, the fact of fusion is not really synchronically recoverable for speakers, except to the extent they might have cross-dialectal experience and awareness. Moreover, there are portmanteau fusions involving combinations with negation that might appear to be synchronically more transparent, especially *nema* in Macedonian and western Bulgarian from 'not' + 'has' in the negative future (cf. the independent existence of *ne* 'not' and *ima* 'has'; Standard Bulgarian and the eastern dialects on which it is based have *njama,* the expected reflex of the jat [< **ē*], which is synchronically less transparent; but given the *e* ~ *ja* alternation in Standard Bulgarian and the dialects on which it is based, this, too, can be taken as potentially transparent), and there are northern Greek dialects that have δα as a fused form of 'not will,' from δε 'not' + (θ)α 'FUT.'[34]

By far the preponderance of fusions in the verbal complex, however, comes from combinations of weak object pronouns, in many instances involving a process that is phonologically irregular from a synchronic standpoint. Albanian, for instance, has a special set of outcomes for combinations of indirect object with direct object, illustrated in (7.22), given in Albanian orthography:

(7.22)

IO	DO	Outcome	Meaning ('DO to IO')
më	e	ma	'him/her/it to-me'
të	e	ta	'him/her/it to-you.SG'
i	e	ia	'him/her/it to-him/her it'
i	i	ia	'them to.him/her/it'
u	e	ua	'him/her/it to.them'
u	i	ua	'them to.them'
ju	e	jua	'him/her/it to you.PL'
ju	i	jua	'them to-you.PL'

and similar fusions are found with the DMS *të* plus object pronouns, as in (7.23):

(7.23) të + e (DO) => ta 'that him/her/it'

In Romanian, the combination of 3SG indirect object *îi* 'to him/her' with 3SG direct object *îl* gives a fused *il* 'it/him to-him/her,' and in Greek, indirect object σου 'to

34 And even Greek, to the extent that both δε and θα are affixal (cf. Joseph 2002e), could be said to have a type of fusion here, as these pieces are not separable when they co-occur.

you.SG' combines with direct objects τον/την/το/τους/τις/τα 'him/her/it/them(M/F/N)' to give στον/στην/στο/στους/στις/στα (cf. Joseph 2002e) meaning 'him/her/it/them to-you.' In none of these cases is the combination predictable from regular rules of phonology for the language even though these combinations are obligatorily converted into the fused forms.[35]

With this understanding about the fusion of elements, especially in the portmanteau realizations of pronoun combinations (cf. (7.22) above), Albanian and Romanian can be added to dialectal Greek and dialectal Macedonian (see §7.4.1.2.1) in showing (near-)maximal expansions of the verbal complex, and so too with Balkan Slavic and its negative fusion (see also §7.6.4):

(7.24) a. S' do të ia jep (Alb)
NEG FUT DMS 3SG.DAT.3SG.ACC give.1SG
'I will not give it to him'

b. Nu o să il dau (Rmn)
NEG FUT DMS 3SG.DAT.3SG.ACC give.1SG
'I will not give it to him'

c. Nema da mu go dadam (Mac)
NEG.FUT SBJV 3SG.DAT 3SG.ACC give.1SG
'I will not give it to him'

The basic left-to-right order remains the same even with these fused elements but the realizations of categories in the template are different.

7.4.1.2.2.2 Order-based Deviations

Regarding the specifics of how the elements are ordered, while there is general agreement across the languages on the ordering as seen in the examples in §7.4.1.2.1, some deviations from that ideal do occur, focusing on each of the various pieces of the verbal complex, but especially pertaining to the weak object pronouns.

7.4.1.2.2.2.1 Weak Object Pronoun Ordering First, with regard to the ordering of the weak object pronouns relative to one another and relative to other elements in the verbal complex, it must be noted that, just as modality mattered for the order of negative and DMS in Greek and Macedonian (cf. §7.4.1.2.1), the other pan-Balkan modalities, the indicative and the imperative, must be taken into account, as aspects of weak object pronoun ordering are dependent to some extent on mood.[36] For one

35 The pronominal combination of IO *më* + DO *e* is the norm in Albanian and thus regular in that sense, and is also mirrored by combinations with IO *të* or DMS *të*, and thus regular in another sense. But there are combinations of *ë* with *e* that do not yield *a*, meaning that the pronominal outcomes involve special phonology in the verbal complex; in particular, the adjectival degree marker *më* when occurring before a PC *e*, e.g., as a feminine singular nominative, as in *më e mirë* 'better' (lit., 'more PC good') cannot coalesce, so that **ma mirë* is completely unacceptable in Standard Albanian. (Note that Geg, *mâ* =Tosk *më* 'more' and can be phonetically denasalized to *ma* and *mirë* without the PC is an adverb, so the string in question does occur as such, but with a different composition and function.)

36 There is a bit more to the matter than this, because with the few nonfinite forms that are present in most Balkan languages, such as the form variously called the gerund or gerundive or present active participle or verbal adverb (see §7.7.2.3.1), the weak object pronouns are positioned after the verb,

thing, since the DMS is optional in the sense that the verb alone can constitute a sentence,[37] weak object pronouns can be positioned relative to the verb in the absence of a mood marker or future tense marker. In such instances where the verb is indicative, i.e., present or past, clitics attach to the verb but are positioned differently in some of the languages from instances where there is a DMS or future marker; moreover, if the verb is in the imperative, that forces a different ordering of the weak object pronouns.

In Albanian, Greek, Macedonian, and Balkan Romance, the weak object pronouns precede (i.e., are positioned to the left of) finite (bare) indicative verbs, even if this means that the pronouns are initial in their clauses. The one exception is Macedonian 1st /2nd person 'be,' e.g., *sum mu prijatel* versus *mi e prijatel* 'I am his friend' versus 'he is my friend' (lit., 'am him.DAT friend' versus 'me.DAT is friend'); cf. (7.148). We can also note that the Banat Bulgarian dialect of Romania has clitic ordering as in Romanian, while Banat Bulgarian in Serbia has either Bulgarian or even Serbian type ordering (Nomachi & Browne 2019). The main exception is the Romanian 3SG.F.ACC.WK *o*, which follows the main verb in the perfect and conditional (Vasilescu 2013: 367; cf. also Nicolae 2019 on clitic order in Early Modern Romanian (sixteenth to eighteenth centuries, see Chapter 1, footnotes 42, 60)). With imperatives, the positioning is different, with weak object pronouns following (i.e., positioned to the right of) imperatives, though with some complicating exceptions in Albanian. The Albanian indicative facts, and the Greek, Macedonian, and Balkan Romance indicative and imperative facts are exemplified in (7.25); note that the subject pronouns are optional so that the weak object pronouns can be sentence-initial:[38]

(7.25) a. (ne) e shqiptojmë / *(Ne) shqiptojmë e (Alb)
we.NOM it.ACC say.clearly.IND.1PL
'We say it clearly'

b. (Unë) i godita / *(Unë) godita i (Alb)
I.NOM them.ACC struck.1SG.AOR.IND
'I struck them'

c. (Εσείς) μου το δίνετε / *(εσείς) δίνετε μου το (Grk)
you.PL me.GEN it.ACC give.IND.PL
'You (all) are giving it to me'

d. (Εσείς) δώστε μού το / *(Εσείς) μου το δώστε (Grk)
you.PL give.IMPV.2PL me.GEN it.ACC
'Give it to me (y'all)!'

e. (Jas) go pročitav / *(Jas) pročitav go (Mac)
I.NOM it.ACC read.1SG.AOR.IND
'I read it'

e.g., Grk δίνοντάς μου το '(while) giving me it,' Rmn *spunându-i* '(while) telling-him,' Mac/Blg *davajkji/davajki mu go* 'while giving him it.' Albanian does not have such ordering.

37 There are some restrictions in some of the languages on the aspect of a stand-alone verb; see §7.7.2.1.3.2.2 for discussion.

38 Regarding Bulgarian, see further below in this section (and also footnote 26). Dialectally, Greek shows considerable variation; see Ralli 2006: §9 for a succinct overview of the dialect facts.

f. (Ti) daj mi go / *(Ti) mi go daj[39] (Mac)
you.SG give.IMPV.2SG me.DAT it.ACC
'(You) give it to me!'

g. (El) mi-a spus / *(El) spus mi-a (Rmn)
he.NOM me.DAT-it.ACC told.3SG.IND
'He told me'

h. (Tu) ajută-mă / *(Tu) mă ajută (Rmn)
you.NOM help.IMPV.2SG-me.ACC
'Help me!'

In Albanian, the positioning with imperatives involves a mix of placement to the left of, to the right of, and even in the middle of, the verb. Newmark et al. 1982: 26 describe it thus:

> With imperative verbs, only pronominal proclitics and combinations of the first, third, and reflexive persons are used. In the positive imperative these proclitics may appear before or after the verb (except for the clitic ***e***, which rarely precedes the verb stem in positive imperatives). In the positive plural imperative they usually precede the stem; but they may also appear as enclitics, coming between the stem and the ending, and that whole sequence is written as a single word.

Among the examples they give are the following (with dashes introduced here to show morpheme boundaries, contrary to the usual orthography), with (7.26a), (7.26b), and (7.26d) showing "proclitic" positioning, (7.26c) showing "enclitic" positioning, and (7.26e) showing what can be called endoclitic positioning:[40]

39 Although preposing clitics to imperatives is counter to the Macedonian norm, it does occur in dialects and folklore, as in the song *Ne mi go prodavaj, Koljo, čiflikot* 'Don't sell the farm on my account, K.' (lit., 'NEG me.DAT it.ACC sell.IMPV, K., farm.DEF') or *prvo me mene ubodi, džanam* 'stab me first, my dear' (lit., 'first.ADV *me.*ACC.WK me.ACC.STR stab.IMPV my.soul'), which latter looks like a Wackernagelian archaism. We can also note here the preposing versus postposing of clitics to infinitives in constructions of the type PREP+INF in Judezmo, in that both *por verlos* and *por los ver* ('for to.see.them' / 'for them to.see') are possibilities in what is now the former Yugoslavia, Albania, and western Aegean Macedonia, including Thessaloniki (but also Istanbul and Izmir), while the first order is the only one for Judezmo elsewhere in the Balkans (Quintana Rodríguez 2006: 400). It would appear, however, that this variation has its roots in the Iberian peninsula, where both orders are or were possible in various regions, and sometimes still are today. Still it is worth a comment that the variation in the Balkans is limited to precisely the regions where it occurs, with Istanbul and Izmir as major outlying places.

40 By various tests proposed by Zwicky & Pullum 1983 for distinguishing clitics from affixes, the weak object pronouns appear to be more affix-like than clitic-like (see Joseph 2013a and Sims & Joseph 2019), so that prefixal, suffixal, and infixal might be more apt. "Endoclitic" refers to the positioning of a (true) clitic within a word, a conceivable placement that Zwicky in some work suggested was impossible, though cf. Harris 2000, 2002 on this ordering actually being realized in Udi, and Maisak 2021 on the phenomenon in Andi. See also footnote 63. Here we must also note Liosis 2021, who cites Haxhihasani 1971: 177–178, regarding endoclisis of the personal pronoun: *pas-me-ke bə́nə* (= StAlb *ma paske bërë*) '(really) you have done it to/for me,' where *paske* 'you have' is analyzed as an admirative auxiliary with clitic *ma* 'for.me.it,' which would certainly be the case in Standard Albanian. Given the relative recency of the fusion of PTCP+AUX to mark admirativity, the apparent endoclisis is arguably not on the same level at that found in Daghestanian. See also footnote 63 in §7.5.3.

(7.26) a. Na tho-ni!
us tell-2PL.IMPV
'(Y'all) tell us!'
b. Më thua-j!
me tell-2SG.IMPV
'Tell me!'
c. Thua-j-më!
tell-2SG.IMPV-me
'Tell me!'
d. Më shkrua-ni!
me write-2PL.IMPV
'(Y'all) write to me!'
e. Shkrua-më-ni!
write-me-2PL.IMPV
'(Y'all) write to me!'

In Bulgarian, by contrast to the other languages in their current state, the positioning in general follows a somewhat different motivation not related to the template, but with the result of a different outcome in the imperative from that in a clause with a DMS or a "bare" indicative, i.e., a verb without a DMS or FUT marker. Bulgarian weak object pronouns follow a (non-Balkan) Romance-style Tobler-Mussafia Law sort of distribution of being banned from occurring in initial position, a distribution that could alternatively be seen as a Wackernagel's Law (see §7.5) type of placement requiring weak object pronouns to be in second position within their domain, except that, as Alexander (2000b, 2000c) demonstrates, Bulgarian clitics are positioned relative to the verb and are not in clausal second position, thus contrasting with the situation in BCMS; see (7.27a) versus (7.27b) and note the contrast with Macedonian in (7.27c):

(7.27) a. Az često mu go davam (Blg)
I often him.DAT it.ACC give.1SG
'I often give it to him'
b. Ja mu go često dajem (BCMS)
I him.DAT it.ACC often give.1SG
'I often give it to him'
c. Mu go davam često (Mac)
him.DAT it.ACC give.1SG often
'I often give it to him'

With imperatives, where the verb is typically first in its clause, or with bare indicatives where the verb can be first, the weak object pronouns in Bulgarian follow a sentence-initial verb, as in (7.28a–b), but if there is a word preceding, such as a subject pronoun for emphasis, the weak object pronouns follow that and thus precede the verb, as in (7.28c–d). The relevant factor here is intonation, i.e., the prosodic unit, so that if a nominative pronoun occurs in initial position with its own intonation, as its own prosodic unit, essentially serving as a vocative, as in (7.28e), so that the imperative is part of a new clause altogether, the weak object pronouns follow the verb. Thus, in a more elaborated verbal complex, with the DMS or the

future marker counting as a first prosodic element, the weak object pronouns follow that but precede the verb, as in (7.28f):

(7.28) a. Vzimam go / *Go vzimam
take.1SG it.ACC
'I take it'

b. Vzemi go! / *Go vzemi!
take.2SG.IMPV it.ACC
'Take it!'

c. Ti go vzemi! / *Ti vzemi go! / *Go ti vzemi!
you.NOM it.ACC take.2SG.IMPV
'You take it!'

d. Az go vzimam / *Az vzimam go / *Go az vzimam
I it.ACC take.1SG
'I take it'

e. Ti! Vzemi go! / *Ti! Go vzemi!
you.NOM take.2SG.IMPV it.ACC
'You! Take it!'

f. Da go vzema-š / *Da vzemaš go / *Go da vzemaš
DMS it.ACC take-2SG
'You should take it'

Romani weak object pronouns typically follow the verb, so that their placement is categorically different and at odds with the template seen in the other languages. Examples (7.29a) and (7.29b) are illustrative:

(7.29) a. Ola me ka lav
her.ACC I FUT take.1SG

b. Me ka lav la
I FUT take.1sg her.ACC
'I will take her [for my wife]'

And, as for West Rumelian Turkish, the language is basically suffixing, like the rest of Turkic, and clitics always follow the stressed element; still, elements that can be either fused or separated are more likely to be separated and less likely to show vowel harmony in WRT than in ERT or western Anatolian Turkish.

Finally, with regard to the order of the weak object pronouns relative to one another, the order of dative to the left of accusative is observed across all the languages, even in imperatives, where the position of the weak object pronouns relative to the verb does not affect their order relative to one another. However, in Greek imperatives, the direct object accusative pronoun can occur either to the left of the indirect object or to its right:

(7.30) a. Δώσε μού το / δώς το μου
give.2SG.IMPV me.DAT it.ACC
'Give it to him!

b. Πλήρωσέ του το / πλήρωσέ το του
pay.2SG.IMPV him.DAT it.ACC
'Pay it to him!'

and this ordering is possible for some speakers with the other nonfinite form, the present active participle (gerund(ive)), e.g., δίνοντάς μου το / δίνοντάς το μου '(while)-giving it to-me' (Holton et al. 1997: 304).

7.4.1.2.2.2.2 Order of Other Elements Verbal complex elements other than the weak object pronouns show ordering variations. In Albanian, for instance, a difference in order for the negator and the DMS mood marker can be observed, with both *të mos* and *mos të* occurring; there seem, however, to be differences in nuance of directness or in illocutionary act, and this same type of distinction also occurs in Romani:

(7.31) a. Mos të qesim pengesa (Alb)
MNEG DMS make.public.1PL obstacles
'We should not make the obstacles publicly known'

a'. Ma te holjane, mi daj (Rmi)
MNEG DMS get.angry.2SG.PRS my mother
'Don't get angry, mother!'

a" Hajri ma te dikhe daje (Rmi)
good MNEG DMS see.2SG.PRS mother.VOC
'May you not see [anything] good, O mother [because of what you did to me]'
(title of a song by Esma Redžepova)

b. Të mos qesim pengesa (Alb)
DMS MNEG make.public.1PL obstacles
'Let's not make the obstacles publicly known'

b'. Te ma holjane, mi daj (Rmi)
DMS MNEG get.angry.2SG.PRS my mother
'May you not get angry, mother!'

In Macedonian, the usual order is negation inside of (to the right of) the mood marker, i.e., *da ne*, but if the whole *da* clause is being negated, as in *dojdov Cezar da go pogrebam, a ne da go falam* 'I came to bury Caesar, not to praise him' (lit., 'I.came Caesar DMS him.ACC bury.1sg.PRS and/but NEG DMS him.ACC praise.1SG.PRS'), the order *ne da* is normal. And, dialectally within Greek, in particular in the dialects of Crete and of Asia Minor, the sequence θα μη, with the modal negation marker after the future marker, occurs (Markopoulos 2006: 247), and να θα is reported in Markopoulos 2006: 240–241 for Crete, and in Ralli 2006: 137, apparently without any particular semantic effect. Moreover, Greek attests a question marker μήνα, used when one expects a negative answer, that is listed in LKN online: s.v. as a "popular" form ("λαϊκότροπη"); it appears to be from the question use of μη (see §§7.6.3.2 and 7.8.3.1) with DMS να, and if so would show a noncanonical order of elements, though a derivation from the (prevocalic and pre-stop) variant μην, with an innovative final -α by analogy to the semantically similar τάχα 'supposedly,' is possible too.

Finally, a number of languages have seen a shift in order of an auxiliary verb – i.e., the future marker or 'be' or 'have' in the perfect tense – with the main verb (participle in the case of the perfect). Early Modern Romanian, for instance,

allowed *mîncat-am* for 'I have eaten' (lit., 'eaten-have.1SG'), whereas modern Romanian has *am mîncat*;[41] similarly, in Albanian, while the order 'have' before (to the left of) the participle is usual, e.g., *e kam harruar* 'I have forgotten it' (lit., 'it-I.have-forgotten'), the inverse order, *harruar e kam*, is possible but stylistically marked.[42] And, in the future tense, Medieval Greek, and modern dialects into the late nineteenth and possibly the early twentieth centuries (cf. Thumb 1912: 161), allowed inversion of the future auxiliary, a form of θέλω 'want,' so that instead of its more usual position preceding the main verb (whether infinitival or a DMS-headed form), it would follow, e.g., γράψει θέλω or να γράψω θέλω for 'I will write' (lit., 'write will.1SG' / 'DMS write.1SG want.1SG'). Similarly, one could find into the twentieth century the so-called "future indefinite" (*bădešte neopredeleno*) described in grammars of Bulgarian, e.g., S. Mladenov 1929, Mirčev 1963, with the (conjugated) future marker following the (infinitival) main verb, e.g., *vide šta* 'I will see' (lit., 'see will.1SG'). According to Stojanov 1983: 386, such forms are now used only to render dialectal or archaic effects. There may have been a Wackernagelian (or Tobler-Mussafian) motivation here, especially in Bulgarian but perhaps in the Early Modern Romanian *mîncat-am* type as well; an auxiliary was typically unaccented and thus prosodically weak, so that it could not appear in initial position and would naturally have followed the main verb it was associated with. That this is the case is suggested by the fact that in the Bulgarian past tense, consisting of an *l*-resultative participle and an inflected form of the verb 'be,' the 'be' verb is prosodically weak and cannot stand in initial position; the participle can be initial so that one finds, e.g., *Pišel săm* 'I wrote' but not **Săm pišel*. However, Macedonian (or, more specifically, the western dialects on which the standard is based) has altered the Tobler-Mussafia/Wackernagelian restriction on weak object pronouns, so that they are in line with the non-Slavic Balkan languages: *Go videle* is the required order for 'They saw him' (lit., 'him saw.PL'), a key difference between Macedonian and Bulgarian; see also §5.5.2, §5.5.3, §7.3.3, and §7.4.1.2.2.2.1 for more on the ordering of weak object pronouns and other prosodically deficient elements.

7.4.1.2.3 The Verbal Complex as a Balkanism

Even with all the foregoing detail about differences in the realization of the verbal complex in the different languages, there is still a considerable amount of convergence that is evident in the surface forms that the languages show for this construct. This convergence, as noted above in §7.4.1.2, certainly contributes to an appearance of structural sameness across the languages, but it is fair to ask if there is

41 Pană Dindelegan 2013: 224, 392 gives some examples in modern usage of the inverted order of participle before auxiliary found more commonly in Early Modern Romanian, but considers such instances to be "archaic" or "marginal." This type of Balkan Romance inverted perfect was assigned a distinct meaning in Meglenoromanian; see §6.2.5.4.

42 And, an inverted perfect is the source of the Albanian admirative mood, with a short participial form of the sort found in Geg and Lab preceding a form of 'have'; see §6.2.5.5.

something more than that, and in particular if the structure of the verbal complex is a Balkan feature that has resulted from language contact, i.e., if it is a Balkanism.

For the most part, the particular pieces in the respective verbal complexes of the various languages represent native material within each tradition, e.g., Albanian negation *s'* from earlier *(*ne*) *k*^{w}*id* '(not) at-all,' Macedonian future *kje* from earlier *xъtj-* 'want,' Greek να from earlier ἵνα 'so-that,' the weak object pronouns from earlier pronominal forms, and so on. And in some cases, the particular sequencing of forms in the complex represents earlier ordering, as with the invariant positioning of the DMS mood marker before the verb it is modifying or even the positioning of auxiliary verbs before their associated main verb, given the preponderance of such instances in earlier stages of Greek, Slavic, and Romance, whether in the future tense or the perfect tense. It is the case that this new material came to be positioned in ways parallel to models elsewhere in the language and thus often generally parallel to what is seen in other Balkan languages.

Albanian progressive *po*,[43] for instance, may derive from an earlier **pēst*, a composite of an asseverative **pe* and a form of 'be' (**est*), as suggested by Hamp apud Joseph 2011b, by a semantic change in which it took on a momentarily progressive sense (see §7.4.1.2.2.1 and footnote 29). Once it had that grammatical function modifying the verb aspectually, *po* came to fit in with the overall structure of the verbal complex, being exterior like the other marker *duke* used in the alternative progressive formation and thus with weak object pronouns inside of (to the right of) it. If based on an existing model in terms of order, the positioning of *po* would not indicate anything contact-related regarding order.

Moreover, the position of the negation marker in indicative forms of the verbal complex may be a reflection of a cross-linguistic universal tendency, at least for head-initial languages, by which negation is preferred on or near the left periphery of a clause.[44] The deviation from such left-peripheral positioning that is seen in subjunctives in Albanian (*të mos*), Romani (*te na*, but *ma te*; however see the discussion after (7.12)) and Balkan Slavic (*da ne*), appears to be a possible significant shared feature between these languages inviting an inference of contact involvement in some way. There is, however, little conclusive to go on here, as it is just a matter of a similar choice between two orders and the Albanian positioning of *mos* vis-à-vis *të* has an internal parallel in the order *duke mos* + participle.[45]

43 See §6.2.2.4 for more on *po*.

44 This generalization, originally proposed by Otto Jespersen (see Jespersen 1917), is what Horn 1989: chapter 7 (see also Horn 2001: xxi) refers to as the "NEG FIRST" principle; note such phenomena as "NEG-Raising" (*I don't think he will come = I think he won't come*) or the *can't seem to* construction (*I can't seem to win = It seems that I can't win*), where the leftward positioning of the negation may be a reflex of this putative universal tendency. However, all the examples are Indo-European, and Turkish is an obvious counterexample. The generalization may say more about an Indo-European bias in some areas of linguistics than about actual universals.

45 Greek θα μη, discussed in §7.4.1.2.2.2, is probably not relevant here in part because it is dialectal in regions not contiguous with Albanian or Balkan Slavic (Crete and Asia Minor). Also, since θα derives from θε να and μη is the expected negation with να (cf. (7.12a)), θα μη makes sense from an etymological standpoint (making the prevalence of the indicative negator before θα, i.e., δε θα, of

With regard to the positioning of the weak object pronouns, it must be noted that adjustments in the ordering of weak object pronouns have been an on-going process for centuries in all of the languages, with a strong internal (and universal, or typological) dimension to it. The differences seen, for instance, in Balkan Slavic weak object pronoun positioning is indicative of fairly recent developments leading away from an earlier BCMS-type configuration to the fixing of the clitics to the verb phrase in Bulgarian (Alexander 2000b, 2000c; cf. Pancheva 2005), to the completely non-Wackernagelian/Tobler-Mussafian situation found in Macedonian. Similar observations have been made about the emergence of the Greek situation, as discussed especially by Janse 1993, 1994, 1998; Condoravdi & Kiparsky 2001, 2004; and Pappas 2004ab. Such historical observations pointing to language-internal developments as playing a key role in the emergence of the current state of affairs with weak object pronoun ordering in the Balkan languages do not mean that language contact could not be responsible for some of the parallels. The fact that precisely western Macedonian, where contact with Albanian, Aromanian, and Greek is strongest and most complex, has sentence-initial clitics found nowhere else in Slavic but found in those other languages, argues for contact-induced change. There are also some specific details that match between Greek and Macedonian for which a contact explanation is suggestive (see §5.5), and such is the case as well with a matching of a detail of pronoun ordering in Albanian, Greek, and Judezmo, as discussed in §7.5.4.

Moreover, the positioning of the DMS after the future marker (Grk θε να, Alb *do të*, Slv *kje*/*šte*/*če da*, Rmn *o să*, etc., Aro *va s'*, Megl *[ă]s*, Rmi *ka te*) can be seen as being a feature of complementation in each language, inasmuch as complements typically follow their governing element, and the order would reflect that fact; however, occasional instances of να θα occurring in Greek might suggest that speakers were in a position to make choices about the order.[46] So it could be that the overwhelming occurrence of θε να / *kje da* / *do të* instead of να θε / *da kje* / *të do* in fact represents a contact-induced convergence, and certainly any "stray" new elements brought into the fold of the verbal complex could have been modeled on orderings observed (actually, heard) by speakers in other languages. Still, any innovative elements could just as easily be modeled on patterns already emerging in each language.

Finally, the fusion of elements seen in developments such as Grk θα from θε να or BSl dialectal *kja, ža*, *šča*, *šta*, etc. from **xtje da* can also be explained by reference to language-internal forces involving vowel and consonant assimilations and reductions, all representing fairly natural paths of development.

some interest, even though δε would be expected with θέλει, the source of θε; the left-peripheral positioning of negation (see footnote 44) may be at least in part responsible).

46 The -α in the θα of να θα suggests that this may actually represent να θε να, with θε being more like a main verb co-occurring with a preceding modal marker να. This clearly warrants more investigation but given that the facts do not point unambiguously to contact involvement – our main interest here – we leave this for a future study.

The conclusion, therefore, is that, for the most part, it is difficult to prove language contact here as shaping the form and ordering of the pieces of the verbal complex, as the convergence here seems largely to be a mix of language-internal and universal typological factors, with perhaps no more than just some help from contact. However, the sentence-initial clitics of Macedonian do indeed look like a Balkanism.[47] Other parallelisms in the verbal complex that result from these various contributing factors give the appearance of structural commonality across the Balkans regarding this most central sentential unit in each of the languages. In such a situation, then, the form of the verbal complex in any one of the languages may well have been enhanced and reinforced through contact with speakers of another language.

7.4.2 Word Order in Clauses

As with the elements internal to phrases, so also with regard to the major constituents within clauses the Balkan languages are characterized by relatively free constituent order. From a structural point of view, even though in all of the Indo-European Balkan languages, the unmarked word order[48] tends toward S(ubject)-V(erb)-O(bject), all six possibilities for the ordering of the major constituents, i.e., VSO, VOS, SVO, OVS, SOV, and OSV, can be found in all of the Balkan languages. Nonetheless, certain patterns are favored for various types of syntactic and narrative strategies (emphasis, topicalization, focus, contrastive thematization, etc.), often in conjunction with another syntactic process, namely object reduplication, the cross-indexing of an object by a weak pronoun discussed at length in §7.5.1. Despite this freedom, which might make it hard to observe any changes in clausal word order, there are some contact-related developments to be noted that pertain specifically to the order of elements in a clause.

For instance, Matras 2009: 245 observes that Romani changed to a verb-medial order, moving away from the verb-final pattern reconstructed for earlier Indic (and attested in some stages). He sees this "as a result of contact with Byzantine Greek," a language where verb-medial structures were quite common, and argues that it "comes along with the re-organisation of all subordinate clause types."

Similarly, in Turkish in general, the word-order tendency is strongly for verb-finality.[49] However, West Rumelian Turkish and Gagauz both show SVO or even

47 A somewhat similar construct, though admittedly not determining an independent sentence like the Balkan verbal complex, but figuring in argumentation for a sprachbund, is what Watkins 2001: 54–55 calls the "'chains' of particles [i.e., connectives] and local and anaphoric pronouns after the first stressed word of the sentence" in his ancient Anatolian sprachbund consisting of Anatolian Indo-European languages (especially Hittite, Palaic, and Luvian) and the non-Indo-European Hattic and Hurrian. We can also note here that BCMS dialects in Romania tend to have sentence-initial clitics as in Romanian.

48 This is not the place for a detailed discussion of the problematic aspects connected with the notion of an unmarked or basic word order, except to signal our awareness of these problems. See Philippaki-Warburton 1985 for an insightful consideration of the relevant issues with regard to Greek, for instance.

49 Ergüvanlı 1984, however, in her discussion of word order in Turkish, offers numerous examples of non-verb-final order in contemporary standard Turkish, especially in conversational registers.

VSO tendencies, presumably due to contact with SVO Indo-European languages. The same is true for the Turkish of Komotini (Trk Gümulcine), which is an ERT dialect (Petrou 2018). Friedman 1982c: 70 supplies several examples of various SVO using the verbs *gitmek* 'go' and *söylemek* 'say' from Macedonia, and Ibrahimi 1982: 53 gives examples from Kosovo (see also example 7.120). Menz 1999 has many examples of non-SOV order using full verbs in Gagauz under the influence of various Indo-European contact languages, and Petrou's 2018 examples also use many full verbs. Matras 2009: 251, writing on the WRT of Gostivar, however, makes the following observation:

> Changes in word-order structure are usually assumed to be gradual, emerging via loss of pragmatic specialization of secondary word order variants. Macedonian Turkish retains its SOV order, contrasting with the surrounding languages in the Balkans, but it does show convergence in the position of the copula in existential predications. While in Turkish the copula is enclitic, Macedonian Turkish places the non-verbal predicate after the copular. (Matras & Tufan 2007; Tufan 2007 [*sic*, for 2008])

He gives these examples to illustrate the innovative Balkan Turkish word order:

(7.32) a. Hised-ıl-mes ki vardır sonbaar
feel-PASS-NEG.AOR COMP exist.COP.3SG autumn
'It does not feel like [= that there is] autumn'

b. Siz i-dı-nız ev-de
you COP-PST2PL house-LOC
'You were at home'

In each case, standard Turkish would show verb-final order (*sonbahar vardır* / *evdeydiniz*).[50] While the copular clitic examples are certainly the most striking, the fact is that in general Gagauz, most WRT, and even the ERT of Greek Thrace all show far more SVO order than StTrk. As Petrou 2018: 303 observes for Komotini (Trk Gümülcine), while StTrk permits a wide variety of orders for pragmatic purposes such as focus or topicalization (see footnote 49), the examples in the relevant Balkan Turkish dialects involve unmarked order and would not occur as such in StTrk.

There are other convergent aspects of Balkan clausal word order but not necessarily as a result of language contact so much as other types of factors. For instance, the predicate-initial pattern seen in the presentational sentence-type discussed in §7.8.1.2, e.g., Grk Νά ο Γιάννης 'Here's John,' Mac/Blg *Eve/Eto ja knigata* 'Here's the book,' Alb *Ja këtu* 'Here it is,' Aro *Ia calu* 'Here is the horse,' is probably a matter of the shared pragmatics of the construction in question, in that the predicate (νά, *eve, eto, ja, ia*, etc.) is the key element pointing towards the nominal being presented as coming into view, and thus most likely pragmatically favored to occur in initial position.

These examples, however, have pragmatic implications not necessarily relevant to the Balkan Turkish examples.

50 In conversational WRT in North Macedonia, as already noted, there are many other non-verb-final examples, e.g., *Gittık Selânikte* 'we.went to.Thessaloniki,' which also has locative for dative (Kakuk 1972: 245).

Moreover, there is a fair amount of convergence in word order of subordinate clauses with a DMS, in two ways. First, as noted in §7.4.1.2, the DMS always precedes the subordinate verb and cannot be separated from it by anything other than prosodically weak ("clitic") material.[51] Second, as discussed more fully in §7.7.2.1.3.1, the DMS markers do not sharply delimit a subordinate clause boundary, so that subordinate-clause material, most notably subject nominals such as *našite* in (7.33a), η ομάδα μας in (7.33b), *ti* in (7.33c), *me* in (7.33d), or *tini* in (7.33e) can "leak" to the left of the DMS giving orders like $Subj_1$-$Verb_1$-$Subj_2$-DMS-$Verb_2$ in Greek and Balkan Slavic and at least some dialects of Romani (e.g., Kočani Arli) and Aromanian (e.g., Ohrid):

(7.33) a. Iskam našite da spečeljat (Blg)
want.1SG ours.DEF DMS win.PFV-3PL.PRS
'I want our (team) to win' (lit., 'want ours (= team) that they.win')'

b. Εμείς ελπίζουμε η ομάδα μας να νικήσει (Grk)
we.NOM hope.1PL the.NOM team.NOM our DMS win.PFV.3SG
'We hope that our team will win.'

c. Sakam ti da odiš (Mac)
want.1SG.PRS you.NOM DMS go.2SG.PRS
'I want *you* to go'

d. Ov mangela me te khelav čučeko (Rmi)
he want.3SG.PRS I DMS dance.1SG.PRS chochek
'He wants *me* to dance the chochek'

e. Mini voi tini s' nedz (Aro)
I want.1SG.PRS you.NOM DMS go.1SG
'I want *you* to go'

However, this is not the case for Albanian or Romanian. In Albanian, a complementizer (*që*) is required if $Subj_2$ is present, as in (7.34), and in Romanian, the complementizer (*că*) is required if the subject is explicit, as in (7.35):[52]

(7.34) a. Shpresoj të jesh mirë

b. Shpresoj që ti të jesh mirë

c. *Shpresoj ti të jesh mirë
I.hope COMP you.NOM DMS are.2SG.SBJV well
'I hope that you(a)/YOU(b) are well'

(7.35) a. Vreau să plece
want.1SG DMS leave.3SG
'I want him to leave'

51 This is a crucial difference between Balkan Slavic and BCMS, which does permit such separation. Under the influence of BCMS, nonclitic material sometimes comes between *da* and the verb in contexts such as Macedonian news reports. Romanian, unlike other Balkan languages, also permits the intensifier *mai*, lit., 'more,' to the right of the DMS, e.g., *am uitat să mai fim oameni* 'we have quite forgotten to be humans' (a song by Vescan Daniel Leonard, Mahia Beldo, and Dorin Oswin).

52 In Romanian and Albanian, the complementizer can be dropped, but the construction is extremely marginal, and, for Albanian, "foreign" (*i huaj*). By contrast, in, e.g., Macedonian, the complementizer is ungrammatical: *Sakam [*deka] ti da odiš* for 'I want [*that] you to go' is abnormal, not normal. There is thus a fundamental difference between Albanian and Romanian, on the one hand, and the other Indo-European Balkan languages, on the other, with regard to "leakage" of this sort.

b. Vreau că Ion să plece
want.1SG COMP John DMS leave.3SG
'I want John to leave'

The positioning of the DMS relative to the verb it governs reflects the historical sources of the subordinating marker, but the differences in its "porousness" cannot be explained simply by the fact that in each language it is prosodically weak and has an affix-like status which makes it essentially like a part of the verb and not a clear marker of a clause-boundary. Rather, the fact that (Geg) Albanian and Romanian have the strongest remnants of an infinitive (Joseph 1983a) demonstrates that the "porousness" of the DMS is dependent on additional factors. The fact that Albanian and Romanian share a restriction different from Balkan Slavic, Greek, and SDBR suggests a combination of substratum and areal factors.

7.5 The Syntax of Clitics, Historical and Otherwise

The clitic elements whose historical existence is justified in §7.3.3 share a certain positional tendency, generally occurring in second position in their domain, positioned after the first accented element. This tendency, referred to already in §7.4.1.2.2.2.1, can be elaborated upon here. It was noted and discussed at some length by Wackernagel 1892, so that it is often referred to as "Wackernagel's Law" positioning, with the phenomenon in general being referred to as "second position clisis" and the elements so positioned as "second position clitics." The examples in (7.36) show such placement of the object pronouns from (7.1):

(7.36) a. Hitt: nu-mu memēr (*Annals of Mursilis*, Yr 10, l. 33)
and-me.DAT said.3PL
'And they said to me'

b. Grk: ἄνδρα μοι ἔννεπε (Hom *Od*. 1.1)
man.ACC me.DAT say.IMPV.2SG
'Speak to me of the man'

c. Skt: rámadhvam me vácase (Rigveda 3.33.2a)
rest.IMPV.2PL me.DAT speech.DAT
'Rest for my speech'

d. OCS: tako mi bozi (Supr. 74:21)
thus me.DAT gods.NOM
'By the gods!'

e. OIr: ní-m-charat-sa (Wurzburg Glosses 5^{c}6)
not-me-love.3PL-EMP
'They do not love me'

This tendency is generally taken to reflect a Proto-Indo-European syntactic phenomenon (whether a "rule" in some technical sense or a construction or a result

of other sorts of processes), that is, an aspect of PIE syntax that can be reconstructed and thus taken as part of the inheritance into the different branches of the family (see Fortson 2010: 161–163, and Wanner 1987: 79–85, for instance). More accurately, though, such a Wackernagel's Law phenomenon involves placement of the clitic element in second position within a specified domain, and the nature of the domain depends on the particular clitic. Some are positioned relative to whole clauses and have a sentence as their scope, some are positioned relative to phrases, and some are positioned relative to words; still, the key fact is that within their specific domain, they occur in second position, in some sense. The sentence-, phrase-, and word-connective element defined by Grk τε, Skt *ca*, and Lat *que* (among other forms in other languages) 'and' allows for the reconstruction of a PIE *k^we, and this element gives clear evidence of such a second-position placement tendency, exemplified in (7.37) from Sanskrit:

(7.37) a. Word coordination, e.g., [NN-**ca**]:
vāyav indraś **ca** (Rigveda 1.2.5a)
Vāyu.VOC Indra.NOM and
'O Vāyu and Indra!'

b. Phrase coordination, e.g., [X N] [ADJ-**ca** N]$_{NP}$
calitaḥ gacchatā **ca** tena (Hitopadeśa 3.4)
set.out.M going.INS and him.INS …
'He set out and by him, while going, …
(lit., 'by-(the-)going and (by-)him', i.e., 'and by the going him')

c.i. Sentence coordination, e.g., [NP VP] [X-**ca** VP]
āpaḥ prṇīta bheṣajám várūthaṃ (Rigveda 1.23.21)
waters.VOC fill.IMPV.2PL healing protection
mama jyók **ca** sūryaṃ dṛśe tanvè
my long and sun.ACC see.INF body.DAT
'O waters, grant in fullness healing, protection for my body,
and (grant for me) to see the sun for a long time'

c.ii. Sentence coordination, e.g., [NP VP] [DET-**ca**-N$_{NP}$ VP]
sa saṃvardhitaḥ. taṃ **ca** múṣikaṃ khāditum (Hitopadeśa 4.6)
he.NOM nourished.NOM that.ACC and mouse.ACC eat.INF
anudhāvan biḍālo dṛṣṭaḥ
running.NOM cat.NOM seen.NOM
'He was nourished. And a cat was seen running after that mouse to eat it.'

This brief discussion provides the necessary proto-language foundation for understanding the historical background to the syntax of Balkan clitics. All of the Indo-European Balkan languages inherited not only some specific clitic forms – pronouns, connectives, and some verbal forms (see (7.1)) – but also some specific clitic behaviors. Any convergences or divergences found in the latter instantiations of these languages must be judged against that starting unity they showed by virtue of their Indo-European patrimony. The changes in Wackernagelian positioning with regard to Balkan clitics are crucial for our interests here.

The relevant domain for the positioning of clitic pronouns changed between early stages of Greek and Slavic and later stages of the respective languages, moving in the direction of becoming more restricted. Ancient Greek – see Condoravdi & Kiparsky 2001: 31 – and Common Slavic, as represented by OCS, had the clause as the domain for the placement of clitic pronouns, whereas different domains are found in the modern Balkan continuations of these languages. The details are given in §7.4.1.2.2.2.1, but of note here is the fact that in Greek, with some dialectal exceptions, as detailed in Condoravdi & Kiparsky 2001, clitics are positioned relative to the verb, and not relative to elements in the clause. Within Balkan Slavic, there are differences between BCMS on the one hand and Macedonian and Bulgarian on the other, with BCMS being conservative in this regard and maintaining a clause-based placement rule (cf. the examples in (7.27)), and the other languages innovatively positioning clitics relative to the verb (see §7.4.1.2.2.2.1). Moreover, there are further differences between Macedonian and Bulgarian here, as illustrated above in (7.27) and (7.28), in that clitics show a prosodically driven mobility in Bulgarian but a more grammatically determined mobility in Macedonian (as in Albanian, Greek, and Balkan Romance), with the result that sentence-initial and phrase-initial clitics are possible in Macedonian (as also in Albanian, Greek, and Balkan Romance) but not in Bulgarian.

Not enough is known about the prehistory of clitics and their placement in Albanian to allow for a definitive judgment on their development, but the positioning, as stated above, is like that found in Greek and Macedonian, so that a change from at least the PIE starting point can be safely inferred for Albanian too. The same can be said as well about Balkan Romance, which shows a Greek-Albanian-Macedonian type of clitic placement that is different from what occurs in Latin. Interestingly, the Balkan Romance pattern, with some notable exceptions (e.g., the postpositioning in Romanian of the feminine accusative singular *o* 'her' in the perfect, i.e., *am vazut-o* 'I.have seen-her,' not found elsewhere in Romance), is found in other Romance languages, especially Spanish. But even in other Romance languages, a restriction of the domain most relevant for the placement of clitics has occurred, with a shift historically from the clause to the verb, as discussed by Wanner 1987.

While some of the Romance facts outside of the Balkans possibly point to a Romance impetus for the Balkan shift, there still seems to be a key feature based on clitics that differentiates the languages of the Balkans and provides a basis for identifying *Balkan language* in the way that we do throughout this work. In the Balkans, the isogloss of a clausal domain for clitic positioning versus a verb-centered domain participates with such features as the occurrence of an invariant future marker based on the verb 'want,' the presence of certain reductions in the nominal case system, and more widespread use of finite complementation. It thus leads to a grouping of languages that includes Albanian, Greek, Balkan Romance, and Balkan Slavic. As with other features, this one cuts through BCMS in such a way that Torlak is transitional between Bulgarian and Macedonian, on the one hand, and the rest of BCMS on the other (see footnote 51). Thus, a verb-based clitic

positioning, as opposed to a pattern of placement defined relative to the clause, becomes one of the defining features of a *Balkan language*, as the term is employed herein (i.e., a language in the Balkan sprachbund), as opposed to the merely geographic designation of *language of the Balkans*.

Clitic or clitic-like elements figure in many pan-Balkan phenomena and so are discussed at various places throughout this work. For example, the definite article of Balkan Slavic, Balkan Romance, and Albanian has Wackernagel's Law properties in terms of its placement, but despite a few observations in §7.4.1.1, it is treated in §6.1.2.2.1, as it is involved in the realization of a key morphological category in the noun phrase. Similarly, weak dative or genitive pronouns mark possession in various of the languages, but in that regard they are presented in §6.1.1.2.2 as again pertaining to a key aspect of the morphology of the noun phrase. And, there is a phonological dimension to clitics with respect to their prosody, thus occasioning discussion in §5.5.3 (but also §8.1). Consequently, only a portion of what clitics do in the Balkans falls directly under the rubric of syntax. In what follows, then, remaining clitic phenomena – one very important syntactic one along with some minor ones – are discussed in this chapter on syntax.

7.5.1 Object Reduplication[53]

Object Reduplication (henceforth, OR) refers to the co-occurrence of a clitic, i.e., short or weak, pronoun, with the same grammatical role in a clause as a nonclitic direct or indirect object. Thus, the weak object pronoun co-occurs with the object (full, strong pronoun or ordinary noun or noun phrase), and agrees with it in gender, number, and case. The following examples from Macedonian and Greek are typical:

(7.38) a. Mu go davam molivot na momčeto. (Mac)
him.DAT it.ACC I.give pencil.DEF to boy.DEF
'I give the pencil to the boy'

b. Με κοιτάζουν εμένα (Grk)
me. ACC.WK look.at.PL me.ACC.STR
'It's me that they are looking at'

OR is characteristic of all the Indo-European Balkan languages, though its degree of integration into the grammar ("grammaticalization"), and thus its function, differs from language to language. OR in the Balkans shows varying degrees of encoding, as a pragmatic or a grammatical device, on the basis of areal rather than genealogical relation: it is more thoroughly embedded in the grammar, i.e., is more thoroughly "grammaticalized," in the western Balkans than in the east, where it has a more pragmatic function. Moreover, the variation shown by Balkan Slavic,

53 This section is based on Friedman 2008c; the reader is referred to Kallulli & Tasmowski 2008, in which volume Friedman 2008b appears, for more on this phenomenon in the Balkans. The phenomenon also goes by other names, including *clitic doubling, clitic reduplication, object doubling* and *proleptic* or *resumptive pronouns*. See below on the prosodic distinction between this and other similar-looking structures, such as Clitic Left Dislocation (Cinque 1977, 1990).

Balkan Romance, and Albanian points to the areality of this feature and to southwestern Macedonia as the core contact zone (Friedman 1994a).

Reduplicative phenomena involving clitic object pronouns have been noted as characteristic of the Balkan languages since Miklosich 1862: 7–8. Important early work was also done by Seliščev 1918, who pointed out the relevance of definiteness or determinedness of the object for reduplication (though see §7.5.1.2.1 below on specificity). Sandfeld 1930: 192, in mentioning the phenomenon, makes the point that French sentences like *Je le connais, cet homme là* 'I know him, that-there man,' which represents an "after-thought" construction (Hyman 1975), are fundamentally different from Balkan OR. A comparison of Sandfeld's example with its Macedonian equivalent, *Jas go znam onoj čovek/čovekon* 'I know that man there/that-there man' (lit., 'I him know … ') shows that in written French the comma is obligatory, indicating a prosodic pause and syntactic clefting, whereas in Macedonian, the clitic pronoun is an obligatory and prosodically integrated part of the clause (cf. Asenova 2002: 106–107). Thus, while superficially similar and even related in their reference to the same object in the respective discourse, they are really quite different. The same can be said regarding other such discourse-sensitive structures involving dislocated elements on the left periphery, such as so-called Clitic Left Dislocation, e.g., English *John, I can't stand him!*, topicalization, e.g., *John, I can't stand!*, or after-thought specification, e.g., *Where did you put them, the drugs?* as they have intonational breaks that distinguish them from the smooth contour of the Balkan OR.

In what follows, the most salient facts concerning OR in the Balkans are presented: first an overview of the situation in each language or language group, including some dialect information, then some discussion of various pragmatic considerations pertaining to the construction (specificity, aboutness, focus, factivity), followed by a sketch of the relevant historical facts and some ways in which ideology has interacted with the doubling of objects.

7.5.1.1 Language-by-Language Outlines

The following sub-sections are intended to provide the basic facts for each language, with occasional additional details. These facts are supplemented by examples and discussion in the subsequent sections that treat various pragmatic, historical, and ideological dimensions of OR.

7.5.1.1.1 Balkan Slavic

As the sections below make clear, Macedonian and Bulgarian differ in important ways regarding the realization of OR. Those BCMS dialects bordering on Macedonian and Bulgarian also have OR to varying degrees. The facts below make it clear that the current areal situation reflects the historical spread of OR, from the regions with the most consistent and grammaticalized reduplication of

substantival objects, in Macedonian, to their pragmatic encoding in Bulgarian and finally their general restriction to pronominal objects in dialectal BCMS.

7.5.1.1.1.1 Macedonian

The normative prescription for OR, as represented, e.g., in Koneski 1967: 335, states that reduplication is required for definite direct objects and for all indirect objects in Macedonian, as in (7.38a) above; that sentence would be incorrect without the weak object pronouns (i.e., *davam molivot na momčeto). In this way, OR in Macedonian can be said to be grammaticalized, i.e., fully integrated into the grammar, and not (just) a discourse-sensitive pragmatically grounded phenomenon. In colloquial Macedonian, however, the dative clitic can be omitted, which signals focus on the indirect object (Petroska 2008). It is also possible for OR to occur with indefinite direct objects in order to indicate specificity (Friedman 1993d; Petroska 2008). Such an example even occurs in Koneski's 1967: 231–232 normative grammar.

For Balkan Slavic, Seliščev 1918: 246–256 (cf. also 1925: 45) noted that OR is consistently realized in the Macedonian dialects west of the Vardar but is rarer to the east. Citing examples from Daniil's *Tetraglosson* (cf. Leake 1814: 383–402), he found that in most but not all cases, reduplication in the Slavic of the text, which is Ohrid Macedonian, is present also in the Greek, Albanian, and Aromanian, showing the convergent aspect to OR among these languages. In Macedonian, the standard language reflects its western dialectal base in this respect, whereas the situation in the eastern dialects is more like that in Bulgarian (Vidoeski 1960/61: 23).

7.5.1.1.1.2 Bulgarian

OR is prescriptively disallowed in Bulgarian except in the impersonal existential use of *ima/njama*, literally 'have/not.have,' e.g., *răkata ja njama* 'there's no hand'; thus the Standard Bulgarian equivalent of (7.38a) would not require the pronouns and would be simply *davam moliva na momčeto*. OR can be used to disambiguate case relations (Stojanov 1983: 192–193), see §7.5.1.2.2 below. This last function is not obligatory, as such ambiguity can be purposefully exploited, e.g., in press headlines for the purpose of grabbing attention such as *prostitutka ubi policaj*, literally, 'prostitute killed.AOR policeman,' where who killed whom will only be disambiguated in the article.

Keremedčieva 1993: 297–299 surveys much of the available Bulgarian dialect literature which, unlike studies of the standard language, does make extensive use of spoken narrative, albeit not as spontaneous as in some of the modern urban corpora. Nonetheless, her observations are consistent with the previous observations, namely that OR occurs relatively infrequently in eastern Bulgarian, and the frequency increases as one moves west. We can also note Cyxun 1974, who makes the point that the exclusion of OR from standard Bulgarian is artificial (cf. also Orzechowska 1973; also Minčeva 1968 for a historical treatment).

7.5.1.1.1.3 *Varieties of Bosnian-Croatian-Montenegrin-Serbian*

In OR, as in many other crucial features, the dialects of Gora (southwesternmost Kosovo) pattern with Macedonian. The Torlak dialects have reduplication of pronominal objects, but the reduplication of nominal objects does not occur in the dialect literature; still, speakers report that it is possible. In Niš, the phenomenon is much rarer in the town than in the surrounding villages (Toma 1998: 311). Pronominal reduplication also occurs in southern Montenegrin dialects (Ivić 1958: 17). BCMS dialects of Sretečka Župa, just to the east of Gora in Kosovo, which lack the definite article (and other features) connecting Gora with Macedonian, display some substantival OR; cf. (7.39), from Pavlović 1939: 256–257, versus (7.40), from Pavlović 1939: 252:

(7.39) Traživ nekoj koj može da mu odgovori ruskomu caru.
he.sought someone who can DMS him.DAT respond Russian.DAT king.DAT
'He sought someone who could respond to the Russian king.'

(7.40) On ne ćav da ide za magare da potraži, teke žena
he not wanted DMS go for donkey DMS seek and.so wife
otišla te mu dovela magare.
went and him brought donkey
'[The wife has told him to go to the neighbors' to borrow a donkey in order to get wood] He didn't want to go ask for the donkey, so the wife went and brought him the donkey.'

7.5.1.1.2 Greek

OR is generally optional in Greek, governed by pragmatic factors having to do with topic and focus (see §7.5.1.2.1). Holton et al. 1997: 432 characterize OR as "the device that removes the object from the comment (new part) of the sentence and renders it part of the background [known] information)." Thus, in (7.41), Μαρία is topical and would be an appropriate answer to a question like "Does John love Mary?," where Mary is thus part of given information:

(7.41) Την αγαπάει τη Μαρία
her.ACC.SG.F loves.3SG the Mary.ACC.SG.F
'He loves Mary'

As in Macedonian, there is no prosodic boundary in such a sentence, but such breaks are possible with different pragmatics; thus την αγαπάει, τη Μαρία, with "comma intonation" before the nominal, is the after-thought construction mentioned in §7.5.1 ('He loves her, Mary that is').[54] One locution where OR is

54 A left dislocation construction is also possible, with an intonational break, and intonation interacts with word order to give different effects. Janse 2008: 171 gives a particularly perspicacious table of the possibilities, where O = object and V = verb, and "clitic doubling" refers to the co-occurrence of a pronoun with the object:

required is with the pronoun όλος ‘all, everyone/thing’ as object (Holton et al. 1997: 194–195):

(7.42) τα ξέρει όλα / *ξέρει όλα
them.ACC.NTR.PL knows.3SG all.ACC.NTR.PL
‘He knows it all’ (lit., ‘them he-knows all-things’)

Although it is found across all of Greek, including Asia Minor Greek dialects (Pontic and Cappadocian; cf. Janse 2008), according to Ilievski 1973, OR is particularly characteristic of northern Greek dialects (see also §7.5.1.2.1).

7.5.1.1.3 Albanian

The requirements on OR in Albanian are similar to Macedonian but more restricted as to the contexts where it occurs (see §7.5.1.2.1). A newly introduced or emphasized direct object will not be reduplicated (Sh. Demiraj 2002: 227). Buchholz & Fiedler 1987: 443 (cf. also Buchholz 1977: 180) states that OR with definite direct objects in Albanian is optional with canonical (SVO) word order, as in (7.43), but omission of the reduplicating pronoun is not possible with indirect objects, as in (7.44):

(7.43) Agim-i po Ø/e vështron hënë-n.
Agim-DEF PROG Ø/it watches moon-ACC.DEF
‘Agim is watching the moon.’

(7.44) Çfaqj-a *Ø/i pëlqeu Agim-it.
show-DEF Ø/him pleased Agim-DAT.DEF
‘The show pleased Agim (= Agim liked the show).’ (Buchholz 1977: 180)

Buchholz 1977: 188–189 points out that even indefinite direct objects can trigger obligatory OR if they are preposed, as seen in (7.45) and (7.46) for ‘They want everything ready’:

(7.45) *Çdo gjë* *Ø/*e* duan gati.
every thing Ø/it they.want ready

(7.46) Ata Ø/*e duan gati çdo gjë.
they Ø/it they.want ready every thing

Cf. also Kallulli 2008: 248, who argues that topicalization has an effect on whether OR is required or not in Albanian, but not in Greek. As for Albanian dialects, Muhurr, as a representative Central Geg dialect (Ylli & Sobolev 2002: 63, 69–70) and Leshnja, as a representative Northern Tosk dialect (Ylli & Sobolev 2003: 42, 48), show differences that pattern with neighboring languages with which they are in contact; Muhurr, in the Dibra (Mac Debar) region parallels Macedonian

topicalization OV +clitic doubling -boundary pause
topic left-dislocation OV +clitic doubling +boundary pause
backgrounding VO +clitic doubling -boundary pause
topic right-dislocation VO +clitic doubling +boundary pause.

OR usage, whereas Leshnja, in the Skrapar region, shows OR patterns more like Greek.

7.5.1.1.4 Balkan Romance

Balkan Romance, like Balkan Slavic, shows a range of realizations of OR, from a nearly fully grammatically determined version in SDBR to a more pragmatically determined version in Romanian.

7.5.1.1.4.1 Romanian

OR is more limited in Romanian than in other languages (Graur et al. 1966: 144–147; Farkas 1978: 93–96; Steriade 1980; Tasmowski-De Ryck 1987: 382–383). The factors of definiteness, position, and humanness affect the realization of reduplication: (a) definite direct and all indirect substantival objects must reduplicate when preverbal; (b) definite or personal pronominal objects must reduplicate; but (c) postverbal direct objects cannot reduplicate if not governed by *pe* (which for substantives is limited to humans). There are also contexts where it is normatively prohibited, and here the additional factors of partitivity and specificity, come into play;[55] that is, reduplication is not permitted with nonpersonal indefinite pronouns, and nonpersonal substantives unless pre-verbal and marked as both specific and partitive (Farkas 1978: 93–96; Tasmowski-De Ryck 1987: 382–383).

Moreover, the social context is crucial; Friedman 2008c: 54 notes, for instance, that as in Bulgarian, formality matters, so that "in a formal invitation sent by an embassy, an expression such as [(7.47)] would be used and reduplication would not be used" (Tasmowski p.c.):

(7.47) X are onoarea de-a invita pe doamna profesoară YZ.
X has honor of-to invite PE Mrs. professor YZ
'X has the honor of inviting Professor YZ … '

The connection between formality and OR makes sense, as Friedman, drawing on the insights of Tasmowski (p.c.), argues, since OR entails a degree of familiarity with the topical object, whereas formality implies a certain (social) distancing. As described by Tasmowski-De Ryck 1987, the involvement of topicality as a factor in Romanian OR, but not the sole determiner, is reminiscent of the way explicitness and aboutness condition OR in Bulgarian in the account given by Leafgren 2002 (see §7.5.1.2.2 below).

7.5.1.1.4.2 Aromanian and Meglenoromanian

SDBR is important here since these Balkan Romance varieties show most clearly the contact nature of OR. The Aromanian of North Macedonia and Meglenoromanian pattern with Macedonian, and thus contrast with Romanian in requiring all definite direct objects to be reduplicated. The Aromanian of

55 Partitivity also optionally allows for reduplication.

Aminciu (Grk Métsovo) in Greece, however, follows the pragmatic constraints of Greek rather than the grammatical requirements of Macedonian. Thus, examples (7.48a) and (7.48b) show reduplicated direct objects in Aromanian and Meglenoromanian respectively, parallel to Macedonian usage and different from Romanian usage (here and in further examples with reduplication elements in italics):

(7.48) a. Auš-lu nu vrea s-l' *u*-aspargă *k'efe-(a)* a fičor-lui. (Aro (Kruševo))
old-DEF not wanted DMS-him it-spoil pleasure-DEF to child-DEF.DAT
'The old man did not want to spoil the child's pleasure' (Gołąb 1974: 37)
b. *lă* loa *bucium-ul*, lă turi shi zisi (Megl)
it.ACC took log-DEF it.ACC threw and said
'He took the log, threw it [away] and said: … ' (Papatsafa 1997: 27, 1999: 15)

On the other hand, examples (7.49a–c) show that the Aromanian of Aminciu (Grk Métsovo) patterns with Greek regarding fronting of an object (see §7.5.1.1.2):

(7.49) a. Kinele muske fčorulu. (Aro (Aminciu))
dog.DEF bit boy.DEF
'The dog bit the boy' (Beis 2000: 382)
b. *Fčorulu lu* muske kinele. (Aro (Aminciu))
boy-DEF him.ACC bit dog-DEF
'As for the boy, the dog bit him' (Beis 2000: 232)
c. *Tute* nu *lji* shtiu.
everything not it.ACC.PL I.know
'I don't know everything' (Beis 2000: 449) (Aro (Aminciu))

Example (7.49c) parallels the Greek of (7.42), but also Romanian *El le ştie pe toate* 'He knows everything' (lit., 'he it.ACC.PL knows on all.PL'). In this locution, Balkan Romance is closer to Greek than to Balkan Slavic; see §7.5.1.1.7 below, and (7.54), regarding the Romani of Agía Varvára.

7.5.1.1.5 Judezmo

Judezmo has OR with preposed objects, as the proverb in (7.50a) from Macedonian Judezmo indicates (Kolonomos 1995: 267), which parallels exactly the Macedonian equivalent in (7.50b):

(7.50) a. *Il palu tuertu* la lumeri *lu* indireche (Jud)
the stick crooked the fire it.ACC straightens
b. *Kriv stap* ogn-ot *go* ispravuva (Mac)
crooked stick fire-DEF it.ACC straightens
'A crooked staff is straightened in the fire'

Otherwise, however, Judezmo does not show reduplication in contexts where Macedonian does.

OR is found in Spanish, as in some other Romance languages outside the Balkans, and the phenomenon as realized in Judezmo does not differ significantly from Spanish, although Wagner 1914: 130–131 observes that reduplicated object pronouns occur more frequently in Constantinople Judezmo than in Spanish, a possible contact effect.

And example (7.50) would not have reduplication in Modern Spanish. Still, the parallels with Spanish must be taken seriously so that even though Judezmo fits in with the pattern of OR in the Balkans in general, this could just as easily be a function of Judezmo's Hispanic heritage as its Balkan aspect.

7.5.1.1.6 West Rumelian Turkish

According to Friedman 2008c, West Rumelian Turkish does not differ from the rest of Turkish regarding OR, which is to say that there is no evidence of the phenomenon.

7.5.1.1.7 Romani

Romani occupies a middle position similar to that of Greek, which was its earliest and most significant Balkan contact language historically. Thus, Romani OR tends to occur with preverbal objects and topicalized object pronouns. Examples (7.51) and (7.52) are typical in this respect:

(7.51) *O melalo pani* na piena *le* ni o džungale ruva.
the dirty water not drink it.ACC nor the bad wolves
'Even wicked wolves do not drink dirty water' (Jusuf 1996: 125)

(7.52) *Man* ma axmize *man* kidisave bucende
me don't embroil me this work.PL.LOC
'Don't mix me up in this business' (Bugurdži, Jusuf 1974: 14)

Examples (7.53ab) show how OR in Skopje dialects of Romani does not correspond to the strongly grammatically determined Macedonian OR system with which it is in intimate contact. They were recorded in July 1994 from a single broadcast of a Skopje Romani radio music-request program (*Gili pali gili* 'Song after song') in which the announcer switched freely back and forth between Romani (7.53a) and Macedonian (7.53b), consistently reduplicating in Macedonian, but not Romani:

(7.53) a. O Ajnuri thaj o Džemo … bahtaren e pranden
the Ajnur and the Džemo congratulate the marriage.ACC
e Ramijeske thaj e Mirsadake
the Rami.DAT and the Mirsada.DAT
'Ajnur and Džemo … congratulate Rami and Mirsada on their marriage'

b. Naza i Oli … im *go* čestitat *brak-ot*
Naza and Oli them it congratulate marriage-DEF
na Rami i Mirsada
to Rami and Mirsada
'Naza and Oli … congratulate Rami and Mirsada on their marriage'

In the Romani of Agía Varvára, a suburb of Athens, the dialect has calqued the Greek construction using 'all' (Igla 1996: 161), as in (7.42) and (7.49c) above:[56]

56 Igla notes that the plural accusative resumptive pronoun in Romani shows this to be a literal loan-translation and not a reflection of the Romani system, which would treat *sa* here like English *everything*, i.e., as a singular.

(7.54) Džanes *len* *sa*
you.know it.ACC.PL all
'You know everything'

In most Balkan Romani dialects there is no lexical verb meaning 'have,' and instead an existential construction is used that requires an accusative weak object pronoun. This pronoun reduplicates the possessor, which is in the nominative case if a substantive but the accusative case if a pronoun, as in (7.55) and (7.56). This is a unique type of reduplication as far as the Balkans are concerned.

(7.55) *I* *daj* si *la* duj čhave
the.NOM mother is her.ACC two children
'The mother has two children'

(7.56) *Man* si *ma(n)* duj čhave
me.ACC is me.ACC two children
'I have two children'

See also Bubenik 1997 for additional Romani examples.

7.5.1.2 Pragmatic Conditions on Object Reduplication

The previous section makes clear some of the grammatical constraints governing OR in the various languages. But, as stated already, there are also pragmatic constraints at work. An important study in this regard is Lopašov 1978, focused entirely on Balkan OR. He finds (p. 123) that OR constitutes essentially the same phenomenon in Macedonian, Albanian, Romanian, Greek, and Bulgarian, but that while the initial (pragmatic) impetus for the construction was the same, the way it is ultimately realized in each language is not. Aside from the differences in degree of grammaticalization, Lopašov argues that both language-external and language-internal factors have encouraged these developments. From this, it follows that the conditions governing OR differ language by language, according to language-specific factors. Still, Lopašov 1978: 26, 57, 58, cited in Asenova 2002: 110, gives various general tendencies for OR (following the translation of Friedman 2008c: 40):

i. most often marked with a definite article;
ii. more often pre-verbal than post-verbal;
iii. especially common when the object is a personal pronoun;
iv. indirect objects are reduplicated more often than direct objects;[57]
v. objects that are not definite are not reduplicated.[58]

Moreover, while Lopašov recognizes that the causes of OR could be independent in each language (see also Keremedčieva 1993: 297–299), he

57 Asenova 2002: 110 points out, however, that in Bulgarian dialects the frequency of OR is equal for direct and indirect objects.

58 This last is not the case. See §7.5.1.2.1. Note also that Aromanian patterns with either Macedonian or Greek, Meglenoromanian patterns with Macedonian, and Romani and Judezmo are more like Greek.

nonetheless firmly holds to the view that language contact was at least responsible for its expansion. Finally, he ranks the languages as to degree of grammaticalization of OR, from most to least as follows: Macedonian, Albanian, Romanian, Greek, Bulgarian, an order which accords with the facts and to some extent the history.

7.5.1.2.1 Indefiniteness and Specificity

As the facts in §7.5.1.1.3 indicate, definiteness and indefiniteness interact with the syntactic factors of positioning and grammatical relation in the realization of OR in Albanian. Recognition of the importance of referentiality was furthered by Kazazis & Pentheroudakis 1976 where, basing themselves on Greek and Albanian, they conclude that indefinite direct objects can, or even must, be reduplicated if they are specific, thematic, or contrastive. This work speaks to the problems with Lopašov's point (v) cited above in §7.5.1.2, and supplies a fundamental challenge to older normative accounts of Balkan OR. In fact, Koneski himself has examples of OR with indefinite direct objects even in the prose of his normative grammar (see Friedman 2008b: 41–43). Kazazis & Pentheroudakis 1976 foreshadows later work by scholars such as Kallulli and Leafgren (see below) identifying topicalization or 'aboutness' as a conditioning factor. (Cf. also Berent 1977 on specificity in Macedonian.) Kazazis and Pentheroudakis note, too, that Romanian behaves similarly but with an added complicating factor of animacy. Their key Greek example (p. 400) is (7.57):

(7.57) Σου (τό) πλέκω ένα πούλοβερ
you.DAT (it.ACC) I.knit one sweater
'I'm knitting you a sweater / I'll knit you a sweater'

To be acceptable, (7.57) cannot answer questions like 'What are you doing?' or 'What are you knitting [for me]?' but must instead be in a discourse where *sweater* is already the topic of conversation. They find a similar situation in Albanian, where reduplication can also be expected in cases of topicalization.

Kallulli 2000: 218–219 compares Albanian and Greek and finds similarities and differences, working with the questions in (7.58) and the answers in (7.59), to which the Standard Macedonian and Bulgarian equivalents have been added:

(7.58) a. What did Ana do?
b. What did Ana read?
c. Who read the book?
d. What did Ana do with the book?

(7.59) a. Ana Ø lexoi libr-in. (Alb)
A. read book-DEF.ACC
b. Άννα Ø διάβασε το βιβλίο. (Grk)
A. read the.ACC book.ACC
c. Ana *e* lexoi *libr-in*. (Alb)
A. it read book-DEF.ACC

d. Άννα *το* διάβασε *το* *βιβλίο*. (Grk)
A. it read the.ACC book.ACC

e. Ana pročete kniga-ta. (Blg)
A. read book-DEF

f. Ana *ja* pročita *kniga-ta.* (Mac)
A. it read book-DEF
'Ana read the book'

Kallulli observes that (7.59a) and (7.59b) answer questions (7.58a) or (7.58b) whereas (7.59c) and (7.59d) answer questions (7.58c) and/or (7.58d). Albanian and Greek differ in that Albanian requires the clitic if the sentence is an answer to (7.58c) or (7.58d) whereas Greek permits the omission of the clitic regardless of the question, though with a preference for the clitic in contexts where Albanian requires it. The reduplicated object form in (7.59f) is the only acceptable possibility for Standard Macedonian whereas the Bulgarian norm would prescribe no reduplication, as in (7.59e). As Leafgren's 2002 data demonstrate, however, the expectation is that colloquial Bulgarian would pattern like Albanian and Greek.

Kallulli 2000: 200–223 makes a similar point about lexical focus as marked by 'even,' since the nonfocused item in a simple SVO sentence is the topic. Unlike in Macedonian, a focused object cannot be doubled in Greek and Albanian, while a topical direct object can be doubled in Greek but must be doubled in Albanian; in (7.60), Tirana is the focus, whereas in (7.61), the Pope is the focus and Tirana is the topic:

(7.60) a. Papa Ø vizitoi madje Tiranën (Alb)
Pope.DEF Ø visited even Tirana.DEF.ACC

b. Ο Πάπας Ø επισκέφτηκε ακόμα και τα Τίρανα (Grk)
the Pope Ø visited even and the.ACC Tirana

c. Papata *ja* poseti duri i *Tirana* (Mac)
Pope.DEF it visited even and Tirana
'The Pope visited even Tirana'

(7.61) a. Madje Papa *e* vizitoi *Tiranën* (Alb)
even Pope.DEF it.ACC visited Tirana.DEF.ACC

b. Ακόμα και ο Πάπας (*τα*) επισκέφτηκε *τα* *Τίρανα* (Grk)
even and the Pope (it.ACC) visited the.ACC Tirana

c. Duri i Papata *ja* poseti *Tirana* (Mac)
even and Pope.DEF it visited Tirana
'Even the Pope visited Tirana'

7.5.1.2.2 Object Reduplication and Definiteness and Aboutness in Bulgarian

Two studies since Lopašov 1978 have been done on OR from a functional perspective, both, as it happens, on Bulgarian, Guentchéva 1994 and Leafgren 2002. Guentchéva, building on the work of Ivančev 1978: 137–149, 160–166, discusses examples (7.62) and (7.63) in connection with two conditions often given for OR in Bulgarian, namely OV word order and case disambiguation (cf. §7.5.1.1.1.2).

(7.62) Kučeto *ja* goni *edna* *kotka*.
dog.DEF.N it.ACC. chases one cat.F
'It's the dog that is chasing a cat' (Guentchéva 1994: 111)

(7.63) KRUŠATA risuva deteto.
pear.DEF.F draws child.DEF.N
'The child draws THE PEAR' (Guentchéva 1994: 109)

In (7.62), the pronoun refers to an object that is specific but indefinite; (7.63) answers a question like "What do you see in the picture?" (Guentchéva 1994: 116).

In (7.63), there is a marked word order of OVS but once 'pear' is emphasized (indicated by upper case), it is unambiguous without reduplication (Guentchéva 1994: 109), with 'child' as topic and 'pear' as focus.

However, Asenova 2002: 113–115 states that reduplication in Bulgarian is impossible when true indefiniteness is involved, as with a bare indefinite such as that in (7.64):

(7.64) Prikazka Ø/*ja razkazvaše vsjaka večer.
story Ø/it. F.ACC tell.3.SG.IMPF every evening
'S/he used to tell a story every evening' (Asenova 2002: 114)

For Guentchéva, *thématisation*, i.e., 'topicalization,' characterizes the function of reduplication. Asenova 2002: 113–115, by contrast, argues that when Bulgarian pronouns bear the logical stress they cannot be reduplicated and further that any object that has any sort of 'identifier' (Blg *opredelitel*), essentially a modifier or determiner, can potentially be reduplicated. For the most part, these studies line up well with Kallulli 1999, 2000, in whose view OR serves to mark topicalization in Albanian and Greek, and topicalization and focus are in complementary distribution. Yet, as the above sections make clear, Albanian requires OR in contexts where in Greek (and in Bulgarian) it is optional.

Leafgren 2002, working with a large corpus of colloquial data, has examined actual usage and thus has developed corpus-based insights into speakers' linguistic behavior. His main thesis (p. 197) is that OR marks *aboutness*, usually understood contrastively (i.e., 'about X as opposed to Y'). Bulgarian uses OR, he writes, almost always as an overt marker of topicality when there is clause-level aboutness concerning the object that is unexpected as not being a discourse-level theme. As an example, consider (7.65), from Leafgren 2002: 180: the discourse theme, by which is meant the topic of the context, has to do with a philanderer, but the specific clause with reduplication has adolescent greed as its topic.

(7.65) No izvednăž zad gărba mu ostana ljubopitstvoto kăm
but suddenly behind back.DEF him.DAT remained curiosity.DEF to
ženskoto tjalo, njamaše *ja* poveče *junočeskata* *lakomija* da natrupva opit
female.DEF body not.have it more adolescent.DEF greed DMS accumulate experience
i toj se čuvaše da kazva na prijatelite si …
and he INTR heard DMS says to friends.DEF self.DAT

'But all of a sudden he lost his curiosity about the female body, the adolescent voracity for accumulating experience was now gone, and he was heard to say to his friends … '

Leafgren is able to use the concept of *aboutness* to account for topicality and focus not always being in complementary distribution, as recognized by Guentchéva (cited in Leafgren 2002: 177) and others; consider (7.66), from Leafgren 2002: 149:

(7.66) – Na piano svirja veče s dve răce. – Na pianoto! – Da. – I kakvo sviriš?
on piano I.play already with two hands. on piano.DEF; yes and what.kind you.play
– Razni pieski. Ama *edna* ošte ne săm *ja* naučil, zaštoto e mnogo trudna.
different pieces but one still not am it learn-LPT because is very hard
'–I already play the piano with two hands. – The piano! –Yes. –And what do you play?
–Various pieces. But one I haven't learned yet because it's very hard'

Leafgren is also able to give an example from an oral corpus to show that while reduplicated topics generally are specific, they are not always so; in (7.67), from Leafgren 2002: 176, the discussion centers on markets:

(7.67) *Banan* ne običam da *go* jam.
banana not I.like DMS it.ACC I.eat
'I don't like to eat bananas'

Koneski 1967: 292 cites a nineteenth-century Prilep example from Cepenkov's Macedonian folklore collections (see, e.g., Cepenkov 1972ab):

(7.68) Star*čovek* da *go* pregrnuvaš vo son boles kje te fati
old person DMS him.ACC embrace.2SG.PRS in dream disease FUT you.ACC grab.3SG.PRS
'If you embrace an old person in your sleep, […] you'll get sick'

As in (7.68) from Macedonian and (7.67) from Bulgarian, the direct object here affected by OR is a bare indefinite, contradicting Asenova's claim (see above regarding (7.64)) that reduplication does not occur with bare indefinites. In fact, OR could apply in (7.64), but its interpretation would then differ from a parallel sentence without reduplication.

7.5.1.3 A Further Function: Factivity

Albanian and Greek match Macedonian in many ways, but generally do not use OR in as many contexts. However, there is one place where OR occurs in Albanian and Greek where in Macedonian it typically does not, namely to signal, proleptically, a complement clause to a factive verb or where factivity is an issue. Buchholz & Fiedler 1987: 442 state that in Albanian OR is obligatory for clausal objects with verbs of thinking and perceiving when that object clause expresses a "determined thought"; thus there is OR in (7.69a) in Albanian, but the Macedonian counterpart in (7.69b) is marginal. That is, such a sentence is

possible at least in some colloquial registers,[59] but normally the pronoun *go* would not occur; however, if the complement is preposed, then the pronoun is required, as in (7.69c):

(7.69) a. *E* dija se do të vonohet. (Alb)
it.ACC I.knew that FUT DMS be.late
b. (#*Go*) znaev deka kje zadocni. (Mac)
it.ACC I.knew that FUT be.late
'I knew that he would be late'
c. Deka kje dojdeš *go* znam
COMP FUT come.PFV.PRS.2SG it.ACC know.1SG.PRS
'That you will come, I know' (lit., ' … I know it')

Kallulli 2008 explains this phenomenon for Albanian by referring to the observation that a reduplicated complement is a topic and thus treated as a given, as something presupposed and therefore a fact. Thus, for example, while OR would be required with *di* 'know,' it would not occur with *besoj* 'believe'; however, if the belief were accepted as fact, OR would apply. Thus, for example, in (7.70), with the resumptive pronoun, the qualification in square brackets is ungrammatical, whereas without the resumptive pronoun, a nonfactive reading renders the contradiction in square brackets acceptable; that is, with OR, the clausal object is presented as a fact, making it infelicitous to contradict it:

(7.70) (*E*) besova se Jan-i shkoi [por në fakt nuk shkoi]
it I.believed that J.-DEF left [but in fact NEG leave.2SG.AOR]
'I believed (it) that John left.'

The facts are similar for Greek, both with ξέρω 'know' (cf. (7.71a)) and with πιστεύω 'believe' (cf. (7.71b)):

(7.71) a. (*Το*) ξέρω ότι ο Γιάννης έλεγε ψέμματα
it know.1SG that the John.NOM was-telling lies
'I know (it) that John was telling lies'
b. (*Το*) πίστεψα ότι ο Γιάννης έφυγε
it I.believed that the John left
'I believed (it) that John left'

Here too, as in the examples in (7.70/7.71), OR is optional in Greek but is expected in Albanian. Moreover, in Greek, whereas OR occurs with the factive verb μετανιώνω (που) 'regret (that),' it does not with the "verb of strong emotion" λυπάμαι (που) 'feel.sorry (that),' as noted in Holton et al. 1997: 453, where it is suggested that the often factive complementizer που might be better taken as a causal adverb (i.e., 'feel.sorry because').

59 The Bulgarian equivalent would also be unacceptable. Some speakers of Macedonian from the southwest accept (7.69b), but this is not sanctioned in the literary norm.

In Macedonian, such nuances would be rendered by the choice of tense form in the subordinate clause, specifically confirmative versus nonconfirmative (see §6.2.5.1).

7.5.1.4 Diachronic Development of Object Reduplication

With regard to the history of OR in the Balkans, there may be hints of the construction in optional doubling for clarification in discourse-driven Left Dislocation constructions in Classical Greek (Janse 2008) and Koine Greek (de Boel 2008), but distinct signs of it as a syntactic process appear only in Postclassical Greek. Tzitzilis 2000: 258–259 documents this construction in Medieval Greek. Ilievski 1973 notes that the construction is attested in Vulgar Latin, so that the more robust emergence of it in Greek dates from a time when contact with Latin was already becoming significant. With regard to Albanian, the pre-sixteenth-century situation is unknowable, and OR is not well established in the earliest texts (Asenova 2002: 105). As for Slavic, the construction may date back to OCS (Ilievski 1973),[60] but it does not really gain traction until the early modern period; indeed, eighteenth- and nineteenth-century evidence indicates that pragmatic factors were still relevant even in western Macedonian (Ugrinova-Skalovska 1960/61). Ilievski 1988b: 148–149 believes that some of Ugrinova-Skalovska's examples of the absence of OR may reflect Serbian or Bulgarian influence, but argues nonetheless that even if internal factors were in part responsible for the rise of OR in Balkan Slavic, the absence of the phenomenon from Slavic outside the Balkans points to language contact as a contributing or even decisive factor. Friedman 2019b, using a corpus of modern spoken Macedonian, analyzes the occurrence of object reduplication in the most common obscenity involving 'mother' as the direct object and finds that, aside from Serbian influence, the curse provides evidence for an earlier state of affairs, when OR was more pragmatically conditioned in Macedonian. Thus, for example, the direct object is more consistently reduplicated when it comes before the verb. And, the interpretation of the dialectology of the degree of grammaticalization in Balkan Slavic also argues for language contact as playing a key role in the spread and establishment of OR. The core of the phenomenon is seen in western Macedonia, but the Meglenoromanian and North Aromanian evidence suggests the possibility of a Romance impetus.

60 An example sometimes cited from OCS is from Codex Marianus, Mark 14:51 *i jęsę i junošę*. Taken at face value, this could mean 'and they seized [him] the youth' or 'and the youths seized him,' with the second *i* either a 3SG.M.ACC pronoun or a conjunction, and *junošę* either the (expected) NOM.PL or a scribal error for the ACC.SG *junošǫ*. Diels 1963II: 15 argues for scribal error, and, indeed Codex Zographensis has *junošǫ* while the Grk καὶ κρατοῦσιν αὐτὸν οἱ νεανίσκοι is unambiguously 'and the youths seized him.' The *Dobromirovo Evangelie* (twelfth century), however, has *i jęšę že i junošǫ* 'and they also seized even the youth,' which clearly indicates that the second *i* is an emphatic marker and not a pronoun (Jagić 1954: 74, 1960: 176; Velčeva 1975: 77). Dimitrova-Vulchanova & Vulchanov 2008 argue that there is evidence for OR in the OCS Codex Suprasliensis, but these are reflexives (see §7.5.1.6).

7.5.1.5 Summary Observations About Object Reduplication

OR in the Balkans reveals much about contact-induced change in general and about Balkan contact in particular.

First, it shows how a pragmatic phenomenon can become syntactic and finally grammatically determined, i.e., embedded in the grammar, almost in some instances to the point of morphologization, i.e., being a part of the morphology of the verb. Moreover, its areal distribution and degree of integration into the various Balkan grammars arguably reflects both different diachronic stages and different synchronic systems.

The core for the innovative appearance of OR clearly lies where Central Geg Albanian, western Macedonian, and Northern Aromanian have been in intensive contact for centuries, precisely, that is, in western Macedonia. To account for the situation south of that core, in northern Tosk, northern Greek and southern Aromanian, taking into consideration the early historical attestations in Balkan Latin (and recognizing the difficulty of any certainty about the ancestor of Albanian), it is reasonable to suppose a process there that advanced along a similar path, only slowed by influence from Greek, a more conservative language in this respect.

East of this core, in eastern Macedonian and in Bulgarian, a similar process must have occurred but it has gone only as far as topicalization. However, the beginnings of a purely case marking, i.e., argument-identification, function may be present in Bulgarian with indirect objects lacking the prepositional marker *na*. Such constructions are highly colloquial, and though many such objects are topics, some are not, so that case marking, Leafgren 2002 argues, appears to be the sole motivation:

(7.72) Kakvo mi drănkaš ti mene za turcite (Konstantinov 1895 *apud* Leafgren 2002: 136)
what me.DAT jabber.2SG you me.OBL for Turks.DEF
'Why are you jabbering to me about the Turks, …'

Further, it is also precisely in eastern Macedonian that synthetic dative pronouns are replaced by *na* plus the accusative, as in Bulgarian, creating a similar possibility for a pure case-marking function of reduplication. This tendency to eliminate dative marking is continuing to spread in the present such that younger Macedonian speakers in the west have moved away from dative marking (Markovikj 2019).

Moving further north to Romanian, on the one hand, and BCMS on the other, the phenomenon becomes more restricted. The relevance of such features as humanness and partitivity for Romanian and the restriction to object pronouns (generally denoting humans) in southern BCMS dialects correlate with a higher degree of topicality.

The relative marginality of object reduplication in Judezmo and Romani can be seen to correlate with the marginality of those languages in the Balkans. Topolińska 1994: 121 writes: "Under such circumstances [of a need to be understood in oral

communication in a multilingual environment] the primary candidates for grammaticalization are also those signals that will guarantee successful reference." Unidirectional bilingualism with these languages, i.e., that unlike their neighbors, speakers of Romani and Judezmo learned others' languages but the others did not learn theirs, could well have contributed to the relative rarity of OR; cross-linguistic referentiality would not be an issue, but reverse interference from speaking other languages could affect Roms and Jews.

Second, the patterns of convergence in the core area, even allowing for the possibility of parallel development and the effects of universal principles, are too striking to be attributed to mere parallelism, especially with patterns of multilingualism present in the region. Parallels seen in Tosk, Southern Aromanian, and Greek also suggest a secondary level of contact convergence. A process clearly begun as one of contrast and topicalization, and grammaticalized in the core, remained in the periphery a pragmatic device encoding explicitness.

Third, some of the differences evident in OR, especially, but not exclusively, in Bulgarian and Macedonian, also illustrate the potential for influence from standardization and language ideology on usage and on scholarship. This idea is explored in §7.5.1.7.

7.5.1.6 Reflexive Doubling

Though not a matter of object reduplication per se, reflexive pronouns show a somewhat similar development. They are most frequently objects, and significantly, they can occur in a doubled form. This development requires a separate section; the account here draws on Friedman 2012e.

For Slavic, the doubling of the reflexive pronoun is already attested in Old Church Slavonic using *si*.DAT.WK, e.g., OCS *sebe si* = NTGrk ἐμαυτόν 'myself' (ACC; Supr 343,18), OCS *sebě si* = NTGrk ἑαυτῷ 'for himself' (OCS DAT, Grk DAT; Supr 347,30), OCS *sebě si* = ἑαυτοῖς 'for themselves' (OCS DAT, Grk DAT; Supr 385,23) as well as the OCS *svoemu si* (a dative reflexive possessive) = NTGrk σαυτῷ 'for yourself' (Grk DAT; Supr 336,16). Two points worthy of note in these examples are that (1) the Greek reflexives do not involve (synchronic) doubling, and (2) the dative weak (clitic) *si* is already replacing accusative *sę* (as in the first example: note that *sę* is a strong (full) form; OCS had pronominal clitics only in the dative), with the replacement presumably coming from a perceived quasi-possessive function (cf. nonstandard English *hisself* versus standard *himself*). The form *sebe* is actually a genitive that was already replacing accusatives in many contexts. These examples thus predate other attestations of object doubling in the Medieval Balkans. Such early attested postposed reflexive doubling may well be relevant for the Balkan Romance examples.

For all of Balkan Romance there is an independent pronominal reflexive based on *insu/însu (< ins < impsus < ipsus)* + the reflexive *şi* (< *sibi*), e.g., Aro *el insushi,* Rmn *el însuși* (cf. etymologically related but quite independently developed Frn *lui-même* 'himself,' Sp *el mismo*, < *metipsimo*, Itl *se stesso*; cf. also Frn *soi-même*

'oneself'). Here, too, while Balkan Romance follows a general Romance pattern of building a new form on a referential lexical item, the postposing of the pronominal element is specific to the Balkans. Sandfeld 1930: 189 sees a Greek model here (αὐτός του), but the Greek that he cites lacks the historical reflexive, and he fails to cite older Romanian *sineși* (modern *sine*), DAT *șieși* (modern *sieși*), which resemble BSl (and OCS) *sebe si*. Note also older Romanian constructions of the type *sine-mi, sine-ți,* etc., and also *pe sine insuți* 'yourself' (ACC) (Qvonje 1980: 24).

The combination of personal and reflexive pronoun also occurs in those Bulgarian and Torlak dialects that have lost *sebe*, but in these cases the strong form is the personal pronoun and the short form is the reflexive. In western Bulgarian (e.g., Umlenski 1965) and along the Danube all the way to Silistra (Kočev 1969: 61), as well as in dialects in Thrace (Bojadžiev 1991: 77), including Săčanli, in the Komotini (Blg Gjumjurdžina, Trk Gümülcine) region (Bojadžiev 1972: 116–118), the strong form *sebe* is absent and the possessive clitic *si* is used consistently to refer to the subject, in collocations such as *mene si* (see also §§6.1.1.2.4.3, 6.1.1.2.4.4). The easternmost Torlak dialects of BCMS are like the adjacent Bulgarian ones (Vukadinović 1996: 189), whereas further west, in areas that used to be in contact with Albanian prior to 1878, *sebe* is preserved (Toma 1998: 319–320). The BCMS distinction between a long form dative *sebi* and accusative *sebe* extends as far south as Niš, but not to the nearby villages. In the Torlak dialects of Kosovo, however, e.g., Sirinićka Župa, Sretečka Župa, and Prizren, like Gora, *sebe* is both accusative and dative (Pavlović 1939; Mladenović 1990, 2001). In the southern Montenegrin dialects bordering northern Albania, however, *sebe* is sometimes replaced by a personal pronoun (Pešikan 1965: 155). The relative conservatism of Balkan Romance vis-à-vis the rest of Romance and its use of postposing arguably reflect contact with Slavic.

With regard to the pronouns, a reduplicated form of the type Mac *sebesi*, Blg orthographic *sebe si*, where the accusative *se* exists only as a clitic, can be considered as normal, e.g., Mac *Se vidov sebesi* (cf. Blg *Vidjah sebe si*) 'I saw myself.' The collocation *sebesi* (the constituents are phonologically inseparable – a situation that is already attested in the OCS Codex Suprasliensis) is a specifically Balkan reduplication, since it is not found elsewhere in Slavic.

For Albanian, the reflexive pronoun, *vet*, can occur both plain and doubled, and, moreover, also with personal pronouns, as in *E pashë veten/vetveten* 'I saw myself' (lit., 'it.ACC saw.1SG.AOR self.DEF.ACC/self.self.DEF.ACC') and *Kosova flet (vetë) për vetveten/veten e saj* 'Kosova speaks for itself' (lit., 'Kosovo.F.DEF.NOM speak.3SG.PRS (self) for self.DEF.ACC/self.self.DEF.ACC PC.FEM.DEF.ACC her'). The Albanian situation is conservative in Geg and innovating in Tosk (see §§6.1.1.2.4.3, 6.1.1.2.4.4), and these two developments seem to interact to some extent with local Slavic dialects.

We can also mention in passing Turkish, which in addition to its conjugational means of marking reflexivity by a verbal suffix (-(*I*)*n*-) has an inherited reflexive *kendi*, and expressions such as *kendi kendime* 'by myself,' *kendi kendini* 'oneself' (ACC) are reminiscent of Balkan reduplication. While we have seen that reflexive

doubling precedes contact with Turkish in various Balkan Indo-European languages, at the same time contact with Turkish could have reinforced these tendencies where they are found.

Although both Ancient Greek and Latin had some doubling in reflexive pronouns, the resultant forms were obsolete or unrecognizable by the time Slavic arrived in the Balkans. The development of Balkan Slavic *sebesi* and Romanian *sineşi, sieşi* is connected in both space and time. The position of Albanian *vetvete* is difficult to determine. Although not mentioned in the oldest Albanian grammar (Ismajli 1982), the form is presumably quite old, and, as is the case with the postposed definite article, might be connected with Balkan Slavic and Balkan Romance. For the regions with reflexive doubling, the parallels with Turkish are also suggestive, but given the early Slavic attestations in Suprasliensis, they are probably just parallels, or at most reinforcements of existing phenomena.

7.5.1.7 Object Reduplication and Language Ideology

From the point of view of language ideology, there are a few aspects of object reduplication that are of particular interest. First, de Boel 2008 argues that the construction itself was historically so negatively evaluated in Greek that it was likely kept out of certain literary registers as stylistically too low. Second, there is a curious effect evident in scholarship on the subject, in that various scholars attempt to "prove" that for their particular language, the development has to be "natural" rather than contact-induced, as if contact-induced changes were somehow unnatural (or, ideologically, demeaning, cf. Kazazis 1977). Dimitrova-Vulchanova & Vulchanov 2008, for instance, adduce numerous OCS examples to attempt to demonstrate that Balkan Slavic object doubling is both archaic and independent of Greek, and de Boel 2008, examining Greek, argues for the internal nature of the OR developments in that language. In fact, however, the same ideology that perceives contact-induced change as "unnatural" would also negatively assess any feature perceived to result from it. While such an ideology might well negatively assess any innovation, whether internally or externally motivated, the chronology of the rise of OR in Greek points precisely to periods of contact with other languages in the region. And even for the dialects of western Macedonia, where OR is currently most deeply integrated into the grammar, Ilievski 1973 shows that many centuries later, it was still only weakly attested even in colloquial texts.

Lopašov 1978, citing Orzechowska 1973, observes that in Bulgarian, the pressure from the standard language, influenced by Church Slavonic and Russian, pushed OR down to the colloquial register, and in that way slowed its grammaticalization or at the very least its expansion. A similar ideology of reduplication avoidance was at play with some language planners of Macedonian (S. Risteski 1988: 421–422), where, however, the west-central dialectal basis for the standard vanquished such a restriction. The success in the case of Bulgarian is illustrated by the fact that Leafgren 1992: 287 found no examples of OR in formal expository prose (cf. also Friedman 1994a).

7.5.1.8 Object Reduplication: Conclusion

OR is thus a classically Balkan convergent phenomenon; Romani and Judezmo are on the margins, and the differing degree of grammaticalization in the various languages and dialects reflects the history of its incorporation into different systems under diverse contact situations. The fact that the epicenter of transition from pragmatic to grammatical conditioning is precisely southwestern Macedonia is consistent with a variety of other areal phenomena. Moreover, the ideological anxieties attached to OR serve to attest to the fact that OR both is and is perceived as a Balkanism, i.e., a shared innovation owing to language contact. For additional discussion of OR vis-à-vis the Hungarian definite/indefinite conjugation and other typological considerations see Kallulli 2019.

7.5.2 Weak Object Pronouns in Argument Roles

The Object Reduplication facts discussed in §7.5.1 highlight a construction in which weak object pronouns co-occur with full nominal or pronominal objects (strong forms of the object pronouns). In many instances, the co-occurrence serves a pragmatic function, but some particular collocations of pronoun and nominal in some of the languages serve a purely grammatical function. In either case, however, as well as in cases where the weak object pronoun occurs on its own, without an accompanying full nominal, the weak object pronoun is adding something to the verbal complex by its presence. Such uses raise questions about the role that the weak object pronouns play in signaling argument structure. However, it seems unlikely that language contact is involved here, although the phenomenon is of general linguistic interest. Thus, for example, in instances of verb lability discussed in §6.2.6.2, while full nominal objects are possible, the weak object pronoun can be the only indicator of a transitive reading for a verb.

There is one type of construction that is somewhat like the lability cases but where an overt object nominal is not possible, yet the weak object pronoun occurs. That is, there are idioms in which there is a weak object pronoun, making the verb appear to be transitive, but not necessarily with a transitive interpretation nor real-world referent that the weak object pronoun is linked with. Greek and Albanian offer such examples, given in (7.73):

(7.73) a. πάμε να την πέσουμε στο κρεββάτι (Grk)
go.1PL DMS her.ACC fall.1PL in.the bed
'Let's go for some sleep in the bed' (lit., 'let's.go that her we.fall in.the bed')

b. si ia kalon (Alb)
how it.DAT.it.ACC pass.2SG
'How're you doing? / How's it going?' (lit., 'how for.it.it you.pass,' i.e., 'How do you pass it for.it?')

The object pronoun την in (7.73a) does not refer to a particular female or feminine gender entity. Similarly, the dative *i* of (7.73b) has no apparent direct referent, yet it is there; its function may be more that of indicating an interested party, like the

ethical datives discussed in §6.1.1.2.5, and it is a key part of the idiomatic reading, in that sense, therefore, contributing to the grammatical structure. See also the mention in §7.8.2.2 of nonreferential subjects.

7.5.3 Other Weak Object Pronoun Phenomena

Aside from the so-called ethical dative or dative of interest (see §6.1.1.2.5), there are a few somewhat scattered developments involving weak object pronouns that do not come under the rubric of argument roles.[61] One that is treated elsewhere, since it involves a class of sentences analyzable as verbless, and thus for organizational reasons falls under a different rubric, is the occurrence in presentational sentences of weak object pronouns, generally accusatives but in Greek and Albanian also nominatives; see §7.8.1.2 on this. Also, a related interrogative construction with accusative weak object pronouns is discussed in §7.8.1.3. The other two turn out to most likely not be connected to language contact, but are nonetheless interesting to note as details about the syntax of weak object pronouns in the Balkans.

The first is a striking convergence found in a few of the languages that involves weak object pronouns in a highly particular way as they occur in a minor sentence type. Three languages, Albanian, Greek, and Judezmo, show a synchronically and typologically unusual placement of weak object pronouns in similar constructions. Even though the weak object pronouns in these languages are typically preposed before finite verbs and postposed after nonfinite forms, including imperatives, as discussed in §7.4.1.2.2.2.1 and §7.4.1.2.2.2.2 (and see also §5.5.2 and §5.5.3), in at least some plural imperatives and an imperative-like construction, a weak pronoun object is positioned word-internally, before (i.e., inside of, to the left of) the plural ending; see the data in (7.74)–(7.76):[62]

(7.74) Greek

a. δό’ - μ’ - τι (Thessalian, cf. Tzartzanos 1909, see also Joseph 1989)
give.IMPV me.ACC 2PL
‘(Y’all) give (to) me!’ (lit., ‘give-(to-)me-y’all’)

b. φέρι - μέ - τι (various northern varieties, cf. Thavoris 1977; Ralli 2006)
bring.IMPV me.ACC 2PL
‘(Y’all) bring (to) me!’ (lit., ‘bring-(to)-me-y’all’)

(7.75) Albanian (cf. Newmark et al. 1980; Rasmussen 1985; Joseph 2010b)
hap- e- ni
open.IMPV it.ACC 2PL
‘(Y’all) open it!’ (lit. ‘open-it-y’all’)

61 Farkas & Kazazis 1980 treat the ethical dative as an extra argument of the verb, with the whole action directed at its referent. The role of the interested party, however, is not an argument role in the strict sense as such a party is not an essential participant in verbal scenario; cf. also Steriade 1980.

62 The spacing and hyphens are added here for clarity in analysis and do not reflect standard orthography.

(7.76) Judezmo (cf. Crews 1935: 234)
shte - l - d-e
look.here it 2PL-it
'here it is, you all' (lit., 'look-it-y'all!')

In this last example, *shte* is from Trk *işte* 'look; here; like that' with a native Judezmo embedded object pronoun (*le*) divided by the native 2PL ending *-d*.

Such placement for weak object pronouns is at best rare crosslinguistically, if the pronouns are true clitics, as opposed to being affixes.[63] The rarity is enough to make one consider a contact explanation for it, as is the restriction within Greek to dialects of the north,[64] thus to an area where contact among Greek, Albanian, and Judezmo is plausible.

Nonetheless, despite the striking convergence in form, the facts of (7.74)–(7.76) probably reflect independent origin. For one thing, there are differences in extent that make the comparisons less compelling, in that in some parts of the northern Greek territory, especially Thessaly as described in the earliest account of this phenomenon by Tzartzanos 1909, this interior pronoun pattern is restricted to the verbs δο- 'give' and πε- 'say,' while elsewhere in northern Greece, to judge from Thavoris's examples and discussion, it has a wider distribution; still, Ralli 2006: 147 states it is restricted to monosyllabic imperatival stems, especially βρες 'find,' πες 'say,' δες 'see,' δο(ς) 'give.'[65] For Judezmo, the only example cited by Crews is *shtelde*. Finally, in present-day Albanian, this ordering is standard for all imperatives.

Furthermore, suitable conditions for the creation of such word-internal pronouns were present in each language. For instance, in Greek, forms like δο'μ'τι can be motivated internally as a blend of the singular δο'μ 'give-to.me!' with the non-pronominal plural δό'τι '(y'all) give!,' essentially involving a reanalysis of δο'μ as a stem (i.e., δομ-), and the same can be said for the Albanian, where the evident greater freedom of 2PL *-ni* may also have played a role.[66] Finally, Judezmo *shtelde* may reflect a Spanish-internal metathesis of *-d-* and *-l-*, as it is found in Old Spanish *dalde*, to be parsed as *da-l-d-e* '(you all) give-it-2PL!,' for expected *da-d-le* ('give-2PL-it,' with pronominal *le* split by the metathesized *d*) and it occurs as well in Judezmo (e.g., in Thessaloniki; cf. Symeonidis 2002: 164). Thus, this is more likely

63 It has been argued, by Zwicky 1977 (see also Nevis 1988) that clitics, as syntactic elements, cannot be positioned within a word; such "endoclisis" would be precluded by the Lexical Integrity Hypothesis (on which see, e.g., Bresnan & Mchombo 1995). While a ban on endoclisis is a useful heuristic (see, e.g., Joseph 1988, 1989), Harris 2002, based on contrary evidence from Udi, and Maisak 2021 based on Andi, show that such generalizations are inaccurate. See also footnote 40 in §7.4.1.2.2.2.1.

64 Joseph 1988 cites one example from the standard language, έχε-μού-τε εμπιστοσύνη 'Have-me-2PL faith,' i.e., 'Have faith in me,' but perhaps significantly, it was collected from standard language usage of Thessaloniki, in the north of Greece.

65 It is worth noting, though, that Thavoris's examples include nonmonosyllabic imperative stems, e.g., φέρι! '(you.SG) Bring!'.

66 That is, *-ni* can occur on nonverbal elements, e.g., *mirëdita* 'good day; hello,' thus *mirëditani,* signaling address to more than one person; see §6.2.4.3.4.2 for discussion and references.

merely a superficial parallel that is simply a chance convergence of independent developments in each of the languages involved.

Finally, in some of the languages, weak pronominal forms or at least ones that are phonologically reduced, occur as the objects of prepositions.[67] It is not clear that this is a contact-related phenomenon, since the details are different enough, but their mention here allows for comprehensiveness in the overall treatment of weak object pronouns in the Balkans. Thus in Greek, weak genitive forms are found as the object with the adverbial elements that make up composite prepositions, i.e., the elements that combine with what are generally referred to as the primary prepositions σε 'to, at, in, on,' με 'with,' για 'for,' and από 'from.' When the adverbial elements occur alone, they take weak genitive pronouns as their object, i.e., only personal pronouns, as those are the only pronouns with weak genitive forms. An alternative expression occurs, too, in each case, consisting of a combination of adverbial plus primary preposition with a strong accusative pronoun as the object, strong because weak pronouns do not occur after the primary prepositions. Examples are given in (7.77):

(7.77) a. μαζί μου / μαζί με μένα
together me.GEN.WK together with me.ACC.STR
'together with me'

b. αντί του / αντί για αυτόν
instead him.GEN.WK instead for him.ACC.STR
'instead of him'

c. κοντά σου / κοντά σε σένα
near you.GEN.WK near to you.ACC.STR
'near (to) you'

d. μπροστά μου / μπροστά από μένα
in.front me.GEN.WK in.front from me.ACC.STR
'in front of me'

With the primary prepositions by themselves, only strong accusative pronominal forms are found and weak forms are ungrammatical, e.g., σε μένα 'to me' (cf. *σε με), μ αυτόν 'with him' (cf. *με τον);[68] with nonpronominal objects, the composite form is used and thus (full) accusative objects are found, e.g., αντί για μουσακά 'instead of (lit., 'for') moussaka.' Frasheriote Aromanian in North Macedonia behaves exactly like Macedonian, in that after prepositions only strong form pronouns occur (Markovikj 2007: 90); this is likely to be a (contact-based) innovation, as Romanian allows weak pronominal objects with some prepositions, e.g., *deasupra-mi* 'above-me' or *în juru-i* 'around-him' (Nicolae 2013: 344–345).

67 This wording is intentional as a way of signaling that truly "weak" pronouns are more of a morphosyntactic category than strictly phonologically definable.

68 In some early twentieth-century lyrics in rebetika (a Greek musical genre originating in Asia Minor in the late nineteenth century and compared to the American blues), ostensible weak forms are found with prepositions, e.g., για σε 'for you,' με με 'with me,' but the pronoun actually is audibly stressed, and these phrases are thus [ja sé] / [me mé], suggesting that it is in fact functioning as a strong form, even though segmentally the same as the weak form. Clitics can also be stressed in Macedonian for rhetorical purposes, e.g., in a political debate *gó zadušúvaat grádot* (lit., 'it they.strangle the.city,' referring to pollution in Tetovo), where the stressed proclitic places extra emphasis on the statement as a whole (Bilal Kasami, October 21, 2021, in a televised debate on www.24.mk).

As noted, in other Balkan languages, there are seemingly parallel uses of weakened, i.e., phonologically reduced, pronominal forms with prepositions. In Albanian, the third person singular accusative pronoun *atë* takes the form *të* with certain prepositions, as in *me të* 'with him/her.' The same can occur with the ablative, e.g., *larg sish* 'far from them (M).' However, this phenomenon is restricted to just the third person. In Bulgarian, but not in Macedonian, the first and second person singular accusative pronouns have shortened strong forms, viz. *men* and *teb*. These forms occur with prepositions, e.g., *săs men* 'with me,' *za teb* 'for you' and also occur independently, as in (7.78):

(7.78) a. Men sigurno tam šte me turjat
me.ACC certainly there FUT me.WK.ACC put.3PL
'They will certainly put me there'

b. To men, ako pitaš, mi dumaha za tebe
that me.ACC if ask.2SG me.DAT said.3PL about you
'They said that to me, if you ask, about you'

c. Men me bjaha haresali da me upotrebjat
me.ACC me.WK.ACC were.3PL pleased.LPT DMS me.WK.ACC use.3PL
'They had pleased me [in order] to use me'

While it is not clear what is driving (or did drive) the reduction in these Bulgarian syntagms, it is clear that they are not genuine weak forms. This, along with the absence of anything parallel in Macedonian, makes it appear to be a Bulgarian-internal development. Moreover, the differences between this Bulgarian phenomenon, being only in the first and second person singular, and the Albanian one, found only in the third person, point to both being independent of one another, and, moreover, they are independent of the Greek facts described above, inasmuch as in Greek one is dealing with the use of an already-existing weak form, and one in a different case at that. Thus despite the apparent similarity when these phenomena are described in superficial terms – weak(ened) pronoun objects occurring with prepositions – the differences are such that they are unlikely to have anything to do with one another as far as influence of one language over another is concerned. However, Bulgarian also permits genuine weak forms after the prepositions *vărhu* 'above, etc.,' *meždu* 'between,' *pomeždu* 'among,' and *okolo* 'around, etc.,' and, moreover, the weak forms are dative, e.g., *vărhu mi = vărhu mene* 'above me, over me,' *pomeždu ni = pomeždu nas* 'between/among us,' *okolo mu = okolo nego* 'around him, in the vicinity of it' (Stojanov 1983: 41). Here the comparison with Greek is unmistakable. Thus, while most of the comparisons seem to have language-internal explanations, the comparison of Bulgarian with Greek in this regard is highly suggestive, given that the dative is never used after prepositions in Balkan Slavic except in these instances. This point emphasizes the need to distinguish *surface* from *superficial.* Superficial similarities (varying degrees of various weak object pronouns after various prepositions) cease to be similar when examined in detail, whereas surface similarities (Bulgarian and Greek weak dative pronouns with certain adverbial prepositions) point to the possibility of contact-induced change.

7.6 Negation[69]

Negation in the Balkans shows some characteristics that are of interest both to comparative Balkan syntax and the comparative syntax of the Balkans, in the senses discussed in §7.1. The most relevant facts pertain to indicative negation, to *modal negation* (as a shorthand for *marked modal negation*) given that indicative is the unmarked mood, and to some particular properties of the morphemes involved in these two types of negation.

The Indo-European historical background to these Balkan facts is, as always, an important prelude to a consideration of the Balkan situation. In particular, several branches of Indo-European, specifically Albanian, Anatolian, Armenian, Greek, and Indo-Iranian, show a formal distinction between the markers for indicative and for modal negation, and this distribution gives a basis for reconstructing such a distinction for Proto-Indo-European. Moreover, the evidence of Albanian, with a nucleus of *n-* in the indicative negator (e.g., StAlb *nuk*)[70] and *m-* in the modal negator (*mos*), and of Indo-Iranian, with the same *n-* in indicative negation (e.g., Skt, Aves, Rmi *na*) and *m-* in modal negation (e.g., Skt, Aves *mā*, Rmi *ma*), allows for a reconstruction of the relevant forms as, respectively, **ne* and **meH*$_1$. It follows from this reconstruction that in the other branches with the indicative/modal distinction in negation, there have been innovations as to the specific markers; in the Balkans, Greek, Albanian, and Romani are such languages, with a reflex of the modal **m*-negator (AGrk μή, ModGrk μη(ν), Alb *mos,* Rmi *ma*), and, for Greek, an innovated indicative negator (AGrk οὐ – also οὐκ, οὐκί, and a few other related forms[71] – and ModGrk δεν[72]). Other branches show generalization of one or the other negator such that Slavic and Romance generalize **n*-forms while Tocharian generalized **m-* forms.[73] The distinction between (reflexes of or substitutions for) **ne* and **meH*$_1$ is thus a lexical one, but it is syntactically relevant inasmuch as it is sensitive to mood; thus it belongs here as a matter of selection in word combinatorics.

7.6.1 Indicative Negation

The Balkan languages all show the use of a derivative of the inherited indicative negator with **n* (Alb *nuk*, BSl *ne*, BRo *nu*, Rmi *na*) or an innovative substitute for

69 See also §6.2.4.1ff. on negation and futurity.

70 Also *s'*, from **ne ... k*^{w}*id*, with loss of **ne* (Hamp 1984: 178; see also Hamp 1974), as with *as-* (see footnotes 71, 89); *nuk* might be a loan from Latin *nunquam* 'never,' but Hackstein 2020, takes it to be an inherited item, cognate with OHG *noh* 'and not.'

71 οὐ / οὐκ / οὐκί ultimately derive from **ne H*$_2$*oyu k*^{w}*id* (= **ne* 'not' + **H*$_2$*oyu* '(long) life' + **k*^{w}*id* 'what(ever)'), lit., 'not ever at all' / 'not on your life,' with loss of **ne* in a bipartite negation schema just as in French, where *ne ... pas* gives simply *pas* colloquially. Related to this is also Armenian *oč'*; see Cowgill 1960. The Albanian negative prefix *as-*, found in *as-gjë* 'nothing,' *as-kush* 'no one,' etc., belongs here too – see Joseph 2005, 2022a – and some argue that the free word *as* 'and not' does as well; see Hackstein 2020 for further discussion, also Joseph 2005, 2022a for a contrary view.

72 From AGrk οὐδέν 'not at all' (from οὐ 'not' + δέ 'but' + ἕν 'one,' thus lit., 'not even one'), with regular loss of the initial unaccented vowel.

73 See §7.8.3.2 for exceptions to this statement that have resulted from subsequent borrowing.

it, e.g., Alb *s'* (see footnote 70) or ModGrk δεν (see footnote 72) to negate indicative verbs. For present tense indicatives and past tense indicatives, such usage may seem obvious, and in any case is most likely an inherited feature from Proto-Indo-European. However, there is one less-than-obvious development that the Balkan languages share, namely the use of indicative negation for the innovated future tense, even though those futures are built on a modal form in that the prototype for the Balkan future consists of a form of WANT or HAVE with a DMS and an inflected verb (see §6.2.4.1 for details in each language). In principle, negation of either WANT/HAVE or the inflected verb should be possible in that schema, with the meaning being essentially the same in either case (much as *John seems not to be smiling* and *John doesn't seem to be smiling* are synonymous in English, a phenomenon known as "NEG-Raising"; see also footnote 44). It is therefore significant that in all the Balkan futures, the negation marker is consistently associated with the WANT/HAVE element and not the inflected verb.

In the languages that distinguish between indicative and modal negators, e.g., Albanian, with a future *do të shkruaj* 'I will write,' Greek, with θέλει να γράφω (> θα γράφω) 'idem,' and Romani with *ka džav* (< *kamav te džav*) 'I will go,' the negation of the WANT element would involve the indicative negator (*nuk do të shkruaj* / δεν θ(έλει ν)α γράφω 'I won't write'/ *na ka džav ~ nae man te džav* 'I won't go.' For the negation of the inflected verb under a DMS one might expect the modal negator (*do të mos shkruaj / *θ(έλει ν)α μη γράφω / *ka ma te džav). With limited exceptions – especially the Greek dialects with θα μη discussed in §7.4.1.2.2.2.2 (and see footnote 45) – all one finds, however, is the indicative option (i.e., Alb *do të mos, Grk *θα μη, Rmi *ka ma te are all ungrammatical).

For Balkan Romance, without an indicative/modal negation distinction, the potential for differential realization of negation in the future would be a matter solely of the positioning of the negative marker (e.g., *nu o să şcriu* versus *o să nu şcriu) and it is significant that only the leftmost negation option, the equivalent of negating not the modal element but the indicative element, is all that is found. As for Balkan Slavic, it is somewhat of a special case, since for most varieties, HAVE is deployed in the negated future, but there is also a DMS, so that in principle, as in Balkan Romance, there is a potential difference in the position of the negator, either under (after) the DMS or associated with the HAVE element (e.g., Macedonian *nema da pišam* versus a hypothetical *ima da ne pišam*, which is possible but has a different meaning); again, only the leftmost positioning occurs, the equivalent of indicative negation.[74]

Given that other aspects of the composition of the future tense in the Balkans are due to language contact, as discussed in §6.2.4, this syntactic convergence in the way in which negation is realized in the future is likely to be a matter of contact as well, with the calquing responsible for the spread of the WANT future in the first place being responsible for this further detail of formation as well.

74 The collocation *ima da ne* ('has DMS NEG') is possible, e.g., *ima da ne mrdneš od tuka* 'you must not budge from here' but it has a necessitative rather than a simple future meaning.

7.6.2 Modal Negation and Prohibitions

As generally reconstructed, the modal negator in Proto-Indo-European had as its primary function the expression of prohibitions, i.e., negative commands, as all of the languages with *meH_1* have at least this function. We say "at least" since, as becomes clear below (§7.6.4 and cross-references there), other functions are also found for reflexes of *meH_1* in the Balkans, and across Indo-European. Not all, however, are of Proto-Indo-European age; some reflect innovations that occurred after the break-up of the proto-language.

Albanian, Greek, and Romani all use their respective outcomes of *meH_1* (Alb *mos*, ModGrk μη(ν), Rmi *ma*) in prohibitives, and their agreement on this point is thus a matter of inheritance from Proto-Indo-European. They differ, however, in the verbal forms that occur with *meH_1* in their prohibitives.[75] Albanian and Romani use the imperative, e.g., Alb *mos shkruaj!* 'Don't write! (SG),' *mos puno!* 'Don't work! (SG),' Rmi *ma mudar amen!* 'Don't kill us!' Both languages also permit the DMS + subjunctive, but in Romani *te* must follow *ma* (after *te* the negator is *na*, most likely a contact-based innovation with either Slavic or Balkan Romance which use their *n*-negator after the DMS), e.g., *ma [te] ave/te na ave* 'don't come (SG),' while Albanian permits both orders with *mos* (see §7.4.1.2.2.2.2), and indicative *nuk* cannot co-occur with *të*. Albanian also permits the optative in any person after *mos*, e.g., *mos vdeksh kurrë* 'don't ever die!.' In Greek, however, the imperative is impossible, e.g., *μη γράψε! 'Don't write! (SG),' *μη γράψτε! 'Don't write! (PL).' This ban on the imperative in Greek prohibitions is a deviation from what is found in Ancient Greek, where imperatives as well as subjunctives could be used. In Modern Greek, what is used in prohibitions is the present imperfective (for continuous action) or the present perfective (for punctual action), a form that is subordinate in a certain sense in that it cannot stand on its own (see §6.2.2.3 and §7.7.2.1.3.2.2), but neither is identifiable in any obvious way with a category of subjunctive in the modern language. The imperfective form in question is identical with the simple present tense form, which can stand alone and determine a sentence by itself even (e.g., Γράφω. 'I am writing.').[76] However, in the perfective aspect, it stands apart from other forms but cannot occur in a clause on its own or determine a sentence by itself (i.e., *Γράψω. 'I (do) write.'). It is subordinate in a certain sense in that it cannot stand alone and always requires a particle or verbal marker or subordinating element to support it (see §6.2.2.3 and §7.7.2.1.3.2.2); μη(ν) is such a supporting marker. Some examples of acceptable prohibitions in Greek are given in (7.79):

75 The verbal mood used in Proto-Indo-European for prohibitions is difficult to reconstruct with any real certainty. For the most part, the languages disagree on which form is used, with some using (or allowing) imperative, some subjunctive, and some (really only Vedic Sanskrit) a special form known as the "injunctive" (formally with past tense endings but without the temporal prefix, the "augment," that characterizes Vedic indicative past tenses).

76 By "simple present tense form" we mean the indicative form that can occur unaccompanied by any particles or supporting elements in main clauses, something only possible in present tense forms for imperfective aspect; see §7.7.2.1.3.2.2 for a discussion of how aspect intersects with decisions about mood categories in Greek, and §6.2.4 on mood in general in the Balkans.

(7.79) a. Μη γράφεις έτσι! cf. Γράφεις 'You are writing; you (do) write'
NEG write.2SG.PRS.IPFV thus
'Don't (you) keep writing like that!'
b. Μη γράψεις έτσι! (cf. *Γράψεις 'you (do) write'
NEG write.2SG.PRS.PFV thus
'Don't (you) write like that!'

Deciding what verbal mood was used with *meH_1 in Proto-Indo-European is not just an exercise in reconstruction, as it has relevance for developments in the Balkans. It is hard to say what the original mood in Albanian prohibitives was – imperative or subjunctive or perhaps a now-lost reflex of the injunctive (see footnote 75) – but imperative certainly has been a possibility for at least recorded Albanian history. As for Greek, there definitely was a change within the history of the language from a state with both imperative and subjunctive moods as possibilities to a state with the occurrence only of a non-imperative. For Romani, Sanskrit offers a view of the earlier state of affairs, in that *mā* occurred with the injunctive in Vedic Sanskrit, but in Classical Sanskrit, the injunctive, while possible with *mā*, was receding in favor of the imperative. Interestingly, then, in a construction where Albanian, Greek, and Romani agree in certain respects, and utilize cognate material (*mos*/μη(ν)/*ma*) that shares other functions (see §7.6.3), this presumably internally motivated change, simplifying the options in use, has driven Modern Greek away from an earlier parallelism it had with Albanian and Romani in the syntax of this construction as to the possibility of using imperatival forms, although subsequent Greek influence on Romani may be present (see §7.6.3.3).

Two matters pertaining to prohibitions that are more lexical in nature are discussed in Chapter 4 but bear repeating here, as some elaboration, especially on the syntactic side in one instance, is possible. First, as pointed out in §4.3.3.3, Greek μη(ν) has been borrowed into southern Aromanian and adjacent Balkan Slavic as *mi*; thus secondarily via contact, the range of occurrence of reflexes of *meH_1 in the various branches of Indo-European has been extended. Second, there is the use of the modal negator as an independent one-word prohibitive utterance, as mentioned in §4.3.4.3.2, whereby the **m*-negator in Albanian, Greek, and Romani occurs by itself with a prohibitive meaning; this use is shown in (7.80abc) respectively:

(7.80) a. Mos! 'Don't!'
b. Μη! 'Don't!'
c. Ma! 'Don't!'

Independent *mos*/μη (notably, never with the final -ν)/*ma* is like English *Don't!*, except that there is no verb, just the negator. It is interesting from the Balkan perspective for two reasons. First, there are no apparent Ancient Greek instances of an independent usage for μή (the ancestor of modern μη(ν)) expressing negative actions in a prohibitive way, thus nothing directly comparable to the Albanian or

Modern Greek usage.[77] Similarly, there is no use in Sanskrit for its modal negator *mā* that compares directly with the independent prohibitive use of Romani *ma* and it appears that those other Indic languages that preserve reflexes of *mā* do not use it in this way; *mā* survives, sometimes in an extended form, according to Masica 1991: 389, in Hindi and Punjabi *mat*, Sindhi *ma*, Kashmiri *mā*, *maṭi*, Gujarati *mā*, i.e., mainly "in the center and west; in Eastern and Southern languages other devices have evolved." In Gujarati, the particle is postposed to the verb, but elsewhere it remains preposed.

Thus both Modern Greek and Romani have innovated, either on their own or through contact with one another or with another language, in such a way as to make the independent use of their *m*-negator as a one-word prohibitive utterance possible. Since so little is known about the prehistory of the uses of Albanian *mos*, it is possible to suppose that Albanian is the source of this innovative usage in the Balkans, but there is reason to believe that Balkan Romance, especially Aromanian, may hold the key to how this usage may have arisen in Albanian, Greek, and Romani. That is, as suggested speculatively in Joseph 2002c, the fact that Balkan Romance uses its sole negator *nu* both in prohibitions and as the equivalent of an independent prohibitive means that it provides a model, via a cross-language analogy (i.e., a calque) for the usage observed in Albanian, Greek, and Romani, e.g., with the verb 'do':

(7.81)		BRo	**nu** face!	:	**nu**!	(= NEG 'do!' : NEG!)
	::	Grk	**μη** κάνεις	:	X!	(X => **μη**)
	::	Alb	**mos** bëj!	:	X!	(X => **mos**)
	::	Rmi	**ma** ker!	:	X!	(X => **ma**)

Balkan Slavic has the innovative negative imperative particles *nemoj* (Macedonian and BCMS) and *nedej* (Bulgarian), with dialectal variants, in these functions. In most dialects, these prohibitives must be used either alone or followed by a 2nd person DMS clause. They thus differ from all the *m*- languages in that those languages all have the possibility of the modal negator being used with a DMS clause in any person (the DMS normally – or obligatorily, for some speakers/dialects – precedes Greek μη(ν) and follows for Romani while Albanian *mos* has flexible order, see §7.4.1.2.2.2.2 above), whereas, with the exception of a specific BCMS usage, the Balkan Slavic prohibitive, aside from the negated imperative, must be DMS + NEG + 2nd person verb – which replaces an infinitive in historical terms – or the abovementioned particles + DMS + 2nd person verb. The particle

77 Willmott 2013, in a detailed presentation of the history of negation in Greek, calls the independent use of μή in Ancient Greek rare, but offers some examples of its use from Aristophanes' *Peace* (l. 455–456, 927). Importantly for the point being made here about the functional parallelism of Alb/ModGrk/Rmi *mos*/μη/*ma*, her examples are not prohibitive in their function; they do express emphatic negation, but there is nothing prohibitive in them (see, e.g., the translation by Eugene O'Neill Jr. (http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0038). Hackstein 2020: 26 suggests that this use of *meH$_1$, which he calls a "sentential" usage, might be an inheritance from PIE, since a (somewhat) parallel usage is found with the Tocharian A negator *mar*, which continues, in part, PIE **meH$_1$*. If so, that could mean that Albanian is the locus of diffusion of this usage within the Balkans; still, see also footnote 248 below.

nemoj is derived from *ne* plus the imperative *mozi* 'can, be able,' a usage already attested in OCS:[78]

(7.82) Ne mozi mene ostaviti (Supr, cited in BER IV:610)
NEG can.IMPV me.ACC leave.INF
'Don't leave me!'

Nemoj (with variants) occurs throughout BCMS, Macedonian, and most of Bulgarian, excluding the northeastern dialects that serve as the basis of the standard. Those dialects and the standard use *nedej*, from *ne* + the imperative of *de-* 'do.'[79] These particles can, when functioning alone, take the 2PL marker *-te*, and, in BCMS, also the 1PL marker *-mo*, as appropriate. They can thus also be seen as verbs that occur mainly in the imperative (*-mo* is a jussive). When used alone they mean 'Don't!' or 'Let's don't' (for BCMS *nemojmo*). As can be seen from example (7.82) above, the original construction was used with an infinitive, which is still the case in BCMS. The short infinitive is also a colloquial or older-generation possibility in standard Bulgarian, e.g., *nedej pisa* 'don't write!,' *nedej zaminava sega* 'don't leave now!' (Nicolova 2008). Nicolova notes that the short infinitive is unknown to the younger generation, and some will use an imperative with *nedej,* e.g., *nedej piši* instead of *nedej pisa*. Thus, *nemoj, nedej* (and related forms, such as *nim*) can both introduce prohibitions, as seen in §7.7.2.1.1.1.2, and serve as one-word prohibitive utterances, as in these examples from R. Greenberg 1996b:

(7.83) *Nemoj*, ne pipaj 'Don't! Don't touch!' (Mac)
Nim bre, Argíre 'Don't, hey, Argir!' (Lower Vardar Mac)

As noted in §7.7.2.1.1.1.2.1, the Macedonian of the Ser/Lagadina (Grk Sérres/Langadas) area in northern Greece shows the prohibitive type *nemoj sediš* with 2nd person nonimperative verb occurring with *nemoj*; since these dialects are in territory where the dominant language is Greek, where inflected second-person forms occur in prohibitions, it is possible that Greek influence is connected to the omission of *da* in this *nemoj* construction, although other explanations (e.g., reinterpretation of a short infinitive as a finite form with the addition of a person marker) are also possible.

There are also lexical borrowings and calques that figure in prohibitions. Balkan Slavic and Meglenoromanian have borrowed the Turkish exclamation *sakın* (Blg *sakăn,* Mac *sakan,* Megl *săcăn*) for 'Don't [do it]!' or 'No way!'[80] And, in northern Greek, Balkan Slavic, and Aromanian, influenced by Turkish, there is shared phraseology involving a predicate meaning 'enough; stop' that has a prohibitive value; see §4.3.3.3 for details.

78 Dialectal Russian has *ne mogi* in this usage (BER IV: 610).
79 Cf. OCS *děti*; the root is still used in Bulgarian word formation, but the verb is obsolete. The prohibitive *ne ději(te)* occurs in many OCS manuscripts (Sadnik & Aitzetmüller 1955: s.v.).
80 Etymologically, *sakın* is an imperative meaning 'Take care!,' but it functions as a particle.

7.6.3 Extensions of Modal Negation

Albanian and Greek, as languages that, like Romani, have a special morpheme for modal negation, show other uses for their respective morphemes, some of which are not even fully negative in nature. These functions for Greek are discussed in Joseph & Janda 1999 (see also Veloudis 1982), and the parallels with Albanian are treated in Joseph 2002c. They include nonfinite negation, tentative questions, and in complements to a matrix predicate meaning 'fear.' Romani, despite having a special morpheme, has only some of these extensions. While all of these have an innovative aspect to them, all seem to represent relatively natural extensions of the basic negation function for *meH_1. Indeed, Chatzopoulou 2019 argues that nonveridicality is a way of linking several of the extended uses, especially the question use and the 'fear' complement use. Thus, they could in principle simply reflect independent innovations in each language, although such an account, in the absence of a clear indication of independence for the developments leading to them, is essentially an untestable, and thus uninteresting, hypothesis. Still, the naturalness and possible independence notwithstanding, it is interesting that all three languages show some parallels in these extended functions of their respective forms *mos,* μη/μην, and *ma* as well as in their apparently more basic functions presented in §7.6.2. The parallels discussed here are striking even if, as becomes clear in the sections that follow, the extent to which they represent contact-induced convergences is difficult to determine. At the very least, then, even if not due to contact, they contribute to the superficial syntactic parallelism evident among the Balkan languages.

7.6.3.1 Negation of Nonfinite Forms

One extended use is purely negative in nature, but negates along a different parameter from the original modal, prohibitive negation. That is, in both Albanian and Greek – but not in Romani – the modal negator is used in the negation of nonfinite forms. In Albanian, this means negating the participially based formations, such as the infinitive and the continuative, as in (7.84), where example (c) is from a song by the Kosovar singer Meda (Mehedin Përgjeqaj):[81]

(7.84) a. për të mos punuar
INF DMS MNEG work.PTCP
'(in order) not to work'

b. duke mos punuar
GRDM MNEG work.PTCP
'while not working'

c. A me të thirr a mos me të thirr (Geg)
Q INFM you.ACC call.PTCP Q MNEG INFM you.ACC call.PTCP
'Should I call you or not call you?'

81 The song's title is the line in example (c), which serves as a refrain throughout; the song can be accessed at https://urldefense.com/v3/__https://www.youtube.com/watch?v=UqZzCDQujH0&list=PLWs12Wo_LK96W0DcaCloELZHia2121DFP__;!!KGKeukY!lGprpnCbatBG9c_-TqM5_PM9y3IHm89oktrJcEMGCapFdQzDb3B5bPoUv1l-NRbE$.

d. me qenë apo mos me qenë (Geg)
INFM be.PTCP or MNEG INFM be.PTCP
'To be, or not to be … '

In Greek, μη(ν) negates the active and the mediopassive participles, as in (7.85):

(7.85) a. Μην έχοντας ιδέα, έφυγε
MNEG have.GRD idea left.3SG
'Not having any idea, he left'

b. Μη δεχόμενος ούτε ένα δώρο έσκασε το παιδί στα κλάματα
MNEG receive.MP.PTCP even one gift broke.out.3SG the child in.the cries
'Not receiving any gifts, the child burst out in tears'

This usage is an innovation in Greek. In Ancient Greek, μή could negate infinitives and participles, but so too could the indicative negator οὐ, with the choice depending on the value of the infinitive or participle: modal or factual. By contrast, the Modern Greek indicative negator, δεν, cannot be used with either of the nonfinite forms in (7.85), i.e., *δεν έχοντας / *δε δεχόμενος.[82] Thus. the job of participial negation is done by μη(ν).[83] It is tempting to see the innovative restriction on participial negation in Greek – μη(ν) only – as somehow connected with the use of *mos* with participles in Albanian, though the chronology of this restriction has not been determined nor is it clear whether this use of *mos* in Albanian extends back into Albanian prehistory. Romani does not have any similar uses of *ma*, e.g., *ma phirindor 'not walking.'

In Balkan Romance, there is a division of labor in negation that is reminiscent of the Albanian/Greek situation with, in essence, a special negation schema for participles that differs from indicative negation. That is, in Romanian (Pană Dindelegan 2013: 207; Manea 2013: 559), the general negator *nu* negates all finite forms and the infinitive, while the prefix *ne-* is used to negate the non-infinitival nonfinite forms, i.e., the gerund, the supine, and the participle, e.g., *necontrazicându-l* 'not.contradicting-him,' *de neauzit* 'not to.be.heard.of,' *neimaginat* 'unimaginable' (lit., 'unimagined'). This pattern of functionally distinct negators in Romanian is different from what is seen elsewhere in Romance (cf. Spanish *no incluyendolo* 'not including it,' with the general negator *no*, or French *ne le voyant pas* 'not seeing him,' with the general negation schema *ne ... pas*), so that it reasonably represents a Balkan Romance innovation. That it is somewhat parallel with the distribution of distinct negative markers seen in Albanian and in Greek makes for intriguing speculation; while a contact account cannot be ruled out, this convergence may simply speak to the naturalness of such a division of labor within the expression of negation.

82 See §7.7.2.2.2 on the decline in the number and use of participles in Greek.

83 It is not the case that μη(ν) negates all nonfinite forms in Modern Greek if the imperative is accepted as nonfinite, as argued by Joseph 1978/1990, 1983a; rather, as noted in §7.6.2, for the negative of an imperative, i.e., a prohibition, μη(ν) is used with a finite form rather than the imperative. This fact also means that μη(ν) cannot be said to negate only nonfinite forms. The proper generalization thus seems to be that the indicative negator δεν can be used only with finite forms and not with nonfinite forms.

7.6.3.2 Tentative Questions

Another extension of *meH_1 forms is their use in introducing tentative questions, discussed more fully, with examples, in §7.8.3.1 (and see also §7.4.1.2.2.2.2). These are questions that seek not just a yes or no answer but some degree of confirmation of an expected 'yes' or 'no' answer, depending on context. Such questions do not involve negation per se, but they seem to start from a presupposition of the speaker not being completely positive about the response; the speaker's doubt can thus be seen as a form of weak negation. This usage is found in Ancient Greek, as in (7.86):

(7.86) μή σοι δοκούμεν τῆδε λειφθῆναι μάχῃ (A. *Pers*. 344)
Q you.DAT seem.1PL this.DAT leave.PASS.INF battle.DAT
'Do you think that we were simply outnumbered in this contest?'
(lit., 'We seem to you to have been left behind in this battle, no?')

This usage has thus been a part of Greek for millennia. It may have been an innovation within early Greek, as this particular usage of *meH_1 is not found in other ancient Indo-European languages. An interrogative usage is found for Classical Sanskrit *mā,* but it does not seem to have a dubitative, tentative quality. Its occurrence in Albanian and Romani means further that this phenomenon, if not simply involving two independent parallel innovations, could represent the influence of Greek on Albanian and Romani at some point.[84] The calquing could also extend to Balkan Slavic, Balkan Romance, and Romani where MNEG + DMS generally renders the Grk MNEG, while in Romani it is an alternative for MNEG (or MNEG + DMS). It is generally held that there are no strong indications of shared innovations linking the ancestors of Greek and Albanian as a subgroup within Indo-European, so that treating this parallel as an innovation those two languages underwent together would go against this general Indo-European-internal dialectological picture.[85] Thus, a contact-related explanation may well be a better possibility in this case, especially since these markers are conversational elements and thus fit in with the pattern of ERIC loans discussed in §4.3.4. In that case, with the documented chronology of the usage in Ancient Greek, Greek could be the source language for Albanian and the other languages via partial calquing, with Greek usage as the model, but note that Modern Greek μη(ν) does not occur with the DMS in this function, whereas the DMS is present in Albanian, as well as the other Balkan languages. Romani can use either DMS + indicative negator or plain modal negator or even MNEG + DMS here. For Albanian and Romani, calquing would have led to a semantic shift in *mos/ma*, extending its range of functions to match those of Greek μη.[86]

84 Such usage is not mentioned by Masica 1991: 389ff. for the rest of Indic.

85 Though see Joseph 2013b for an exploration of some tantalizing features shared between Greek and Albanian, some of which may well be innovations, contrary to the more usual consideration of the two as totally independent branches of Indo-European. This line of reasoning is developed more fully, with a large number of additional candidates for shared innovations, in Hyllested & Joseph 2022.

86 See §4.3.10 on such contact-induced semantic shifts, referred to there as "isosemy," with numerous examples.

Somewhat parallel to this interrogative usage is that seen in Macedonian with *da ne*, and Albanian with *të mos*, in both instances thus a DMS + MNEG structure, as well as Romani *te na* and *ma*, and *ma te* in (7.87):

(7.87) a. Da ne si nešto bolen? (Mac)
DMS NEG are.2SG somewhat sick

b. Të mos jesh i sëmurë? (Alb)
DMS MNEG are.2SG PC sick

c. Te na sijan nasvalo? (Rmi, Skopje Arli)
DMS NEG are.2SG sick

d. ma sijan nasvalo (Rmi, Skopje Arli)
MNEG are.2SG sick
'You aren't sick, are you? / Aren't you feeling well?'

e. ma te si nekoj mulo? (Rmi, Boretzky & Cech 2016: 45)
MNEG DMS is.3DG someone dead.M
'Someone hasn't died, have they?'

Aromanian has the DMS *s'* + *nu* in this same meaning and context (cf. Bara et al. 2005: 192).

7.6.3.3 Modal Negators in 'Fear' Clauses

A further extended use of *mos,* μη(ν), and *ma* is one that is loosely associated with negation without involving negation proper. That is, the modal negator can occur in the complement of verbs or nouns of fearing, without imparting a negative meaning to the complement verb. This use can be said to be somewhat associated with negation due to the negative connotations connected with expressions of fear, though the actual occurrence of *mos* or μη(ν) amounts to a pleonastic usage as they do not add logical negation to the sentence, but only reinforce the negativity of the 'fear' expression. As is evident from (7.88abc) below, Albanian, Greek, and Romani actually have quite similar syntax in this construction, with a subordinating element, the indicative complementizers *se* and *kaj* in the case of Albanian and Romani, respectively, the DMS να and *te* in the cases of Greek and Romani, respectively, followed by the modal negator morpheme, and then the complement verb. It is worth noting that the use of Albanian *mos* after the complementizer *se* is quite at odds with the usual negation with *se*, inasmuch as *nuk* or *s'* would be expected, and conversely, the DMS *të* would be expected as the subordinator cooccurring with *mos*, not *se*. For Romani, both possibilities exist, i.e., both the DMS *te* and the indicative COMP *kaj* can be used with MNEG *ma*. Moreover, if the complement verb is to be negated (e.g., 'I fear that he might not come'), the indicative negator is used in all three languages, as shown in (7.88def):[87]

(7.88) a. Φοβάμαι να μην έρθει (Grk)
fear.1SG DMS MNEG come.3SG
'I fear that he might come'

87 For Romani, DMS + MNEG (*te na*) and COMP + NEG (*kaj na*) are also possible.

b. Kam frikë se mos vjen (Alb)[88]
have.1SG fear that MNEG come.3SG
'I have a fear that he might come'

c. Darava te/kaj ma avel (Rmi)
fear.1SG DMS/COMP MNEG come.3SG
'I have a fear that he might come'

d. Φοβάμαι να μη δεν έρθει (Grk)
fear.1SG DMS MNEG NEG come.3SG
'I fear that he might not come'

e. Kam frikë se mos nuk vjen (Alb)
have.1SG fear that MNEG NEG come.3SG
'I have a fear that he might not come'

f. Darava te/kaj ma na avel (Rmi)
fear.1SG DMS/COMP MNEG NEG come.3SG
'I fear that he might not come'

All three languages can also use *mos*/μη/*ma* by itself as the subordinating element, presumably a complementizer, so that (7.89acd) are acceptable; the negation used is still the indicative negation, as in (7.89be):

(7.89) a. Φοβάμαι μην έρθει (Grk)
fear.1SG MNEG come.3SG
'I fear that he might come'

b. Φοβάμαι μη δεν έρθει (Grk)
fear.1SG MNEG NEG come.3SG
'I fear that he might not come'

c. Kam frikë mos kam infektuar të tjerë (Alb)
have.1SG fear MNEG have infected PC.ACC others
'I fear that I might have infected others'

d. Darava ma avel (Rmi)
fear.1SG MNEG come.3SG
'I fear that he might come'

e. Darava ma na avel (Rmi)
fear.1SG MNEG NEG come.3SG
'I fear that he might not come'

This complementizer use of μή is found in Ancient Greek (Emde Boas et al. 2019: §43), so the Modern Greek usage in (7.88–7.89) continues that use (and see Tzitzlis 2000: 267 for examples from Medieval Greek). Moreover, given that the Latin modal negator *nē* introduces such clauses (Hale & Buck 1966: §502), and Armenian *mi* is involved in such complements (Joseph 2002c), this use may well be inherited from Proto-Indo-European. As for Romani, the use of *ma* here appears to be an innovation, to judge from the evidence of modern Indic languages of India,

88 Source: <https://www.youtube.com/watch?v=84h9IgROu7E>. Some Albanian speakers, however, feel that the complementizer is required. A Google search did not yield the collocation *kam frikë mos nuk* + VERB, but it is reasonable to speculate that speakers that use *mos* as a complementizer could produce such constructions.

such as Gujarati, Hindi, and Marathi, where a parallel *m*-form is not used in sentences similar to (7.88cf) and (7.89de).

Macedonian offers a similar sort of distinction, though it is expressed with different elements from the languages with the *m*-negator. That is, as shown by the examples in (7.90), the indicative complementizer *deka* can introduce a negated complement of *se plašam* 'I fear,' whereas the occurrence of the DMS *da* with the ostensible negative *ne* shows a pleonasm in that that combination does not give a negated complement from a semantic standpoint; thus (7.90a) parallels (7.88def) and (7.89be), and (7.90b) parallels (7.88abc) and (7.89acd) (cf. also Mitkovska et al. 2017):

(7.90) a. Se plašam deka ne doagja
REFL fear.1SG COMP NEG come.3SG
'I'm afraid he is not coming'

b. Se plašam da ne doagja
REFL fear.1SG DMS NEG come.3SG
'I'm afraid lest he come' (i.e., 'I'm afraid that he might come')

As for the somewhat pleonastic use of *mos*/μη(ν)/*ma*, it is found with μή in Ancient Greek (cf. LSJ s.v. μή, §B.5) so one may conclude that the modern usage is in some way a continuation of the earlier usage. However, given the possible association of **meH*$_1$ with the negativity of fearing as far as complementation with a 'fear' expression in Proto-Indo-European was concerned, the extension of **meH*$_1$ to use as a pleonastic negator is a natural enough development. It is not at all clear, however, how this usage squares with the prehistory of Albanian *mos*, so that there is no principled basis for deciding the issue of the origin of this use within Albanian, and the parallels between Albanian and Greek here may just as well be a matter of independent innovation or inheritance as a contact-related development.

7.6.4 Negative Fusion

Negative fusion refers to the incorporation of a negative marker together with or into a verbal host, e.g., the negative existentials Mac *nema*, Blg *njama* (both from *ne* + *ima*), Aro *nori* (< *nu* + *ari*; Markovikj 2007), Alb *s'ka* (if < *se* + *ka*),[89] all from 'NEG + have.3SG,' and Rmi *nane, nae* (< *na* + *hine*) from 'NEG + be.3SG' (Boretzky & Igla 1994: s.v.), among others.[90] Such fusion, however, is widely attested in both ancient and modern Indo-European (and other) languages, e.g., OIr *ní* 'is not' from **nēst*. It must be concluded, therefore, as discussed in Joseph 2001a, that the facts concerning negative fusion in the Balkans are interesting from the perspective of the "linguistics of the Balkans" or the "comparative syntax of the Balkans," but not from the perspective of "Balkan Linguistics" or "comparative Balkan syntax." Each language reveals

89 Fusion is involved in Albanian under the account in Çabej 2006: s.v. *s*, whereby *s'* 'not' is taken to be a reduced form of *së* that is found in Arbëresh and Arvanitika, itself a development from an unstressed form of *se* 'what?' If, however, *s'* is from **(ne)* ... **k*^{w}*id,* as noted in §7.4.1.2.3, §7.6, and footnote 70, then it is simply the expected outcome of **k*^{w}*id* and thus there is no fusion per se (except perhaps orthographically).

90 See Joseph 1990, 2001a on Greek.

interesting phenomena but their occurrence need not be attributed to language contact, and thus they are not relevant to the concerns of the investigation of the sprachbund as a contact-based phenomenon. As with so many apparent syntactic convergences, however, their presence does contribute to a certain sense of similarity among the Balkan languages from a superficial syntactic standpoint.

7.7 Clause Combining

Syntactic units of the size of clauses, with predicates and at least implicit subjects, can be combined in various ways but involve either coordination, where the clauses stand in a functionally equal relationship to one another, or subordination in some form or other, where the clauses are in a functionally unequal relationship. In coordination, either clause could in principle stand alone as a well-formed utterance, whereas with subordination, one clause can stand alone, the main clause, while the other is dependent on that main clause. Thus, under the rubric of subordination are subsumed both complementation, where the dependent clause-like unit serves to fill out the argument structure of the main (superordinate) verb, and adjuncts. The notion of adjunct takes in both adjectival clauses, i.e., relative clause structures, where a clause-like construct is adjoined to and modifies a noun head, and adverbial clauses of various kinds – temporal, purpose, causal, etc. – that are adjoined to and modify predicates or even entire clauses.

Clause-combining strategies can be areally based. For instance, one feature that is found in various language groups of South Asia that has been claimed to help define a South Asian linguistic area and that may have thus diffused through contact is the use of nonfinite invariant adverbialized forms of verbs – labelled as co-verbs, con-verbs, or conjunctive participles – to link events together serially (Masica 1976, 1992, 2001). It is thus reasonable to consider this aspect of the syntax when examining the Balkan sprachbund. On the other hand, as Slobin 1986 has shown, Turkic cyclically borrows finite subordinators from Indo-European contact languages. Accordingly, we survey here various ways in which clause-combining is effected in the Balkan languages, with particular attention to those that appear to be contact-related.

7.7.1 Coordination and Parataxis

The most interesting pan-Balkan aspects regarding coordination of clauses do not have to do with the syntax per se of such combinations, but rather with both the particular lexical items used as the linking elements and the functions of such clausal combining.[91] Regarding the former, as shown in §§4.3.3.4 and 4.3.4.1, there are numerous words

91 There is a coordinative construction related to but different from simple coordination that has Balkanological significance but does not fall under the rubric of "clause-combining"; this is the comitative agreement construction, on which see §7.9.3.

that are borrowed between the languages that serve to join clauses together into compound sentences and to join sentences together into coherent discourses. As to the latter, we note that parataxis, the juxtaposition of semantically subordinated elements without overt subordinating markers and possibly even without any overt coordinating markers, is found in all the languages and plays an important role in Balkan syntax; it is tied, functionally at least, to subordination and thus in some instances cannot be treated adequately without some reference to subordination, a subject treated in §7.7.2. The different functions of paratactic combinations, with or without overt coordinative elements, are treated in the sections that follow.

7.7.1.1 Functionally Subordinative Parataxis[92]

The most common function for parataxis is its use in place of subordination, with the paratactic clause following and playing the role semantically of a complement or an adjunct, as the case may be. Since the DMS is generally a subordinator (but not always, see §6.2.4.3.2 on insubordination), in parataxis the DMS is absent and there is generally an overt coordination marker, e.g., BRo *şi, shi, de*, Alb *edhe, dhe, e*, BSl *i, a, ta, pa*, Grk και (κι/κ' before vowels), Rmi *thaj*, *vi,* and borrowings from all the relevant contact languages, Turkish *ve* (a borrowing from Arabic), *(h)em*, borrowed into Turkish from Persian and thence into the Balkans, and native *dA* (OT *taqı*) all meaning 'and' with various nuances and clause restrictions.[93] Some of these items are colloquial, and, with the advent of standard languages, some now have a folkloric flavor.

As Sandfeld 1930: 197ff. notes, one can find examples such as those in (7.91) where the connection between the clauses is causal, i.e., an event happened and that caused some other event, but the clauses are linked only by a coordinating word meaning 'and,' *de* in the Romanian example, *ta* in the Bulgarian or Macedonian, και in the Greek:[94]

(7.91) a. Ce ai uitat de te ai întors înapoi? (Rmn)
what have.2SG forgotten and REFL have.2SG returned back
'What did you forget (so) that you returned?'

b. Kakvi sa tija jabulki ta sa tolko skapi (BSl)
what.kind are.3PL these apples and are.3PL so expensive
'What kind of apples are these that they are so expensive?'

c. Τί έχεις και είσαι πάντα συλλογισμένο (Grk)
what have.2SG and are.2SG always sad
'What (in particular) do you have (so) that you are always sad?'

92 See also §4.3.3.4 on the borrowing of complementizers, where a case involving 'and' is mentioned.

93 See Ingria 2005: 78–91 and Nicholas 2005: 95–99 for considerable discussion of the specifically complementational uses of και in Greek, with references to the relevant literature.

94 Some of Sandfeld's examples are in nonstandard orthography or are dialectal. We have chosen to leave these as they appear in his text.

Sandfeld further points to a special convergence in a related causal expression involving the verb 'do' in a lead clause with a coordinated following clause giving the meaning 'why is it that … ?':

(7.92) a. Ce facem de rămânem aici? (Rmn)
what do.1PL and remain.1PL here
'Why do we stay here?' (lit., 'What are we doing and we stay here?')

b. Kakvo pravjăt tam, ta igrajăt? (Blg)
what do.3PL there and play.3PL
'Why are they playing there?' (lit., 'What do they do there and they play?')

c. Sh bim e rrim? (Alb)
what do.1PL and stay.1PL
'Why do we stay?' (lit., 'What do we do and we stay?')

d. Tí κάνεις και ξετυλίγεις τα σκοινιά (Grk)
what do.2SG and untie.2sg the ropes
'Why do you untie the ropes' (lit., 'What do you do and you untie the ropes?')

Parataxis also serves other functions in the various languages. Especially interesting is the use of parataxis in a complement function, as a means of replacing what earlier had been infinitival complementation, for instance with verbs of causation and compulsion, as in (7.93), from Sandfeld op.cit.:

(7.93) a. Il făcu de spuse (Rmn)
him made.3SG and said.3SG
'He made him say' (lit., 'He made him and he said')

b. Έκαμε τη μάνα κ' έριξε το παιδίν της στο νερό (Grk)
made.3SG the mother.ACC and threw.3SG the child her in.the water
'He made the mother throw her child into the water' (lit., 'he made the mother and she threw her child into the water')

c. E porositi e i bëri një palë (Alb)
him ordered.3SG and him.DAT made.3SG a sword
'He ordered him to make a sword' (lit., 'he.ordered him and he.made a sword')

This use even occurs following verbs with the meanings 'can,' 'begin,' and 'want,' as in (7.94), from Sandfeld 1930: 199, except as marked:

(7.94) a. Πιάνει και του λέγει (Grk)
takes.3SG and him says.3SG
'He begins to say to him' (lit., 'he.takes (= begins) and he.speaks to him')

b. Πώς εμπορεί και κοιμάται τώρα? (Grk)
how can.3SG and sleeps.3SG now
'How can he sleep now?' (lit., 'how can.he and he.sleeps now?')

c. Možeet i pra'jeet pajvani ot pesok (Mac, Ohrid region)
can.3PL and make.3SG shackles from sand
'They know how to make shackles from sand' (lit., 'they.can and they.make … ')

d. Desha e ju gënjeva (Alb)
wanted.1SG and you.PL.ACC fooled.3SG
'I wanted to fool you' (lit., 'I.wanted and I.fooled you')

e. Zu e pi (Alb; Gjinari 2007: Map 142)
take.AOR.3SG and drink.PRS.3SG
'He began to drink' (lit., 'he.took and he.drinks')

f. Vrui de v' arăshi (Aro)[95]
wanted.1SG and you.PL.ACC fooled.1SG
'I wanted to fool you'(lit., 'I.wanted and I.fooled you')

Parataxis with these verbs is noteworthy because they are the ones that maintained infinitival complementation the longest (see §7.7.2.1.1.2). Moreover, owing to the fact that they do not appreciably affect the event structure of a proposition, infinitival use with such verbs seems typologically particularly natural, while coordination, i.e., parataxis, by contrast, seems unnatural, inasmuch as coordination would seem to imply the occurrence of two separate events.[96]

7.7.1.2 Mitigating Parataxis

Quite common in the Balkans is a paratactic turn of phrase in which a negative clause is followed by a positive (main) clause connected to it by 'and'; this combination gives a mitigating meaning of 'X almost happened' or 'no sooner had X happened when … '. Examples, from Sandfeld 1930: 196, are given in (7.95):

(7.95) a. N' apuc bine a scăpa din una şi dau peste alta (Rmn)
NEG grab.1SG well INFM escape from one and give.1SG over other
'I barely escaped from one (thing) and (= when) I encounter another'

b. Nu tricură putsăne dzănle šă lăndzidză păntru muarte (Aro)
NEG passed several days and languished for death
'Scarcely had several days passed and (= when) he languished in death'

c. Ošte ne izdumal i lisicata go džasnala odzade (Blg)
still NEG fabricated and fox.DEF him pushed behind
'He had not finished speaking and (= when) the fox pushed him from behind'

d. S ndënjti shumë kohë edhé na i vjen
NEG stayed.3SG much time and here him.DAT comes.3SG
vdékja dhe plakut (Alb)
death too old.man.DAT
'He was not very alive for long and (= when) death came to him, the old man, too'

e. Τρεις ημέραι δεν παρήλθον και τα βουνά αντελάλησαν (Grk)
three days NEG passed.3PL and the mountains resonated.3PL
'three days had not passed and (= when) the mountains resonated'

This expression is rare in the Romance languages, so its appearance in Romanian and Aromanian is noteworthy in the light of its occurrence in neighboring

95 Aromanian *de* is more polysemous than some of the other conjunctions, but the basic principle here is the same.

96 For instance, with a verb subordinated to 'begin,' there is not really a separate act of beginning since all events begin somewhere at some time. So also with 'can,' in that all events that occur entail the ability to occur.

languages. The use of *i* 'and' here does not appear to occur elsewhere in Slavic.[97] Sandfeld is inclined to see this as a contact-related point of convergence in the languages, and while it is a reasonable inference, his argument is more circumstantial than direct. However, given that all of Slavic has the contrastive coordinating conjunction *a* in addition to the noncontrastive coordinating conjunction *i* – a distinction not found in the relevant non-Slavic languages – the possibility of reverse interference seems quite plausible at least for Balkan Slavic. See also §7.7.2.1.2 for a paratactic use of verbal nouns. We can also note here a similar usage in Turkish: *Daha araba gelecek de, bineceğiz de gideceğiz* (Tietze 2002: 543) 'As soon as the car[t] comes, we'll get in/up and we'll go' (lit., 'already/more/still car[t] come.3SG.FUT and, mount.1PL.FUT and go.1PL.FUT').

7.7.1.3 Juxtaposed Parataxis

The Balkan languages also show a syntagm involving the juxtaposition of two verbs, separated only by a negation marker, with no overt marker for coordination (or subordination) in the meaning 'whether VERB or not.' This construction thus involves two potentially clause-defining elements, the verbs, used in a function that is subordinative in the broader discourse, i.e., "there is an event X, whether it is the case that VERB or it is not the case that VERB." Thus even though it is discussed more fully in §4.1, this construction is relevant here as well; illustrative examples are given in (7.96) – see §6.2.2.4.3, footnote 245 regarding (c):

(7.96)
a. φύγει δεν φύγει (Grk)
leaves NEG leaves
'whether one leaves or not'

b. spune nu spune (Rmn)
says NEG says
'whether he says (so) or not'

c. Peniš se ne peniš se, šte te jam (Blg)
foams INTR NEG foams INTR FUT you.ACC eat.1SG
'whether you foam or not, I'll eat you' (see §6.2.2.3.3, footnote 245)

d. Vjen s' vjen aq më bën (Alb)
comes NEG comes so me.ACC does.3SG
'whether he comes or not, I don't care'

e. cu/di vreare, cu/di nivreare (Aro)
with/of wants with/of NEG.wants
'whether he wants to or not'

f. kyere no kyere (Jud)
wants NEG wants
'whether he wants to or not'

g. saka nejkje (Mac)
wants NEG.wants
'whether he wants to or not'

97 Thus, for example, in Russian, the contrastive coordinator *a*, which is also preserved in Balkan Slavic, could be acceptable, but *i* would not be used.

h. ister istemez (Trk)
wants wants.NEG
'whether he wants to or not'

i. ladž na ladž (Rmi)
shame NEG shame
'whether it is shameful or not'

This paratactic convergence seems to have originated, as argued in §4.1, in the one specific token shared broadly across all the languages, namely that with 'want.' Besides the examples cited in (7.96e–h) from Aromanian, Judezmo, Macedonian, and Turkish, this expression occurs also in Greek (θέλει δεν θέλει), Bulgarian (*šte ne šte)*, Romanian (*vrea nu vrea*), Albanian (*do s'do*, cf. *domosdo* 'certainly, for sure, etc.'), and Romani (*mangela na mangela*). And this particular instantiation may well be traceable back to a classical prototype as it is found in both earlier Greek and in Latin.[98] This highly particularized construction therefore has a place in the consideration of contact-related parataxis convergence in the Balkans.

7.7.1.4 Parataxis as a Balkanism

Given the parallels with parataxis evident in the preceding sections, it is reasonable to ask if parataxis represents a syntactic Balkanism. This question is complicated by the fact that earlier stages of at least some of these languages show evidence of parataxis. For Greek, Blass & Debrunner 1961: 247–248 identifies parataxis in Hellenistic papyri and the New Testament, as in (7.97a), and Sandfeld himself, 1930: 198, cites a New Testament example (7.97b):

(7.97) a. ἐὰν τολμήσωσι καὶ καταβῶσι (UPZ I 64.10 (156 BCE))
if dare.3PL.SBJV and go.down.3PL.SBJV
'if they dare to go down' (lit., 'if they.dare and they.go.down')

b. αποτολμᾷ καὶ λέγει (Romans 10:20)
be.bold.3SG and say.3SG
'he dares to say / he is so bold as to say' (lit., 'he.is.bold and he.says')

It is interesting to note that (7.97b) has various treatments in the modern Balkan languages, as illustrated in (7.98a–i) below:[99]

(7.98) a. smelo reče (Mac, Bible 1990a)
daringly says

b. drăznuva i kazve (Blg, Bible 1912)
dares and says

c. se osmeljava da kaže (Blg, Bible 1995b)
INTR dares DMS says

98 In the case of Turkish, the usual meaning of VERB.GPRS + VERB.NEG.GPRS is 'as soon as, no sooner than,' e.g., *gelir gelmez* 'as soon as s/he arrives.' The meaning with *ister istemez* is thus unusual for Turkish and could well be a calque on an Indo-European language and treated as an idiom rather than a construction.

99 Bibliographic information is given under "Bible" in the list of references.

d. guxon edhe më tepër e thotë (Alb, Bible 1872, 1930, orthography modernized)
dares and more much and says

e. guxon të thotë edhe më tepër (Alb, Bible 1980)
dares DMS says and more much

f. merge cu îndrăzneala pîna acolo că zice (Rmn, Bible 1908, 1924, 1962b)
comes with daring so.far.as there that says

g. cuteadzã shi dzãtsi (Aro, Bible 2004)
dares and says

h. bi darakoro vakergjas (Rmi, Bible 1995a)
without fear.GEN said.3SG.AOR

i. tromarelpe te mothol (Rmi, Bible 1990b)
dares DMS says

The borderline nature of hypotactic uses of paratactic conjunctions is also noted by Večerka 2002: 58–60 with regard to examples from Old Church Slavonic. Among others, Večerka (p. 59) cites example (7.99) from the *Sinai Psalter* (eleventh century):

(7.99) reče i pridǫ pьsьję muxy (OCS)
spoke and came canine flies
εἶπεν καὶ ἦλθεν κυνόμυια (Grk equivalent)
spoke and came dogfly
'he spoke and [so] the dogfly came'

We can also observe that MacRobert 1980: 194, 212 gives some examples of coordination where subordination would be expected on historical grounds in Middle Bulgarian texts (with reference to the Balkan situation):

(7.100) a. izvolitь … i posaditь (Rila 94–95)
chooses and seats
'[whom God …] chooses to seat [on my throne]'

b. sę pokousit …i posramit (Blg. Charters 66 (1429))
INTR tries …and shame
'[whoever …] tries […] to shame …

Such instances of parataxis suggest that this phenomenon in later stages of the language may have its roots in earlier usage; indeed, Blass & Debrunner 1961: 248 write that "coordination has gone still further in M[odern]Gr[eek]," implying that they view the modern phenomenon as an extension of the earlier usage.

This evidence, however, need not mean that contact with other languages was irrelevant. For one thing, there is too much agreement on detail – especially the typologically unusual use of parataxis after 'can,' 'want,' 'begin,' and the paratactic use of "what are you doing and … " for 'why (is it that) … " discussed in §7.7.1.1 – to argue for independent origin in the different languages. For Sandfeld 1930: 199, the earlier Greek evidence was a basis for viewing parataxis as emanating from Greek, in line with his general approach to Balkan convergences. But other scenarios exist. Contact could well have had an enhancing role in

bringing an already-existing construction into increased use. Parataxis is considered by many observers to be a particularly "vernacular" characteristic: Blass & Debrunner (p. 247) say as much, discussing it under the rubric of "Parataxis in the Vernacular," and Sandfeld (p. 196) clearly holds this view: "*un trait qui se retrouve un peu partout en langue populaire*."[100] This means that it would be expected to be common in everyday conversational interactions. Moreover, parataxis can be seen as a useful L2 discourse strategy in that it allows the speaker to be minimalistic on the syntactic front since situational pragmatics can fill in the connection between clauses. Thus, parataxis could have spread in the Balkans through regular conversational interactions among speakers of different languages who had some command of their interlocutor's language without necessarily being fully bilingual.[101] It is, in a sense, another reflex of a move towards analytic expression found in many language contact situations.

7.7.2 Subordination

Various types of subordinate constructions deserve mention in a consideration of clause-combining in the Balkans. Modifier and adjunct structures, such as relative clauses and various adverbial clauses are relevant here. However, primary importance must be given to complementation, i.e., the use of subordinate clauses or clause-substitutes to fill argument positions, i.e., subject and, especially, object, in a sentence. Complementation in the Balkans historically and to some extent synchronically, too, involves both finite and nonfinite structures; and paramount in a consideration of Balkan complementation are developments involving the reduction or outright loss of one particular nonfinite form, the infinitive, and its replacement by finite complement structures. These have long been discussed as pan-Balkan features and have played an important role in comparative Balkan syntax, as documented below. Consequently, the lion's share of the discussion in this section is directed at them, though attention is given as well to other relevant developments involving subordinate clause structures.

7.7.2.1 Complementation

As noted above, developments with the infinitive are the key issue regarding complement structures in the Balkan languages. In what follows, the facts concerning the fate of the infinitive in the various languages are presented, with attention as well to the causes of these developments and the role of Balkan language contact in them. The nonfinite forms are discussed first, and then the array of finite complementation found in the Balkans, but it should become clear that it is sometimes difficult to discuss the infinitive in isolation, i.e., without some reference to the finite

100 'A construction that is found rather everywhere in colloquial language.'
101 Recall the evidence of the conversationally based ERIC loans in §4.3.

replacements for it. Thus, some alternation in the presentation of the nonfinite and the finite facts cannot be avoided. Similarly, in discussing the synchrony and diachrony of the infinitive as the primary means by which complementation was achieved, it is impossible to ignore some uses of the infinitive that are not complementation in the strict sense but rather reflect adjunct and not argument uses.

The key distinction of finite versus nonfinite has a grounding in traditional grammar, pertaining to the presence versus absence of overt marking for categories such as person, number, and tense, and such a characterization, especially regarding person, is followed here.[102] For distinguishing particular nonfinite forms, especially infinitive versus participle, a further consideration of the typical functions for a given form is necessary; thus, complementation is a typical function for an infinitive, while (adjective-like) modification is a typical function for a participle.[103] These, and other, characterizations of (non)finiteness admittedly are imperfect and fraught with difficulties, but for the most part they work for the languages and developments under consideration here and so they are adopted without further comment.[104]

7.7.2.1.1 Nonfinite Complementation: Synchrony and Diachrony

Since the earliest work on Balkan linguistics, starting no later than Kopitar 1829 and Miklosich 1862 and with discussion in Sandfeld-Jensen 1900, and mention in Seliščev 1925, with more in Sandfeld 1930, the developments with the infinitive have been duly noted as a common Balkan feature.[105] In particular, there is mention of the "*extinction de l'infinitif*" ('extinction of the infinitive') by Sandfeld and "*der Mangel des Infinitivs*" ('the lack of the infinitive') by Miklosich, both characterizations referring in different ways to the fact that there are few to zero infinitives in use in most of the Balkan

102 In other words, at issue is limitation along some parameter (feature), hence "(non)-*fin*-ite," based on the root of Latin *finis* 'limit, boundary.'

103 There can of course be language-specific characteristics that distinguish among nonfinite forms; Pană Dindelegan 2013: 211, for instance, states that in Romanian, "the infinitive is singled out by the following features: (a) mixed marking (suffixal and analytic); (b) the ability to encode a temporal distinction; (c) its occurrence in contexts common with the subjunctive … ; (d) formal differentiation from the nominal infinitive."

104 For instance, in some languages, forms that seem to be nonfinite along some dimensions can show tense, as with infinitives in English (where *to leave* versus *to have left* in indirect discourse corresponds to present versus past tense; cf. *I believe John to leave for work every day at 7:00* versus *I believe John to have left for work yesterday at 7:00*), or voice, as in Ancient Greek (where there are active, middle, and passive infinitives, as opposed, for instance, to voice-neutral infinitives in Sanskrit). Also, some forms appear to move from finite to nonfinite in given languages across different historical stages, as argued by Joseph 1983a: chapter 2 for imperatives in Greek. Thus other language-particular facts, such as the placement of weak object pronouns relative to the form in question (relevant for Greek and Macedonian) or choice of negator (relevant for Greek), may be needed to help identify finite and nonfinite forms. In any case, however, care must be taken to distinguish forms characterized in this way, especially regarding person marking, from the impersonals (see §7.8.2.2). See Joseph 1983a: chapter 2 for a discussion of these and other problems with the definition and identification of (non)finiteness in particular languages and cross-linguistically.

105 See Chapter 2, passim, for fuller consideration of these works.

languages. These statements reveal that this phenomenon is multi-faceted, with both synchronic and diachronic aspects. The former is implicit in Miklosich's use of *Mangel* 'lack.' That is, it is a structural fact, i.e., a synchronically valid typological property, of these languages, that there is no category, and there are no forms, answering to the infinitives that occur, for instance, in Western European languages; such a category is absent or lacking in most of the Balkans.[106] The latter is seen in the historical fact that the earlier stages of most of the languages demonstrably show use of infinitival forms which are not to be found in contemporary stages, so it is accurate to talk in terms of there having been a disappearance, an "extinction" à la Sandfeld, of a category and associated forms. What the Balkan languages have instead is, as Miklosich 1862: 6 put it, "*Ersatz desselben durch ein mit einer Conjunction verbundenes Verbum finitum*" ('the replacement of the same [i.e., the infinitive] by means of a finite verb combined with a [subordinating] conjunction').

Thus there are really two historical developments – and two concomitant synchronic facts – to note regarding infinitives in the Balkans: their loss historically, leading to the absence of such a category and of such a set of forms in the modern languages, and their replacement by finite forms, leading to a synchronic situation with mostly finite complementation in the various languages.[107] Regarding the latter, it should be noted that a finite form of some sort is not the only possible replacement for an infinitive: in various Balkan languages, deverbal nouns have also moved into slots vacated by the infinitive; see §7.7.2.1.2 for details.

7.7.2.1.1.1 Synchrony of the Infinitive in the Balkans

In a sense, discussing the synchrony of the infinitive per se is easy, since there are basically no infinitives to speak of, though there are some notable exceptions to this generalization and some details at the edges of infinitival usage to be noted. The following survey of usage in the various languages makes this clear.

7.7.2.1.1.1.1 Greek In contemporary Greek, both in the standard language and in most of the regional dialects, there is but a single verbal form lacking person and number marking that is used in a complement-like grammatical function, and only

106 This means, moreover, that another significant typological fact about these Balkan languages is that they show redundant person and number marking on complement verbs in contexts where many languages, both Indo-European, e.g., French, English, or Russian, and non-Indo-European, e.g., Hebrew, Arabic, or Finnish, get by with marking just on the matrix verb. The situation in Turkish is more complex (see Göksel & Kerslake 2005: 93ff), and WRT can sometimes be like the other Balkan languages in this respect.

107 In various ways to be made clear in what follows, Albanian offers some exceptions to these generalizations, as would Romanian, under the analysis of *să* given in Pană Dindelegan 2013: 211–212 in which, parallel to the analysis of Miller 2002: §4.4 for Greek subjunctives with the marker να (contra which, see Joseph 2019f, 2020c), some instances of Romanian subjunctives with *să* are taken to be nonfinite. (See also Kim 2010, claiming [unconvincingly] that some Macedonian *da*-complements are nonfinite.) There are also formal remnants of infinitives in some of the languages, as well as new nonfinite developments, to be discussed below.

one such function at that. This is the verbal form found as a perfect tense formative with the auxiliary verb έχω 'have,' e.g., active έχω λύσει 'I.have loosened,' nonactive έχω λυθεί 'I.have been.loosened.'[108] It is not clear, though, in what sense this perfect formative could truly qualify as an infinitive; while these forms are by some lights "complements" to the auxiliary, the AUX + λύσει/λυθεί combination behaves like a unit, so that weak object pronouns are positioned in relation to the auxiliary, before with finite forms of έχω, as in (7.101a), and after with the nonfinite participial, as in (7.101b):[109]

(7.101) a. το έχω λύσει (*έχω το λύσει / *έχω λύσει το)
it.ACC.N have.1SG loosen.PRF
'I have loosened it'

b. έχοντάς το λύσει (*το έχοντας λύσει / *έχοντας λύσει το)
have.ACT.PTCP it.ACC.N loosen.PRF
'having loosened it'

Moreover, there is no subordinating element accompanying these forms as there is with other complementation in Greek. As a result, they have no properties that would identify them as infinitives and no uses other than in the perfect system. There is thus no economy or special insight to be gained in the grammar by categorizing them as infinitives, as opposed to some other suitable category. In fact, Joseph 1983a: 77–80, building on a suggestion of Hesse 1980: 13, argues that forms like λύσει and λυθεί are best treated as perfective *participles* and shows that the participial interpretation fills a gap in the array of participles in the language.[110] By contrast to what is seen with the perfect tense system, in all complement structures and in most adjunct structures there are finite verbs, marked for person and number.[111]

Although a fully functional infinitive is absent in most varieties of Greek, there are some marginal dialects in which infinitives are still to be found, in some instances with some interesting alterations. In particular, as summarized by Joseph 1983a: 73–74 – drawing on earlier sources – the Greek of southern Italy, known as Grico (also spelled Griko) in Apulia and Greco (also spelled Greko) in Calabria has an infinitive that is used as the complement to the modal verb 'can' (*sonno*, with the form *sodzo* in some villages), to the verb of compulsion 'make' (*kanno*), and to at least some verbs of

108 And so also in the pluperfect (e.g., είχα λύσει 'I had loosened'), future perfect (e.g., θα έχω λύσει 'I will have loosened'), and with all forms throughout the perfect system.

109 Exceptions to this generalization are adverbs and the subject, which can intervene, e.g., έχει η Μαρία φύγει 'Maria has left,' έχω ήδη λύσει το σκοινί 'I have already loosened the rope.' This same situation applies to Macedonian, e.g., *toj ima dve-tri čaši ispieno* 'He has drunk two or three glasses [apparently; based on the way he is weaving between lanes on the expressway].'

110 See §7.7.2.2.2 for more on participles. High-style participles of Katharevousa origin are excluded from consideration here (though some grammars, e.g., Holton et al. 1997, include them in a comprehensive, pan-style, account of the Greek that average educated speakers are exposed to and can use).

111 An exception is the present active participle, e.g., έχοντας in (7.101b), which functions as a sentence adverbial, setting the circumstances under which the event in a main clause takes place. See §7.7.2.3.1 on such verbal adverbs across the Balkans.

perception (e.g., *akuo* 'hear'), and it can also occur in some nominalization uses; these usages, however, are not exclusive, since finite complements are also possible with these verbs. Similarly, in the variety of Pontic Greek known as Romeyka, still spoken by Muslims in eastern Turkey, as described by Sitaridou 2013, 2014ab, 2021, the infinitive has a fairly robust use, albeit often with personal endings added onto the historical infinitive, what Sitaridou refers to as an "inflected infinitive," which is also a characteristic of Turkish infinitival constructions, e.g.:

(7.102) a. íxa ipína
had.1SG say.INF.1SG
'if I had said … '

b. íxe ipína
had.3SG say.INF.1SG
'if I had said … '

c. prin písin to faí
before do.INF the food
'before cooking the food'

Among the verbs that still govern infinitives in Romeyka are negated *poró* 'can,' negated *θélo* 'want,' and (as in (7.102ab)) *éxo* 'have' as a conditional mood auxiliary, personal in (a), impersonal in (b), all of which typically or exclusively have their subject understood as identical to the complement subject (the so-called "like-subject" condition). Such inflected infinitival forms were reported for the same area in the late nineteenth century by Deffner 1878, leading ultimately to considerable controversy a century or so later as to whether Deffner had reported these forms accurately and whether these were infinitives or not (see Tombaidis 1977); Sitaridou's work, and that of Mackridge 1987 before her, would seem to decide the issue in favor of taking the forms to be infinitives, though altered so as to allow personal endings to be added onto the infinitive proper, a feature consistent with Turkish as the dominant contact language.

The upshot is that some outlying dialects of Greek do show infinitives even today, although not as robustly as in some earlier stages of the language, and subject to later contact influences in the case of Romeyka, but such is not the case with the dialects of mainland Greece, including the standard language. This geographic distribution of the occurrence of infinitives within the Greek-speaking world is significant, and the relevance of infinitival geography in Greek is taken up below in §7.7.2.1.5.

7.7.2.1.1.1.2 Balkan Slavic Considerations similar to Greek regarding the absence of an infinitive hold for Balkan Slavic, and for some of the languages to an even greater extent than for Greek.

7.7.2.1.1.1.2.1 Macedonian Macedonian in the contemporary standard language shows a total absence of an infinitive, with at best a few lexicalized remnants, as in *može bi* 'maybe' (etymologically, 'it.can (to.)be'), and no other uses, productive or otherwise. There is some infinitival usage reported into the twentieth century for regional dialects and folk poetry in various constructions, especially modals and

prohibitives. Peripheral Macedonian dialects closest to Serbian and Bulgarian, e.g., Tetovo, Kriva-Palanka, Gorna Džumaja (now Blagoevgrad, in Bulgaria), Gevgelija, as well as Tikveš, have – or had – infinitive remnants similar to what is still found in Bulgarian.

For instance, Stankiewicz 1986a: 211, drawing on Stoilov 1904: 210, writes that "In the Macedonian dialect of Gorna Džumaja, the truncated form [of the infinitive] is used only after *stíga*."[112] Gorna Džumaja is in the Ser-Nevrokrop dialect region, located entirely in Bulgaria and Aegean Macedonia, and *stiga* 'it's enough; enough (of)' is one of the few predicates that allows for infinitival complementation for at least some speakers of standard Bulgarian (although nowadays *stiga* is followed by a preterite [Nicolova 2008]; see §7.7.2.1.1.1.2.2), so that contact with Bulgarian may well have enhanced the dialectal Macedonian infinitival use. But even so, the infinitive here is restricted to the singular, since Stankiewicz (p. 211) notes that this construction "is conjugated in the plural by means of the second person plural ending of the aorist," e.g., singular *stíga kopá* 'Enough (of) digging(, you)!' with an infinitive, but *stíga kopáhte* 'Enough (of) digging(, you all)!' with the inflected aorist form. Presumably the infinitive, which is homophonous with the aorist, was susceptible to being reanalyzed as an aorist, i.e., as a finite form, and that reanalysis led to the overtly inflected form in the plural.

A similar pattern is seen with prohibitives. Koneski 1981: 151 cites a prohibitive *nemoj nosi* 'Don't wear!' from folk poetry, where *nemoj* is a verb-like element that governed an infinitive as its complement, as its etymology would lead one to expect.[113] He further cites derivatives of *nemoj* occurring with infinitives from the Gevgelija region, *nim se kosi* 'Don't get angry!,' and from the Tikveš region, *numu lafi* 'Don't speak!' Gevgelija is in the Lower Vardar dialect region, which borders on the Pirin Macedonian dialects to the east, which are in contact with Bulgarian, where infinitival remnants are marginally more alive (see §7.7.2.1.1.1.2.2). Tikveš, however, is entirely within the Republic of North Macedonia, in the East Central group (bounded by the rivers Crna and Vardar on the west and east, respectively, and the mountains of the Greek border on the south). Yet, there are signs here too of reanalysis in that for expected dialectal *nemoj vika* 'Don't shout!,' with an infinitive, Koneski (p. 152) notes the possibility of *nemoj vikaj*, with an inflected imperatival form *vikaj* in place of the infinitive. And, Miletič 1934 gives similar sorts of examples from Macedonian dialects of Nevrokop (modern Goce Delčev), Ser (Grk Sérres), and Lagadina (Grk Langadas), with second-person indicative forms, not imperatives, following *nemoj*, e.g., *nemoj sediš tuka* 'Don't you.sit there!' (cf. §7.6.2; Standard Macedonian would be with the DMS *nemoj da sediš tuka* or the negated imperative *ne sedi tuka*).

112 "Truncated" here refers to the absence of the infinitival ending *-ti*, as seen in *možebi* just cited.

113 *Nemoj* and *nemojte* historically derive from imperatives, singular and plural respectively, of the root *mog-* 'can' negated with a prefixal *ne-*, so that they are etymologically 'you can't … !,' i.e., 'you should not' or 'you must not,' prohibitively interpreted.

Finally, some infinitives occur outside of these regions and uses. Stankiewicz 1986a: 210, citing Miletič 1934, states that so-called "long" infinitival forms in *-ti*/*-t* "are fairly well preserved in folk poetry, and have been found occasionally in the speech of older people." And, Koneski *ibid*. points to infinitives in the early nineteenth century writers Kiril Pejčinovikj and Joakim Krčovski, both of whose writings were significantly influenced by Church Slavonic. Remnants of the infinitive in *-ti* survived in the southwest margins (Kostur, Lower Prespa) as verbal nouns (cf. the situation in Aromanian §7.7.2.1.1.2.3.1) into the early twentieth century (Koneski 1981: 152–153). But beyond the uses sketched here, infinitives are totally absent from modern Macedonian.

7.7.2.1.1.1.2.2 Bulgarian As signaled in the previous section, infinitival usage in contemporary Bulgarian is greater than in Macedonian, though increasingly marginal. The infinitive can still occur as an option with some verbs at least in literary or old-fashioned usage, and thus is at least recognizable to educated Bulgarians of older generations, even if such forms occur very rarely colloquially; *moga* 'can' and *smeja* 'dare,' affirmative and negated, are mentioned in various sources as possible infinitival governors, and some add *stiga* 'enough' (lit., 'arrive'), which is used at least by older educated speakers (e.g., *stiga pi* '[we've had] enough to drink'). The infinitive can also be found as an option in the prohibitive construction with *nedej/nedejte* 'don't … ! (SG/PL),' where the infinitive serves as the complement to the prohibitive word (historically a verb; see §7.7.2.1.1.2.2). Rudin (*apud* Joseph 1983a: 122) even offers an example of an infinitive with *moga*, from a translation of *I, Robot* (by Isaac Asimov):

(7.103) ne možeš go stigna
not can.2SG it.ACC reach.INF
'You can't catch up with him'

Various scholars have noted, too, the dialectal instances of infinitival use in the twentieth century. S. Stojanov 1983: 385–386 cites various nineteenth-century authors from eastern or central Bulgaria, and states that nowadays authors use the construction for dialectal or historical coloring. Stankiewicz 1986a: 210 notes that long forms of the infinitive, those with the ending *-ti* or *-t*, "occur … in idiomatic phrases that have a modal meaning (e.g., *ne je čut*, *ne je videt* 'one cannot hear,' 'one cannot see')," i.e., "(it) is not to.see/for.seeing,"[114] and, citing Miletič 1903, gives examples of the infinitive used as a verbal noun after a preposition, e.g., *voda za piti* 'water for drinking,' *vol za klati* 'ox for slaughtering.' Miletič 1934 reports on Bulgarian speakers from Thracian villages who had moved there centuries earlier from the Rhodopes and had intact (long-form) infinitives as part of their (then-) current usage; nonetheless, they showed signs of reanalysis leading to inflected forms in their place, as with Macedonian forms cited above (§7.7.2.1.1.1.2.1), for instance from Ser (Grk Sérres) and thereabouts. Dialectally, there are also inverted futures

114 See also Mirčev 1963: 212, 1978: 235.

with an inflected 'want'-based auxiliary following the infinitive (S. Mladenov 1929), e.g., *vide šta* 'I will see,' *vide šteš* 'you will see,' etc.

Stankiewicz further considers the truncated ("short") infinitive to be "well preserved in various areas of Bulgaria," though he does not say which exactly, and notes that it "seems to combine only with modal verbs … *móga* 'I can,' *šta* (which expresses the so-called 'indeterminate future') and *nedéj*/*stiga* 'do not' (which express the attenuated form of a negated command)." He continues: "Of the three auxiliaries, only the last (or its dialectal equivalents *nekáj*, *nemój*, *mi*, *n'álaj*) is productive, whereas *šta* with the infinitive has disappeared in various Bulgarian dialects and is also going out of use in the literary language."

Stankiewicz's description is interesting as he steers a middle ground between dialects and literary language and thus invites comparison with accounts to be found in a representative sampling of various grammars of standard Bulgarian since the mid-twentieth century, some of which focus more on literary usage while others of which purport to give colloquial usage. Collectively, they show a fair degree of agreement but also some significant disagreements. The chronological progression affords a glimpse of the historiography of Bulgarian grammatical description as well as the extent to which infinitives are recognized by different researchers as part of the living language and the specific parameters of their use.

Beaulieux 1950 is actually the second edition of a work first published in 1933. It is thus somewhat older than the others and, perhaps as a result, is aimed at a level of usage more oriented towards the literary language. In this work, infinitival usage is recognized (pp. 329–330) as complement to the verbs *moga*, *smeja*, *nedej(te)*, and *stiga*, and, rarely, in the indeterminate future. Other than labeling the future forms as "*très rares*" ('very rare'), he says nothing about the degree of use in these contexts. The Bulgarian Academy Grammar (Stojanov 1983) gives the same characterization of infinitival use, although also recognizing the *voda za piti* 'water for drinking' usage; moreover, it emphasizes the unsystematic nature of the infinitival uses.

Scatton 1984: 12, 29 mentions a "vestigial infinitive" but states explicitly (p. 49) that it "is used in a very limited number of constructions and is quickly going out of use." One area where it is used is in negative imperatives, e.g., *nedéj govori* 'don't (you) speak!,' though it is a "less frequent" option than "preposing *ne* or … using *nedej(te)* plus *da* plus 2nd person present tense" (p. 46). In terms of it functioning as complement to main verbs, Scatton notes its use with *moga* 'be able to,' *ne smeja* 'dare not,' but makes no mention of *stiga.*

Feuillet 1996 aims at coverage of literary usage and states explicitly (p. 209) that the infinitive "*est une survivance du XIXe siècle qui a practiquement disparu de la langue littéraire (on peut encore le trouver dialectalement)*."[115] He recognizes uses, however, with *nedej(te)* in negative imperatives, and as a possible complement to *moga* and *smeja*, but he says the complement uses are "*réellement sentis comme archaïques*" ('currently felt as archaic'), adding that "*seul ɔa + forme conjuguée*

115 "(It) is a survival of the 19th century that has practically disappeared from the literary language (it can still be found in dialects)."

serait possible de nos jours."[116] Feuillet also discusses the inverted future, saying: "*Dans un état de langue archaïque, on pouvait former le future avec l'infinitif* + ща [šta] *conjugué au present*,"[117] but branding that usage as "*practiquement impossible de nos jours*" ('practically impossible nowadays').

Hauge 1999 states categorically that "Modern Bulgarian has no infinitive" (p. 129). While recognizing a use of an apparent uninflected form in prohibitives with *nedej(te)*, Hauge writes (p. 111) that such forms from "a formal synchronic point of view … are 3rd-person singular aorists, although historically they are shortened infinitive forms."[118]

Alexander 2000b, while stating (p. 49) that "Bulgarian has no infinitive," nonetheless acknowledges the use of the infinitive with *nedej* and *stiga*. She recognizes, however, that there are individual differences between speakers as to the degree of use of these forms: "This usage is common with some speakers and rare with others" (Alexander 2000c: 14). She also mentions (2000c: 257) two fixed phrases in which "*bí* (functioning as an irregular truncated infinitive form of *săm*) also appears": *móže bi*, said to be "exactly equivalent to English 'maybe'" and "the question marker *da ne bí* …, which allows a speaker to express a supposition in the form of a question … often … with potential negative consequences for the speaker."[119]

Finally, Nicolova 2008 gives basically the same characterization of the extent of infinitival use as the Academy Grammar, i.e., with *moga* 'can,' *smeja* 'dare,' *nedej* 'don't,' and – mostly in early twentieth-century archaizing poetry – conjugating *šta* 'FUT.'

The upshot of this survey is that some use of the infinitive must be acknowledged for Bulgarian in the modern period, even if that use is somewhat restricted to a handful of constructions and quite likely impossible for some speakers, especially younger ones, and Hauge's point about the reanalysis of the infinitive as a finite aorist form is arguably valid for at least some speakers, especially younger ones. The absence of *stiga* from some of the descriptions is striking, inasmuch as other accounts acknowledge it as a predicate that can govern an infinitive, and VAF heard it in Sofia in the course of fieldwork in the 1990s. A study based on actual contemporary use in informal contexts would be desirable, and Alexander 2000bc comes closest to that, but is still impressionistically based. It may be significant, for instance, that most of the examples cited in the grammars involve overt negation (*ne moga*, *ne smeja*, and *nedej*), and *stiga* also

116 "Only *da* + conjugated form would be possible today."

117 "In an archaic stage of the language one could form the future with the infinitive + *šta* conjugated in the present tense."

118 Presumably he means 2SG, which is syncretic with the 3SG form, since otherwise it would be hard to get the semantics of the combination right; the construction, after all, means that a second-person referent should not act, so a third-person form poses serious problems for compositionality as to the meaning.

119 An alternative would be the derivation of *bi* from the aorist *bi*, with which the short infinitive is homonymous. However, since Bulgarian, unlike Macedonian, never permits the aorist after the DMS (*da*), a short infinitive seems to be the more economical analysis.

has a negative aspect to it, so that the infinitival traces may be moving in the direction of becoming negative polarity items.[120] Still, lacking such a study, the best way to reconcile the different accounts is to embrace the speaker-to-speaker variability that Alexander notes, and add in register differences (literary versus colloquial), so that one can recognize, for the language as a whole, that infinitives are still possible, albeit only marginally. Given the availability of corpora of spoken Bulgarian, an update of the situation as attested in those corpora is a desideratum.

7.7.2.1.1.1.2.3 Torlak BCMS An infinitive seems to be almost completely absent from Torlak BCMS dialects. A. Belić 1905: 478 states explicitly that it is absent, and there is no mention of infinitival use in this dialect region in later standard sources such as Ivić 1958 or Popović 1960. This accords with the judgment of Pavlović 1960: 41, although he does mention a few traces here and there in the area; these come in very restricted, but telling, contexts. For instance, he gives examples from folk songs from Prizren and Gnjilane with infinitives occurring, significantly, after *mog-* and *sme-*, e.g., *ne može odvojit* 'he cannot separate,' *ne sme uzet* 'she dare not take (it) out.' The restriction to these controlling verbs matches aspects of Bulgarian usage, and interestingly, these verbs are negated, as many instances cited for Bulgarian are (see §7.7.2.1.1.1.2.2 and footnote 120). In these songs, moreover, the infinitive seems to be used only where metrically convenient, suggesting an overall lack of robustness for the category. A. Belić 1905: 645 also notes that occasional remnants of the infinitive occur in Kosovo.

7.7.2.1.1.1.3 Balkan Romance The situation with the infinitive in Balkan Romance is not unlike that seen in Balkan Slavic, with a limited range of uses, and differences among the languages in the extent to which an infinitive is used. Aromanian, Meglenoromanian, and Romanian are surveyed here.[121] As in other matters, Aromanian is more like Macedonian and Romanian is closer to Bulgarian, although Meglenoromanian here is closer to marginal Macedonian, and thus, Bulgarian.

7.7.2.1.1.1.3.1 Aromanian Sandfeld 1930: 174, basing himself on materials from not later than the early twentieth century, reports that Capidan 1925a, 1928 "*nie catégoriquement l'emploi verbal de l'infinitif en aroumain*."[122] Similarly, Vrabie 2000: 55 states: "Aromanian does not have an infinitive." What Aromanian has in place of an infinitive is finite complement clauses introduced by *s(ă)*, as in:

120 Interestingly, though probably just coincidentally (i.e., a result of typological drift), something similar is happening with the infinitival remains in present-day Romeyka (Pontic Greek of the Trabzon area), as described by Sitaridou 2014b (and see §7.7.2.1.1.1.1).

121 The other Romance language in the Balkans, Judezmo (Judeo-Spanish), shows interesting developments regarding the infinitive, but discussion of that is deferred to §7.7.2.1.4.

122 "Categorically denies the verbal use of the infinitive in Aromanian."

(7.104) a. Pot s- cãntu
can.1SG DMS-sing.1SG
'I can sing'

b. Acãtsarãm s- cãntãm
began.1PL DMS-sing.1PL
'We began to sing'

c. Cãtse nu vrei s- cãntã?
why NEG want.2SG DMS-sing.2SG
'Why don't you want to sing?'

d. Kãshlu ljapseshti s- mãca
cheese.DEF needs.3SG.PRS DMS-eat.3SG.PRS
'The cheese needs to be eaten'

Such a pattern is by now a familiar syntagm that parallels what other Balkan languages have.

There is a systematic and productive deverbal noun in *-re* or *-ri* (Cuvata 2006) that can be formed from all verb stem classes to create abstract nouns, e.g., (using the *-ri* forms in Cuvata 2006) *cãntari* 'singing' (cf. *cãntu* 'I sing'), *cãdeari* 'falling' (cf. *cad* 'I fall'), *dzãtseari* 'saying' (cf. *dzãc* 'I say'), *durnjari* 'sleeping' (cf. *dormu* 'I sleep'), *alinari* 'climbing' (cf. *alín* 'climb').[123] This would not suffice for a category of infinitives except that, as noted in §7.7.2.1, infinitives are verbal nouns, ones that generally have an abstract sense and occur in particular functions and with a clear place in the verbal system. However, these forms do not show any verb-like traits and can have concrete nominal meanings, cf. *cãntari* 'song,' *cãdeari* 'fall,' *durnjiri* 'sleep,' and at least some can take nominal morphology such as definite or plural forms, e.g., *dzãtseari* 'an (act of) saying,' *dzãtseri* 'acts of saying.' As Gołąb 1984a: 106 states, the "function [of the *-re/-ri* form] is that of a noun."

Second, and more importantly, even though there are clear uses of *-re* forms as nouns, e.g., *Nã lipseáshte mãcáre* 'We need food,' it is reported in Vrabie 2000: 65 that "there are native speakers who feel that these deverbals have preserved some of their former verbal meaning. Cf. *Cáshlu va mãcáre* 'The cheese should be eaten,' or 'The cheese needs eating' where *mâcáre* is felt as being a verbal form." Markovikj 2007: 166 has the following example in his discussion of verbal nouns:

(7.105) Va multu imnari pănã Bituli
wants much go.VBLN to Bitola
'It is necessary to go (lit., 'going') to Bitola a lot'

The Macedonian equivalent for (7.105) is *saka mnogu odenje vo Bitola*, which is word-for-word identical, with a verbal noun (*odenje*). In principle, one could argue that such uses of the Aromanian verbal noun preserve an older verb-like usage. However, given the exact parallel with other Balkan languages (see Koneski 1986: 186; see also §7.7.2.1.2), especially Macedonian, where the verbal noun is descended directly from the verbal noun of Common Slavic, the noun-like morphological evidence in

123 These forms of course continue the Latin infinitives in *-re* (see §7.7.2.1.1.2.3.1), and are actually called "infinitives" in some accounts, e.g., Papahagi 1974: 62.

Aromanian suggests that at best such intuitions offer a window into an earlier stage, perhaps still within the grasp of some older speakers, when these deverbal nouns may have had more verbal characteristics and fewer nominal traits; this is especially so when one takes into consideration the recent historical evidence of roughly the past one hundred years, as discussed in §7.7.2.1.1.2.3.1.

It is thus safe to conclude that, as noted by earlier authors, there is no synchronic infinitive in contemporary Aromanian.

7.7.2.1.1.1.3.2 Meglenoromanian In Meglenoromanian, unlike its South Danubian Romance relative, Aromanian, there is reason to recognize a category of infinitive in present-day usage. Atanasov 2002: 227–232 notes many uses for the long-form infinitives, those with *-ri* (cf. Aro *-re*, Lat *-re*). These occur in verbal contexts, especially in complementation, while short forms (without *-ri*) are quite restricted. The long form occurs after *puteari* 'can, be able,' *trubăiíri* 'need,' and *ţireári* 'need,' as illustrated in (7.106):

(7.106) a. Pot sirbíri
can.1SG work.INF
'I can work'

b. Nu mi pot sculari
not REFL.1SG can.1SG raise.INF
'I can't get up' (lit., 'raise myself')

c. Ti poţ dúţiri singur
you can.2SG go.INF alone
'you can go alone'

d. Trubăiá videári ţísta lúcru
need.3SG see.INF this work
'This work needs to be seen' (lit., 'this work needs to.see')

In this last use, in Romanian, with its verb 'need,' *trebuie*, one finds the supine, a nonfinite form that occurs only somewhat occasionally and mostly in fixed phrases in Meglenoromanian (Atanasov 2002: 235). Meglenoromanian can also use the long infinitive in a periphrasis with a fixed (3SG) form of 'want' that generally has a presumptive (evidential) sense, but occasionally can be a simple future, e.g.,:

(7.107) a. Vă iri niputută
FUT.3SG be.INF sick.F
'She must be sick'

b. Lăna nu vă júnḑiri tri ună bluză
wool.DEF not FUT reach.INF for a blouse
'The wool will not be enough for a blouse' (lit., 'will not reach to a blouse')

Finally, there are some exclamatory uses of the long infinitive to express indignation, e.g.,:

(7.108) Ţeástă uşă nu si dişcl'ídi, muríri
this door not REFL opens die.INF
'You could die and this door would not open!' (lit., 'this door doesn't open, [to] die!')

The short infinitive, by contrast, is found only in frozen expressions, mostly in curses, as an affective variant of the long infinitive, and only with the 3SG form of 'want,' e.g.,:

(7.109) Lupu vă ti mâncá!
wolf.DEF FUT you.ACC eat.INF
'May the wolf eat you!'

Interestingly, the more widely used long form, *mâncări*, is also possible in (7.109).

There is not much cause then to think of the short form as a productive infinitive, but the long form has various verbal uses and seems to be alive, even with some limitations on its use. Nonetheless, finite clausal substitutes for infinitives do occur; as Sandfeld 1930: 174 remarks: "*En méglénite … la construction normale est celle avec des propositions subordonnés*".[124]

7.7.2.1.1.1.3.3 Romanian Compared to the other Balkan Romance languages, the infinitive in Romanian is fully instantiated, with multiple forms and a wide range of uses (see Dragomirescu & Nicolae 2016 on more recent innovations). In most of these uses, however, finite forms marked with *să* can substitute, so that it is impossible to discuss the infinitive by itself without some reference to the finite alternatives. The Romanian infinitive occurs in two formally differentiated types and a third analytical type, which are functionally differentiated as well.

There are the long infinitives in *-re*, and the short infinitives, which lack *-re* and have a characteristic vowel suffix, different for each of the main conjugational classes, and the short forms in turn can occur as bare forms or with the "analytical … invariant proclitic marker *a* … [which] functions as an inflectional head (Jordan 2009: 181), similarly to the marker TO in English and to the marker $SĂ_{SUBj}$ of the Romanian subjunctive" (Pană Dindelegan 2013: 211–212).[125] As to function, the long forms have "the inflection and semantico-syntactic behaviour of an abstract noun," e.g., with definite forms and plurals, whereas the short forms "function only verbally" (Pană Dindelegan 2013: 215–216). These verbal functions, both for bare forms and those with *a*, include the following (adapted from Gönczöl-Davies 2008: 121, whence the examples), some of which, e.g., (d) and (f), are not complementation *sensu stricto*:

(7.110) a. subject of a sentence:
A iubi este uman
INFM love.INF is human
'To love is human'

b. complement of prepositions, including *de*, which functions as a complementizer:

i. Pentru a reuşi, trebuie să perseverezi
for INFM succeed.INF must DMS persevere.2SG
'In order to succeed, you must persevere'

ii. Dorinţa de a câştiga este des întâlnită
desire COMP INFM win.INF is densely known
'The desire to win is widespread'

124 "In Meglenoromanian … the normal construction is that with subordinate clauses."
125 The parallel is lexical as well as formal. Romanian *a* can also mean 'to.'

c. complement with indefinites with *a avea* 'to have':[126]
   i. Nu are ce mânca
      NEG has.3SG what eat.INF
      'He does not have anything to eat'
   ii. Nu am cu cine vorbi
      NEG have.1SG with who talk.INF
      'I do not have anyone with whom to talk'

d. directions and interdictions:
   i. A se desface cu grija!
      INFM INTR open.INF with care
      'To be opened carefully!'

e. one future tense formation and the present conditional mood:
   i. voi pleca 'I.will leave'
      vei merge 'you.will go
   ii. aş pleca 'I.would leave'
      ai merge 'you.would go'

to which can be added two more (Zafiu 2013a: 36, 2013b: 576–577, 581):

(7.110) f. prohibitions (negative imperatives):
   i. Pleacă! 'Leave!' ~ Nu pleca! 'Don't leave!'
      IMPV.2SG NEG INF
   ii. Taci! 'Be silent!' ~Nu tăcea 'Don't be silent!'
      IMPV.2SG NEG INF

g. the (optional) use as complement to some modal verbs, especially *a putea* 'can; be able' (bare infinitive) and *a părea* 'seem' (infinitive with *a*):
   i. El poate pleca
      he can.3SG leave.INF
      'He can leave'
   ii. Pare a ploua
      seems.3SG INFM rain.INF
      'It seems to be raining'

For almost all of these uses, finite clauses with *să* in place of the infinitive are possible and in some instances preferred in colloquial usage; compare (7.110bii) and (7.110gii) with (7.111):

(7.111) a. dorinţa să plece
      desire.DEF DMS leaves
      'the desire to leave'

b. Pare să plouă
   seems DMS rains
   'It seems to be raining'

Both options reflect possibilities in the standard language, so that present-day Romanian must be said to show variation in the realization of the infinitive as a

126 These are not complements in the strict sense, though they do "complete" the indefinite pronouns otherwise left hanging; they might be better classified as infinitival relative clauses.

category. Importantly, there are dialectal, stylistic, and syntactic dimensions to the variation.

The range of uses noted in (7.110a–g) essentially covers the syntactic dimension, except that not all of these contexts are equally hospitable to the infinitive. Pană Dindelegan 2013: 221 observes that "when the form in question occurs as the argument of the verb, the degree of replacement by the subjunctive is higher if the subject of the embedded verb is different from the matrix subject," as in:

(7.112) Vreau să plece Ion$_j$ cât mai repede
want.PRS.1SG DMS leave.SBJV.3SG Ion.NOM as more quickly
'I want Ion to leave as quickly as possible'

A consequence of this observation is that "like-subject" conditions between embedded and matrix verb favor infinitives, and indeed the modal verb *a putea* 'can,' prototypically a "like-subject" predicate, is one of the more robust venues for the infinitive, noted as an infinitival context in most older and more recent descriptions of contemporary Romanian.[127] These conditions do not rule out the occurrence of finite forms in place of the infinitive – *poate să plece* 'he can (that he) leave' is perfectly acceptable – but they do enhance the likelihood for the occurrence of an infinitive. And, enhancement is the right way to characterize the situation, since it is a matter of percentage of use; Pană Dindelegan 2013: 221, referring to Schulte 2007: 292, 294–295, 303–304 for "notes on the proportions," observes that "there are certain [other] verbs which also select the infinitive, but not in the same proportion as *putea*."[128]

As far as dialects are concerned, there is a north-south split in the degree to which the infinitive is found as opposed to iso-functional finite clauses. As might be expected, regions that were Ottoman for a longer period (Wallachia, Moldova) are less likely to use infinitives. Pană Dindelegan 2013: 221 describes it thus:

> The infinitive is best preserved in the northern area of Maramureş and Crişana, especially in quasi-frozen structures, after modal verbs (*putea* 'can', *trebui* 'must', *vrea* 'want', *avea* 'have') and aspectual verbs (*da* 'be on the point of', *începe*, *prinde*, *a se pune* 'begin') (Farcaş 2006)).

The regions noted here are at the northerly geographical boundaries of the Balkans, where contact is with languages with robust infinitives. It can be noted that some of these options, for instance the infinitive with *trebui*, are not possible in the contemporary standard language.

The stylistic dimension to infinitival usage was noted by Nandriş 1945: 184, who stated that the infinitive "is sometimes used by the best authors for a stylistic

127 Joseph 1983a: 162, 166 cites, for instance, Nandriş 1945, Pop 1948, Iordan 1954, and Kazazis 1965 for this fact, confirmed by Pană Dindelegan 2013; exceptionally, Gönczöl-Davies 2008 does not seem to mention this.

128 Such verbs include *binevoieşte* 'is willing,' *caută* 'tries,' *continuă* 'continues,' *începe* 'begins,' *îndrăzneşte* 'dares,' and *reuşeşte* 'manages.'

effect." Pană Dindelegan 2013: 221 adds the element of text type, and especially content, offering the following description:[129]

> A tendency to return to infinitive structures, on the Romance pattern, can be noted in texts showing educated usage – especially those belonging to journalistic, scientific, juridical, and administrative styles, but also in other types of scholarly texts.

Given that education seems to play a role, there is thus a social and possibly even socioeconomic dimension here as well (and see below, §7.7.2.1.1.2.3.3, for more on the notion of a "Romance pattern").

Overall, though, the infinitive in at least some (higher) registers of Romanian appears to be healthier than elsewhere in the Balkan sprachbund.

7.7.2.1.1.1.4 Albanian Taking a purely synchronic view of present-day Albanian usage as far as an infinitive is concerned, one has to first of all distinguish Geg usage from Tosk usage, as these two main dialects differ as to the form and extent of use of an infinitive. To get the best picture of the Albanian situation, however, it is important to broaden the temporal scope somewhat to take in pre-modern Albanian.

Geg Albanian has a fully productive infinitive that has the superficial form consisting of a morpheme *me*, presumably connected in some way with the preposition *me* 'with,' followed by the participle, e.g., *me qeshun* 'to laugh.' Moreover, this formation has a wide range of uses appropriate to an infinitive, e.g., complementation (to verbs, nouns, and adjectives), especially object complementation, though subject complementation is possible too, expression of purpose and other adverb/adjunct functions, and the like:[130]

(7.113) a. Fillova me qeshun
begin.AOR.1SG INFM laugh.PTCP
'I began to laugh'

b. Ka pasë fatin me njohun shumë njerëz të letrave
has.3SG have.PTCP luck.DEF.ACC INFM know.PTCP many people PC letter.PL.OBL
'He has had the good fortune to know many men of letters'

c. Mund të jetë e vështirë me ngrënë dhe me pi
can DMS be.SBJV.3SG PC difficult INFM eat.PTCP and INFM drink.PTCP
'It can be difficult to eat and to drink'

d. Me i ikë rrezikut nuk âsht ligështi por urti
INF it.DAT leave.PTCP risk.DAT.DEF NEG be.3SG cowardice but prudence
'To avoid danger is not cowardice but prudence' (Camaj 1984: 247)

129 Note that the quotation refers to the usage as a "return," i.e., a conscious archaism, in this case based on the prestige of Western European Romance.

130 Not all Geg dialects have all the possible uses of the Geg infinitive. Thus, for example, after verbs of motion and necessity, East Central Geg uses the subjunctive rather than the infinitive. Moreover, in many Geg dialects, the subjunctive competes with the infinitive (see Gjinari 2007: Maps 430–432; cf. also Boretzky 2014).

e. Gjergji shkoi përjashta me mësue
Gjergji went.3SG.AOR abroad INFM study.PTCP
'Gjergj went abroad (in order) to study'

The infinitive is also used in the Geg future formation with *kam* 'have,' as discussed in §6.2.4.1.4, e.g., *kam me qeshun* 'I will laugh' (lit., 'I.have INFM laugh.PTCP').

By contrast, Tosk Albanian has an infinitive-like construction that consists (at least from a historical perspective) of the preposition *për* 'for' + PC *të* + participle, e.g., *për të pirë* '(in order) to drink.'[131] This form is especially evident in the standard language. This Tosk infinitive is used in many of the same ways as the Geg infinitive is, e.g., as complement to verbs, nouns, and adjectives, as in (7.114abc) and in the expression of purpose, as in (7.114d):

(7.114) a. Mos luftërat shërbejnë për të shitur armë?!
NEGM wars.DEF serve.3PL for PC sell. PTCP weapons
'Don't wars serve to sell weapons?!'

b. propozimi për të ndryshuar pjesën e rezolutës
proposal.DEF for PC change.PTCP part.DEF.ACC PC resolution.DEF.GEN
'the proposal to change part of the resolution'

c. Është e vështirë për të thënë
is PC difficult for PC say.PTCP
'It is difficult to say'

d. Shkova në Petrograd për të raportuar mbi gjëndjën
went.AOR.1SG to Petrograd for PC report.PTCP about situation.DEF.ACC
'I went to Petrograd (in order) to report on the situation'

As mentioned in §6.2.4.1.4, there is also a future use for this infinitival with *kam* 'have,' found in various Tosk dialects (Gjinari 2007: Map 126), despite the general occurrence of a 'want'-based future in Tosk: *kam për të punuar* 'I have to work (hence I will work).' This substantival origin is probably the historical source, and the infinitival quality is similar to that found in certain constructions in Macedonian and Aromanian with deverbal nouns (see footnote 131 and §7.7.2.1.2).

As discussed more fully in §7.7.2.1.3, in many of these uses in the standard language, the infinitive competes with finite variants; (7.115), for example, has a finite complement paralleling the infinitive usage of (7.114a):

(7.115) Këto shërbejnë të zgjerojnë kuptimin
these serve.3PL DMS broaden.3PL understanding.ACC
'These serve to broaden understanding'

Tosk *do me thënë* 'that is to say' (*do* 'wants.3SG' + *me* 'infinitival marker' + *thënë* 'say.PTCP') is a Gegism, but with Tosk phonology (Černjak 1973: 12).[132] Similar

131 The combination of PC + participle functions on its own as a verbal noun, with *të* serving a nominalizing function. The preposition *për* 'for' corresponds exactly to BSl *za* 'for,' Aro *ti* 'for,' e.g., Alb *ujë për të pirë,* Mac *voda za pienje,* Aro *apã ti beari* 'water for drinking,' where *të pirë, pienje*, and *beari* are all deverbal nouns.

132 This phrase might represent a calque on Western European models, perhaps based on French *c'est à dire* 'that is to say,' structurally with a preposition and an infinitive and appropriate semantics, and *ça veut dire* 'that means' (lit., 'that wants to-say'), with use of the verb 'want' in a similar expression.

Gegisms in Tosk are the discourse expressions *me thënë të vërtetë* 'to tell the truth' and *me thënë të drejtën* 'to be honest' (lit., 'to say right.DEF').[133]

These forms can be described as true infinitives, not only in morphologically lacking person and number marking, but also functionally in serving typical infinitival uses. On the other hand, Albanian deverbal nouns (PC + PTCP), in addition to occurring with constructions with 'for' (cf. footnote 131), can decline like nouns, with which the expected PC is *të*.[134] Unlike other substantivized participles, however, the infinitive (or, perhaps, supine) can support weak object pronouns, e.g., *për t'i bërë* '(in order) to make them.' However, in these constructions *për* is crucial for indicating intent. On the other hand, substantivized participles generally observe a pattern of having passive meaning if formed from a transitive verb and active meaning if from an intransitive (e.g., *i goditur* 'one who is hit' from *godit-* 'hit (someone)' versus *i ikur* 'one who has escaped' from *ik-* 'leave, flee'); the Tosk infinitive, however, is inherently active – *për të goditur* '(in order) to hit,' *për të ikur* '(in order) to flee' – and must be overtly converted into a nonactive form via the voice marker *u*, e.g., *për t'u goditur* '(in order) to be hit.' This nonactive form, however, is not available in all Tosk varieties; in Arvanitika, for instance, such forms are impossible, so that *për t'u bërë 'to be done' is ungrammatical, suggesting that the specifically verbal *për të bërë* construction is restricted mostly to the standard language.

These infinitives or infinitival uses in contemporary Albanian derive from a far more complex historical situation; this diachronic side is taken up in §7.7.2.1.1.2.4.

7.7.2.1.1.1.5 Romani There is no native infinitive in any Romani dialect. However, some Romani dialects of eastern Bulgaria that have Turkish conjugation for Turkish verbs (see §6.2.1.1.1 and Table 6.18) have adopted the Turkish infinitive in Turkish verbs as in example (7.116), cited in Friedman 2020b: 244, based on materials in RMS:

(7.116) Rači lijom o grastis gijom te ajda-ma-a
yesterday take.AOR.1SG DEF horse.ACC went.AOR.1SG DMS ride-INF-DAT
'Yesterday I took my horse and went for a ride'

All other Romani dialects with infinitival constructions clearly developed them after the exodus from the (linguistic) Balkans as a result of contact with languages that have infinitives.[135] As Beníšek 2010: 47 puts it:

133 Cf. the verbal adverb in standard Bulgarian, a feature absent from the dialectal base of the standard language that was consciously introduced (see §6.2.2.4).

134 See §7.7.2.1.1.1.4 above, and Joseph 1983a: 89–91, for more discussion on the analysis of *për të* synchronically.

135 See Friedman & Joseph 2019 and sources cited therein.

> Romani dialects do not have any inherited infinitive. The common ancestor of Romani dialects in the Greek-speaking area, recently labelled Early Romani (ER) (cf. Matras 2002; Elšík and Matras 2006), is assumed not to have had the category of infinitive. Its role in the complementation of modal as well as other verbs was supplied by a subordinate clause with a finite verb in the subjunctive mood, introduced by a non-factual complementiser (see Elšík and Matras 2006:84). This construction still characterises a majority of dialects, particularly, but not only, those constantly spoken in the Balkans.

There are Romani dialects that have innovatively created infinitives, generally based on what were once subjunctive phrases, but these have not arisen in the Balkan sprachbund.[136] Thus Balkan Romani has only finite complement clauses headed by the complementizer *te*, as in (7.117), taken from the south Balkan dialect of Parakálamos (Beníšek, p. 48, drawing on Matras 2004):

(7.117) Kam-ama te av-av-Ø demosiγráfos
want-1SG DMS become-1SG-SBJV journalist
'I want to become a journalist'

The earlier diachrony of the infinitive in Romani that gave rise to this present-day situation is presented in §7.7.2.1.1.2.5.

7.7.2.1.1.1.6 West Rumelian Turkish The synchronic situation with nonfinite complementation in West Rumelian Turkish (WRT) parallels what is seen in other Balkan languages. As discussed by Friedman 1982c: 31 (and chronicled therein, passim), WRT shows numerous Balkan features, and among those features is the "use of the optative-subjunctive in place of other finite and nonfinite verbal forms" (see also Kakuk 1960 on WRT in Bulgaria). For instance, he cites (7.118ab), with a person-marked optative form, where Standard Turkish would have an embedded verbal noun in *-ma* with a possessive suffix rather than a verbal ending, as in (7.119ab):

(7.118) a. Lâzımdır calışalım (WRT)
necessary.COP.3SG work.OPT.1PL
'We need to work'

b. Sonra başlayacam çalışam (WRT)
after begin.FUT.1SG work.OPT.1SG
'Then I'll begin to work'

(7.119) a. Lâzım çalışmamız (StTrk)
necessary work.VBLN.1PL.POSS
'We need to work'

b. Sonra çalışmaya başlayacağım (StTrk)
after work.VBLN.DAT begin. FUT.1SG
'Then I'll begin to work'

136 Matras 2002: 161 notes that for the non-Balkan dialects that have "new infinitives," they involve "the reduction of person agreement in the finite complement clause of modal constructions, and the generalization of just one single form, based on a form that is selected from the present paradigm."

WRT does have nonfinite forms: an infinitive is cited in Friedman's corpus (*düzletmek* [= StTrk *düzeltmek*] 'put in order') and one sentence is given with a verbal noun in a purpose expression:

(7.120) Sen de celi misin yıkanma
you too come Q.2SG wash.VBLN
'Are you coming to wash, too?'

On the other hand, Rentzsch et al. 2018, 2020 give a comprehensive account of complementation involving potentiality in Ohrid Turkish, where the focus is on Mac *znae da* = WRT *bilir* + OPT versus StTrk *-Abilir*(-) 'can' (lit., 'knows [how] to') which is a verbal suffix and not an analytic construction. Here the semantic equivalence of Mac *znae* 'know' and Trk *bil-* 'idem' points to the possibility of mutual contact influence. It can be argued that that semantics of Trk *bil-* were adopted in Mac, and then the analytic construction in Mac was adopted in WRT. Synthetic potentials do occur in WRT, but the use of analytic constructions is clearly due to contact.

It is fair to state that nonfinite forms in complementation are far rarer in WRT than in StTrk, and finite complementation is the norm. The historical developments that gave rise to this situation are surveyed in §7.7.2.1.1.2.6.

7.7.2.1.1.1.7 Summary Regarding Synchrony To summarize, there are some infinitives in the Balkans, but generally only in a very limited way. They are limited as to the languages that have the category to any degree of robustness, limited as to the contexts in which they can occur, and limited as to the extent to which they are used even in those contexts. Macedonian, Romani, Aromanian, and Greek form a cluster with only verbal nouns (or, in the case of Greek, a de-infinitivized analytic verbal formant as well) filling the occasional infinitival slot. Tosk Albanian has usages of its verbal noun that are more infinitival than purely nominal. Bulgarian and dialectal Macedonian have some scraps of remnants. Romanian and Meglenoromanian are a bit more robust, and Geg has a robust analytic infinitive that is also ideologically salient.[137]

The infinitive in WRT is not obsolete, but rather in retreat (a retreat that has been slowed or halted by the spread of access to Standard Turkish). As shown in the diachronic sections that follow, these synchronic limitations reflect the fact not that the infinitive is a nascent category in the Balkans that has yet to be fully established, but rather that there has been a steady decline of a formerly robust infinitive in each of the languages, with preservation in certain domains, dialectal, syntactic, and stylistic.

7.7.2.1.1.2 Diachrony of the Infinitive in the Balkans

Turning now to the diachronic side of the infinitive in the Balkans, it becomes clear that the historical trend has indeed been towards the loss of the infinitive and its

137 A major point of contestation in current Albanian language policy is whether or not to admit the Geg infinitive to the standard. Planners on both sides of the argument see the Geg infinitive as "opening up the standard" to Geg, or, as Rexhep Ismajli of ASHAK commented to VAF: "*Paskajorja është lokomotivi*" 'The infinitive is the locomotive.'

replacement by finite forms, but that these developments have unfolded over a long period of time. As the survey that follows makes clear, this process is best observed in Greek due to the abundance of available data, but it can be traced as well in the other languages for which there is reliable overt or comparative, including dialectal, evidence.

7.7.2.1.1.2.1 Greek Ancient Greek had a robust verbal category of infinitive, characterized by numerous forms and numerous uses. It is not necessary to rehearse here well-known facts, and a glance at any grammar or textbook of Ancient Greek (e.g., Smyth 1920; Emde Boas et al. 2019) reveals eleven different formal categories of infinitive, depending on combinations of tense and voice as well as variation of formation within each category and also across literary dialects. As the discussion in §7.7.2.1.1.1.1 states, such is not the case in Modern Greek, and in fact throughout the history of Postclassical Greek, there has generally been a steady decline in the forms and uses of the infinitive, reaching the modern situation with virtually no infinitive at all. These developments are well known, but a brief overview is given here; for more details, see Hesseling 1892, Aalto 1953, Burguière 1960, Joseph 1983a, Tonnet 1993, and Horrocks 2010.

In the Koine period, as evident in New Testament Greek, the infinitive comes to be consolidated into a single active and a single nonactive form, and in its complementation (i.e., argument) uses it is increasingly replaced by finite subjunctive complements introduced with ἵνα '(so) that' or finite indicative complements introduced with ὅτι 'that,' the latter a type that was possible in Classical Greek as well with certain verbs. Thus ὅτι complements in the Koine era represent an extension of an already-existing structure of earlier Greek, whereas the ἵνα complements represent a new use for what previously had been an adjunct structure (final or result clause). Competition between infinitival and finite complementation was partly lexically determined, with some predicates, such as ἀρχίζω 'begin,' δύναμαι 'can,' μέλλω 'be about to,' ὀφείλω 'ought,' and τολμῶ 'dare,' occurring exclusively with an infinitive in the New Testament, and others, e.g., θέλω 'want' or (εἰμὶ) ἄξιος 'be worthy,' allowing, but not requiring, finite complements.[138] The two complement types could even occur conjoined, a phenomenon found also in Bulgarian, as noted in §7.7.2.1.1.2.2.2, and Early Modern Romanian (see (7.133)):

(7.121)

θέλω	δὲ	πάντας	ὑμᾶς	λαλεῖν	γλώσσαις
want.1SG	but	all.ACC	you.ACC	speak.INF	tongues.DAT
μᾶλλον	δε	ἵνα	προφητεύητε		(1Cor. 14:5)
rather	but	that	prophesy.2PL.SBJV		

'I want all of you to speak in tongues or rather that you prophesy'

Moreover, the infinitive in the Koine era expands in use in purpose expressions, especially with verbs of motion (Blass & Debrunner 1961), and as the object of

138 Compare, for instance, John 1:27 (ἄξιος ἵνα λύσω 'worthy that I.loosen') with Acts 13:25 (ἄξιος λῦσαι 'worthy to.loosen').

prepositions via a productive nominalization of the infinitive with the definite article (the so-called articular infinitive).

Even with the infinitive in good evidence in this period, there are also indications of weakness within the category beyond just the fact of competition with finite forms in some contexts, and the nature of this competition in some instances is quite revealing. Unlike the situation with verbal complementation, where finite forms were possible with certain verbs from early times, papyri of this era show cases of finite verbs and infinitives occurring where they never could in Classical Greek, e.g., after the article τό of the articular infinitive, as in (7.122a), and after the complementizer ἵνα of result clauses, as in (7.122b):

(7.122) a. πρὸς τὸ … δ[υ]νηθῶμεν … ἐκτελεῖν
for the.ACC.NTR are-able.1PL.SBJV finish.INF
'in order for us to be able to finish' (Joseph 1983a: 51)

b. ἵνα πέμψαι
so.that send.INF
'[make someone] (so as to) send … ' (Joseph 1983a: 50)

Such examples might well be hypercorrections, and indicate an uncertainty on the part of some speakers as to the use of these constructions. As such, they suggest that the retreat of the infinitive was more advanced among the less-educated classes of speakers and that preservation of the infinitive was a function of learnèd influence.[139]

The presence of the infinitive, but also its retreat, continued throughout the Byzantine period; nonetheless, in what may be called the Medieval period (tenth to fifteenth centuries), a few new uses for the infinitive arose, some ephemeral and some more lasting. One important one has been discussed in §6.2.4.1.1, namely the future tense formation with θέλω 'want.' In some sense, this development continued and expanded the use of an infinitive in verbal complementation, although in this case with a grammatical value. Also complement-like was the use of the infinitive with ἔχω 'have' in the perfect system (see §7.7.2.1.1.1.1), originally only in a pluperfect formation. The perfect system arose via the transformation of the past tense of the 'have'-based future of Postclassical Greek, which served as a conditional, into a pluperfect through a semantic shift rooted in semantic affinities between conditionals and pluperfects, in the account of Joseph 1983a: 62–64, 2000b, fleshing out the insight of Thumb 1910 (see also Burguière 1960; Aerts 1965).[140]

A further infinitival use that was innovative in this period is a nominalized (articular) infinitive serving as an adjunct to a sentence; this temporal infinitive (also called the circumstantial or absolute infinitive in the literature) gives the

139 This perhaps is not unlike what is seen in present-day Romanian, as described in §7.7.2.1.1.1.3.3, with French standing in as the learnèd model.

140 See §6.2.4.1.1. As far as Greek was concerned, this 'have'-future was probably a calque from Latin, as it arose in the Roman period. Conditionals, while modal, are semantically close to nonmodal pluperfects in that both are "out-of-time" with respect to a past-time action, with conditionals being irrealis and thus not temporal, and pluperfects being anterior, and thus beyond measurable time.

attendant circumstances under which or the time at which the action in the main clause takes place, somewhat like an absolute construction but sometimes, unlike absolutes, connected to a nominal in the main clause. This construction is illustrated in (7.123), where (7.123a), the earliest known example (Mihevc-Gabrovec 1973), is from Constantine Porphyrogenitus's *Book of Ceremonies* (tenth century) and (7.123b) is from the *Chronicle of Morea* (fourteenth century):[141]

(7.123) a. καὶ τὸ ἀνελθεῖν αὐτὸν κρατοῦσι τὰς χεῖρας (I.148.11)
and DEF return.INF him.ACC hold.3PL DEF hands.ACC
'And on his return, they hold his hands'

b. κ' ἐγὼ τὸ ἀκούσει το εὐτὺς ἐθλίβηκα (6066)
and I DEF hear.INF it at.once was.aggrieved.1SG
'And, on hearing it, I was at once aggrieved'

And see Kavčič 2015 for discussion of an innovative use for the infinitive, starting in Postclassical Greek, involving the infinitive in dependent deliberative questions such as τί ποιῆσαι οὐκ ἤδειν 'what to-do(INF) not I-know' (i.e., 'I do not know what to do,' from the seventh-century *Pratum Spirituale* 2996 B, by John Moschos). Furthermore, in the medieval period too, there arose new infinitival forms, most notably εἶσθαι 'to be,' replacing an earlier εἶναι.[142] These new uses and new forms suggest some vitality in the category of infinitive extending into the Medieval Greek period.

No predicate at this point, however, seems to require the infinitive to the exclusion of finite complementation.[143] Still, in this period, the infinitive survives as a complementation option with 'can,' 'dare,' 'begin,' and a few other verbs. Moreover, the evidence of the 1547 Constantinople Jewish Greek Old Testament translation – discussed in §3.2.2.10 in connection with the relevance of social networks in the retention of archaisms – is telling, as it has infinitives in roughly Koine-like uses, as well as an innovative use in the translation of the Hebrew expression 'and NP VERB-ed, VERB-ing,' with the genitive of the definite article as a generalized complementizer or even infinitival marker (see footnote 167), as in (7.124):[144]

141 See also §7.7.2.3.1 for some further discussion of this usage.

142 This new form is variously spelled, reflecting in part matters of orthographic variation but also some likely pronunciation differences. Its earliest attestation, according to Holton et al. 2019: 1750ff., appears to be from the eleventh- to twelfth-century Letter of Nikon (38.3, see Hannick et al. 2014), though they caution that a "lack of textual evidence makes it impossible to establish how long before the 11th/12th c. it came into being." Despite the spelling of the modern standard 3SG form είναι ([íne]), it is not directly related in any way to the ancient infinitive εἶναι.

143 Joseph 1978/1990: chapters 3–4 (see also Joseph 1980) claims that the Object Deletion (*pretty to look at*) and Object Raising (*easy to please*) constructions were among the last to be affected by the replacement of the infinitive, as one finds only examples with infinitival complements through roughly the fourteenth century. However, the dating of some of the texts and the possibility of influence from archaizing language, as well as the fact that the number of examples is very small, around five in each case, make it hard to be certain that there were no finite variants in competition with the infinitival types for this construction in earlier texts.

144 The probative value of the 1547 translation as discussed by Joseph 2000a has been disputed by Krivoruchko 2014; responding to Krivoruchko, Joseph 2019a reiterates and expands upon the argument given in his earlier work. The debate may well continue but it should be clear that the translation must be taken into account in some form when assessing the status of the infinitive in the sixteenth-century Greek-speaking world.

(7.124) καὶ ἐσύντυχεν μετ' ἐκεῖνον ὁ θεὸς τοῦ εἰπεῖ (Gen. 17:3)
and spoke.3SG with him DEF God.NOM DEF.GEN say.INF
'And God spoke with him, saying … '

It can be assumed, then, that the loss of the infinitive as a category was no earlier than the sixteenth century, even if it was in decline prior to that. By around the eighteenth century, as the use of the infinitive per se in the future tense was becoming moribund in a large part of the Greek-speaking world, and the innovative perfect system use was all that remained productively of the earlier infinitive, the motivation for recognizing a category of infinitive evaporated, as argued in §7.7.2.1.1.1.1, and the infinitive as a grammatical category was effectively gone.[145]

Besides the perfect system usage in the modern language, there are some lexical items that derive historically from infinitives, either lexicalized remains that are no longer synchronically distinguishable as productive deverbal derivatives or borrowings, essentially fixed phrases, from learnèd, puristic (Katharevousa) Greek. This lexical material includes inherited infinitival nominalizations like το φαγί 'food' (< τὸ φαγεῖν 'the (act of) eating'), το φιλί 'kiss' (< τὸ φιλεῖν 'the (act of) loving'), and η θανή 'death' (< τὸ θανεῖν 'the (act of) dying,' with a shift of gender), as well as puristic importations, such as φερ' ειπείν 'for example' (εἰπεῖν, old infinitive of 'say') or δούναι και λαβείν 'debit and credit' (old infinitives of 'give' and 'take,' respectively), with some dialects showing additional such forms.[146]

7.7.2.1.1.2.2 Balkan Slavic Old Church Slavonic, taken here as representative of the ancestor to Balkan Slavic, had an infinitive much as Ancient Greek did, but with less morphological diversity and less syntactic flexibility. Still, the infinitive is very much in evidence in the OCS material, with expected uses in complementation, even filling the subject argument slot, as shown in (7.125e) and including a prohibitive construction with *(ne) mog-* '(not) be able' and a permissive construction built on the idiomatic phrase *ne děi(te)* 'permit, let,' uses with certain conjunctions, and uses with various auxiliary-like verbs in the expression of future tense discussed in §6.2.4.1.3:[147]

(7.125) a. onъ že xotę opravьditi sę (Luke 10:29, Zogr)
he.M and wanting.M justify.INF REFL
'and he, wanting to justify himself … '

145 Though various future formations deriving ultimately from the infinitival type, including the eventual present-day future with θα, remain current.

146 Bogkas 1964, for instance, mentions the nominal forms το έρθει 'the coming' (< AGrk aorist infinitive τὸ ἔλθειν), το ιδεί 'the look' (< AGrk aorist infinitive τὸ ἰδεῖν 'the seeing'), and το έχει 'fortune' (< AGrk present infinitive τὸ ἔχειν 'the having') as occurring in Thesprotia, in the northwest of Greece.

147 Even though these infinitives correspond to infinitives in the Greek text that was the basis for the OCS translations, it can be assumed that they reflect possible OCS usage or else they would not have appeared in this form; moreover, some show distinct OCS syntax, such as the dative infinitival subject *ima* in (7.125d) – Greek has an accusative subject here – that point to these being real OCS syntagms.

b. možaaše bo si xrizma prodana byti (Mark 14:5, Zogr)
could.IMPF.3SG for this ointment sold be.INF
'for this ointment could have been sold'

c. něsmь dostoinъ otъrěšiti remene (Luke 3:16, Zogr)
not.am worthy.M unloose.INF strap.ACC
'I am not worthy to unloose the strap … '

d. isplъnišę oba korablja jako pogrǫžati sę ima (Luke 5:7, Zogr)
filled.AOR.3PL both ships so.that sink.INF INTR them.DAT
'They filled both ships so they began to sink'

e. a eže neumъvenami rǫkami ěsti (Matthew 15:20, As)
but whereas unwashed.INST hands.INST eat.INF
ne skvrьnitъ člověka
not defiles.PRS.3SG man.GEN
'but to eat with unwashed hands defileth not a man'

f. nemozi mene ostaviti (Supr 539.8)
not can.IMPV.2SG me.GEN leave.INF
'Do not leave me!'

g. ne děite sixъ iti (John 18:8, Zogr)
permit.IMPV.2PL them.GEN go.INF
'Let these (ones) go their way!'

Moreover, there was another nonfinite form, the supine, used in purpose expressions with verbs of motion, which was in essence a functional "alloform" of the infinitive (so Joseph 1983a: 103, following Meillet 1965: 242). And, in late OCS, this functional division between supine and infinitive began to break down, thereby weakening the nonfinite category comprising the two through the greater restrictedness on, and ultimate loss of, one member.[148] Moreover, there was also encroachment on the infinitive – and supine – by finite clauses introduced by *da*, so that there were, for instance, three ways of expressing purpose with verbs of motion: the historically older use of the nonfinite supine, the expanded infinitival use crowding out the supine, and the further impinging use of a finite *da* clause; these are illustrated in (7.126):[149]

(7.126) a. idǫ ugotovatъ města vamъ (John 14:2, Sav)
go.1SG prepare.SUP place.GEN you.DAT.PL
'I go (in order) to prepare a place for you'

b. idǫ ugotovati město vamъ (John 14:2, As)
go.1SG prepare.INF place.ACC you.DAT.PL
'I go (in order) to prepare a place for you'

148 Within Slavic as a whole, Slovene and Lower Sorbian represent exceptions in that they preserve the supine.

149 Here we follow Gołąb 1964b in distinguishing *da* as COMP (BCMS only) and *da* as DMS (BCMS, BSl) as being both functionally and etymologically distinct. For OCS, we use the gloss DMS when *da* is used in a DMS function. On the use of Romanian *şi* (StRmn 'and') as equivalent to standard *să* as DMS, see Zafiu 2019. Cf. also in this regard Aromanian *śi* as the DMS in some dialects of Aromanian (Bara et al. 2005). Such usage is reminiscent of the fact that Slavic *da* can also function as a coordinating conjunction. At issue is the surface resemblance of the words for 'and' that might have played a role in the convergence.

c. izide sějęi da sěetъ (Matt. 13:3, Zogr)
went-out.3SG the-sowing.NOM.PTCP COMP sow.3SG
'The sower went out (in order) to sow'

Moreover, other OCS uses of the infinitive also have alternative expression with finite clauses introduced by *da*, as shown in (7.127):[150]

(7.127) a. isplьnišę se dьnie da roditъ (Luke 2:6, As)
filled.3PL INTR days.NOM COMP bear.3SG
'The days were fulfilled for her to give birth'

b. něsmь bo dostoinъ da podъ krovъ moi vьnideši (Luke 7:6, Zogr)
not.am for worthy.NOM COMP under roof my come.2SG
'for I am not worthy that you come under my roof'

Thus, the predecessor to Balkan Slavic had an intact and functioning infinitive but with some incipient weakening of category and encroachment from finite alternatives.

7.7.2.1.1.2.2.1 Macedonian Against the background of OCS usage given in §7.7.2.1.1.2.2, the eventual state of the infinitive in Macedonian reflects the demise of this category and the forms that filled it. The actual historical evidence on Macedonian itself is somewhat scanty, but the outlines of the developments are clear.

The breakdown in a distinction as to function in late OCS between the supine and the infinitive continued into Old Macedonian. Koneski, in his discussion of the infinitive in early Macedonian (Koneski 1981: 150–153), cites the example of Luke 18:10 from the Macedonian Gospel of Pop Jovan of the twelfth century:

(7.128) čl̄ka dva vьnidosta vь cr̄kve pomoliti sę
men.DU.NOM two entered.3DU in church.ACC pray.INF INTR
'Two men entered the church (in order) to pray'

where proper OCS usage, as seen in the Codex Zographensis version of Luke, has a supine (*pomolitъ*). While this development expands the range of the infinitive, it also suggests that the nonfinite verbal forms overall were losing some grammatical salience. And indeed there is a reinterpretation of a supine as a finite verb in Matthew (11:1) from the Pop Jovan Gospel cited by Koneski 1981, where originally there was a supine (*propovědatъ*, found in Zographensis). The reinterpretation was aided by phonetic developments (the loss of weak jers) that drove the nonfinite forms in the direction of third-person singular forms (properly in *-tь*):

150 Example (7.127b) makes for a telling contrast with (7.125c) in that both have complements depending on the adjective *dostoinъ* 'worthy'; while the *da*-clause is used in an "unlike"-subject context, dative subjects with infinitives were possible (cf. (7.125d), and §7.7.2.1.1.2.3), so that in principle we can assume that an infinitive could have been used here. Admittedly, the Greek in this verse has an infinitive, a reflex of its own "like"-subject preference for infinitival use, and this may have played a role in the decision as to how to translate the sentence.

(7.129) prěide ot tǫdu ucitъ i propovědaetъ
went.by.3SG from there teach.SUP and proclaim.PRS.3SG
'He went forth from there to teach and proclaims' (for: ' … and to proclaim')

This suggests that finite forms were more in the forefront grammatically than nonfinite forms for some users of the language.

A more usual development with nonfinite forms at this stage was for them to be replaced by finite constructions with *da*. Koneski op. cit. states that verbs of volition, ordering, and asking are among the first to be affected in this way, with *imamь* 'have' and *načьnǫ* 'begin' being affected later. The two constructions that resisted replacement of the infinitive the longest are, not surprisingly given the discussion in §6.2.4.1.3 and §7.7.2.1.1.1.2.1, the future tense formation with forms of *xъtěti* 'want' and the prohibitive formation based on forms of *nemoj*/*nemojte* or derivatives thereof. As noted in §7.7.2.1.1.1.2.1, remnants of infinitival use in prohibitives continued dialectally into the twentieth century. In the contemporary standard language and more widely, prohibitives with *nemoj(te)* take *da* complements and the future formation has moved away from infinitives and ended up with a reduced form of *xъtěti* followed by finite forms.[151] Thus the point at which the infinitival future and prohibitive formations were affected by the infinitive-replacement process and the noninfinitival variants generalized was the point at which the infinitive passed out of Macedonian altogether.

7.7.2.1.1.2.2.2 Bulgarian By contrast with Macedonian, the historical record on Bulgarian is much richer and affords a somewhat better diachronic view of the developments with the infinitive in this language. The starting point is the same as for Macedonian, namely early South Slavic complementation possibilities as seen in Old Church Slavonic (see §7.7.2.1.1.2.2), with robust use of infinitives but some encroachment of finite *da*-clauses on the domain of the infinitive in complementation and other constructions.

By far the most thorough study of the chronological progression of the loss of the infinitive throughout the history of Bulgarian is MacRobert 1980, a dissertation that carefully examines crucial texts both quantitatively and qualitatively at different points in the development of the language to determine what they reveal about infinitival and finite-clause usage. Her findings indicate a real presence for the infinitive throughout much of Bulgarian prior to the modern era but also a clear and steady reduction in the use of infinitival constructions between Old Bulgarian and the present-day situation with just scattered and largely unsystematic remnants of infinitival use (on which see §7.7.2.1.1.1.2.2).

A key document in older Bulgarian literature after OCS is the fourteenth-century Tale of Troy (*Trojanska Priča*), a rendition of the story of the Trojan War based on a Greek version. The language is relatively progressive for the time written in what MacRobert (p. 152), following Mirčev 1963, characterizes as "comparatively

151 Imperfective prohibitives are usually formed with *ne* + IMPV, although *nemoj* + *da* + PRS can also be used. Negated perfect imperatives are unusual and have a nuance of threat.

unconstrained popular language," making the text "a reliable source for Middle Bulgarian." The infinitive is found with a large number of verbs, though the preponderance of instances of infinitives come with *xotěti* 'want,' generally in a future sense, *načęti*/*počęti* 'begin,' and *mošti* 'be able.' With one exception, those three verbs do not have competing *da*-clause complements, the exception being one instance with *xotěti*, most likely in a volitional sense. Variants with *da*-clauses occur with several of the other verbs that also govern infinitives, and MacRobert comments on the variation (p. 164) that "it is not clear that there is any marked difference in sense" between the finite and nonfinite mode of complementation. The overall occurrence of infinitives and *da*-clauses is consistent with what is found in earlier stages of the language but with movement in the direction of more restricted use of the infinitive. Moreover, there are some relevant innovations in usage, such as a *da*-clause in a like-subject context with *xotěti*, or *da*-clauses in expressions of purpose with motion verbs. State documents – charters and the like – from this period, including Wallachian charters written in a variant of Slavonic, though one with some clear colloquial Bulgarian influence on it, show a similar pattern of infinitive and *da*-clause usage.[152] Interestingly, in these documents, not unlike what is seen in Greek as early as the New Testament (see (7.121) in §7.7.2.1.1.2.1) and Early Modern Romanian (see (7.133)), infinitives and finite clauses appear conjoined, controlled by the same verb. MacRobert insightfully comments thus on such occurrences: "Such switches of construction are not the general rule, but they may indicate a lack of confidence about the handling of infinitival constructions and certainly show that infinitive and *da*-clause could be equivalent as complements of the verbs *xotěti*, *izbrati, pristǫpiti*" (p. 211).[153] The "lack of confidence" suggests that infinitives are no longer a part of colloquial usage; in fact, a hypercorrection showing the use of *da* with an infinitive is found in these texts, e.g., *da se tvoriti* '(in order) to-do [this],' a development reminiscent of the occasional use of an infinitive with ἵνα in Postclassical Greek papyri (see (7.122b) in §7.7.2.1.1.2.1).[154]

After surveying a few somewhat later texts, the Prayers of Cserged and the Tikhonravov Damascenes from the seventeenth century, showing further decline in the use of the infinitive, MacRobert concludes (p. 314) that "the infinitive in early seventeenth century Bulgarian was restricted to a very small group of constructions, mainly after the verbs *šta* ['will'] and *može* ['can'] and the [negative] imperative *neděj(te)* [and] even with these its use was optional." This is approximately the contemporary situation, as set out in §7.7.2.1.1.1.2.2, though with further erosion of the viability of the infinitive. The trajectory of the loss of the infinitive in Bulgarian thus mirrors to a large extent what is seen in the other Balkan languages.

152 See Goina 2009 on the language of the Wallachian Charters.
153 These verbs mean 'want/will,' 'choose,' and 'approach,' respectively.
154 MacRobert 1980: 212, from document 101, dating from 1450.

7.7.2.1.1.2.2.3 Torlak Bosnian-Croatian-Montenegrin-Serbian The situation with the infinitive historically in Torlak BCMS has to be viewed against the backdrop of the rest of the BCMS complex. Thus even though BCMS in general is outside the scope of this study, for reasons outlined in §1.1 and §1.2 and noted throughout, a brief sketch is in order, as the dialectology of the infinitive within BCMS is important and quite revealing as to the Balkan character of the infinitival developments.

Most of BCMS shows some uses for the infinitive, though it is more robustly realized in the north and west of BCMS-speaking territory.[155] Usage in the standard languages reveals infinitives as subjects and as verbal complements (e.g., Browne & Alt 2004), though in colloquial usage there is evidence of encroachment of finite complements at the expense of infinitives. Stankiewicz 1986a: 207 states that "the gradual limitation of the infinitive can be observed *in vivo* even in the western parts of Serbo-Croatian territory"; following Pavlović 1960, he sees the alternations as governed by aspect and mood.[156] More recently, B. Belić 2005 argues for a syntactic basis – presence or absence in the matrix clause of a nominal that controls the complement – for the variation between infinitives and finite clauses in Belgrade usage.[157] Whatever governs the variation, infinitives are still widespread in BCMS, so that it is clear that the general absence of infinitives in Torlak reflects a loss of the category and form, such as occurs elsewhere in Balkan Slavic. Torlak thus shows convergence with Macedonian and Bulgarian in this feature – as already observed by A. Belić 1905: 473, 474 – and divergence from the rest of BCMS. At the same time, it is also worth noting here that Makartsev 2021b has noted the tendency to replace infinitives with DMS clauses in the BCMS of Fier, in Albania, whose speakers migrated there from the Sandžak (of Novi Pazar) about a century ago. He reasonably attributed this to the influence of Albanian, with which the speakers are now bilingual. The relevance of the dialectology of the loss of the infinitive in BCMS is taken up in §7.7.2.1.5.

7.7.2.1.1.2.3 Balkan Romance The starting point for Balkan Romance, including Judezmo, is the Latin of classical times, particularly Vulgar Latin, the spoken language (as attested in documents), as opposed to Classical Latin, which was the higher prestige literary form. Classical Latin had well-developed infinitival usage – as subjects of sentences, as complements to nouns, verbs, and adjectives, and in indirect discourse – and a fully productive pattern with the ending *-re* for

155 And, within South Slavic more broadly, more so in Slovene and BCMS as a whole (West South Slavic) than Bulgarian and Macedonian (East South Slavic), and within West South Slavic, more so in Slovene and Kajkavian Croatian than in the rest of BCMS. Since the break-up of the former Serbo-Croatian, the infinitive has been identified as "Croatian" by some language planners, while *da*-clauses are identified as "Serbian." Such distinctions are not based on dialectology but rather on ideology (Friedman 2007b).

156 "Imperfective forms of the verb regularly take the infinitive, whereas the perfective forms, which in South Slavic have a concomitant modal meaning, dispense with the infinitive" (p. 207).

157 This is not unlike the situation in present-day Romanian regarding "like-subject" versus "unlike-subject" conditions mentioned in §7.7.2.1.1.1.3.3.

active voice, for instance. For the most part, this situation carried over into Vulgar Latin, though there were various ways in which a retreat from infinitival use occurred in Vulgar Latin. For instance, in unlike-subject contexts between matrix clause and subordinate clause, Classical Latin employed an infinitive, as in (7.130):

(7.130) Volō tē venire
want.1SG you.ACC come.INF
'I want you to come'

In Vulgar Latin, by contrast, to judge from the absence of an infinitive in the French, Spanish, and Italian versions of (7.130) (*Je veux que tu viennes* / *quiero que vengas* / *voglio che venga*, respectively), there was instead finite complementation. Similarly, Iliescu 1968: 117 observes that finite complements headed by *quod* 'that' occur in Vulgar Latin where Classical usage would have infinitives instead, and Barić 1961 posits a tendency for purpose- and goal-based expressions to use the conditional *si* in Latin in place of infinitives. These alternatives to infinitival usage may have been the seeds of the retreat of the infinitive instantiated in the variety of Vulgar Latin that became Balkan Romance. It is no doubt significant that *quod* and *si* are the sources, respectively, of the Balkan Romance complementizer *că* and subjunctive marker *să*, which head clauses that substitute for infinitives in Balkan Romance. From this brief background, we turn to the historical developments that can be inferred for Aromanian and Meglenoromanian, and which are documented from the sixteenth century onward for Romanian; on the situation in Judezmo, see §7.7.2.1.4.

7.7.2.1.1.2.3.1 Aromanian As the discussion in §7.7.2.1.1.1.3.1 makes clear, the main candidate for infinitival status in Aromanian is the long form that corresponds to the Latin infinitives and the Romanian verbal nouns in *-re*; that is, with one quasi-exception, there are no short form infinitives to be found in Aromanian. The one possible exception is in the expression *va hi* 'perhaps,' which Sandfeld 1930: 174 treats as etymologically "(that) will be," with *hi* as the short infinitive of 'be' (from Latin *fi-[eri]*), but about which Papahagi 1974: s.v. is less certain, marking the etymology from a putative earlier **va fi* with a question mark. Even if Sandfeld is right, however, it is unlikely that any conclusions can be drawn from this one instance, especially since it is a fixed phrase.

And, as also pointed out in that section, even the long forms in *-re* are not obviously verbal, and thus not obviously infinitives in any real sense in contemporary Aromanian. Still, it is noted there that some, presumably older, speakers might feel that the long forms are verbal in nature. If these intuitions signal a real difference among speakers in terms of linguistic competence or the form the language takes for individual speakers, and if thus the variability within the speech community that these sensibilities represent are age-related, then, as apparent-time distinctions, they may offer some insight into the prehistory underlying the present-day Aromanian situation; these intuitions of older speakers can be taken as representing real-time generational differences,

analogous to age-based differences in production that are a crucial piece of variationist sociolinguistic analysis. As such, they can shed light on the history underlying the current state of affairs, and they would suggest that at the time the speakers formed these assessments of the long-form infinitives, these forms showed evidence of being real verbal forms and thus true infinitives.

These intuitions accord with statements made some 120 years ago about the long forms in Aromanian. In particular, as reported in Sandfeld 1930: 174, Weigand 1896: 142 claims to have heard an infinitive of purpose in Aromanian, *tri a lare faţa* '(in order) to wash the face,' with the long form *lare* used in a verbal function, and Papahagi 1974: 724 gives as an example under *lare* the sentence (proverb?) *Caplu di ɣumar nu va lare* 'The head of an ass does not need washing/to-be-washed' (lit., 'the.head of ass not wants/is.in.want.of washing'), where *lare* is interpretable as a verbal form.

What these intuitions and these examples might point to, then, is that the long infinitival forms in *-re* retained their verbal nature longer in Aromanian than in Romanian, and thus were real infinitives until relatively recently, changing their status to that of true nominals within the past 100 years. It is clear that they are nominal forms for all but the oldest speakers today. The fact that, quite parallel to what is found in contemporary usage of the closely related Meglenoromanian, the long infinitive shows clearly verbal behavior provides further support for this interpretation, as it would indicate that before Meglenoromanian and Aromanian diverged from one another, the common SDBR language had a fully verbal infinitive in the form of the long infinitive. Presumably, then, it was the short-form infinitive that gave way to the finite complement structures seen in such abundance in Aromanian, and exemplified in (7.104) in §7.7.2.1.1.1.3.1.

7.7.2.1.1.2.3.2 Meglenoromanian The robustness of the long-form infinitive in Meglenoromanian, as outlined in §7.7.2.1.1.1.3.2, would seem to leave little to say on the matter of diachrony, inasmuch as most of the uses appear to reflect uses found in Latin. Thus, for the most part, there is diachronic stability here rather than the diachronic change that other Balkan languages, including Aromanian, show. Still, the overall Meglenoromanian developments show some interesting aspects in terms of language contact and contact-related change.

In particular, Sandfeld 1930: 174 notes that "*en méglénite … l'infinitif est encore vivant dans quelques emplois que coincident avec ceux où l'infinitif peut être employé en bulgare*" ('in Meglenoromanian … the infinitive is still alive in some uses that coincide with those where the infinitive can be used in Bulgarian'), suggesting that contact with Balkan Slavic is at least in part responsible for the extent to which Meglenoromanian has maintained its infinitive. This claim has considerable credibility, especially since the Macedonian dialects with which Meglenoromanian would have been in contact are exactly those dialects where remnants of infinitival usage was maintained into the twentieth century, as discussed in §7.7.2.1.1.1.2. Sandfeld draws attention to the parallelism in infinitival usage with the verb meaning 'can' in both languages, and there is a further detail that is

noteworthy here. That is, one context for the long infinitive in Meglenoromanian is with the verb *trubăiiri* 'need, must' (see (7.106d) in §7.7.2.1.1.1.3.2), which was borrowed from Slavic, where it allowed infinitival syntax, demonstrably in OCS and presumably later as well. Interestingly, this verb is absent from Aromanian altogether, and while it is present in Romanian, in the form *trebuie*, it does not allow for infinitival syntax (except in the northern dialect areas of Maramureş and Crişana; cf. Farcaş 2006), occurring only with a nonfinite supine form, as noted in §7.7.2.1.1.1.3.2, or a finite *să*-clause; and, *trebuie* is an innovation within Romanian, since Early Modern Romanian used *opu e* (Lat *opus est*) with an infinitive for the expression of necessity. It seems, then, that *trubăiiri* in Meglenoromanian reflects an expanded infinitival usage under Balkan Slavic influence, with the syntax associated with the Slavic word being borrowed along with the word itself.[158]

Thus even with apparent diachronic stability with the infinitive from Latin into Meglenoromanian, there is a language contact dimension to be considered here, both to infinitival retention and to infinitival expansion.

7.7.2.1.1.2.3.3 Romanian In considering Romanian diachronically, there is more to go on than just the inferences that the Comparative Method allows for, as in the case of Aromanian or Meglenoromanian. That is, there is the direct evidence of sixteenth-century materials, which include the 1521 Letter of Neacşu of Câmpulung and various religious translations found in the Codex Voroneţean. As described in Joseph 1983a: 153–159, drawing on Meyer-Lübke 1895, those materials reveal a stage of the language in which the infinitive was in regular use, in familiar contexts such as (object) complement to modal verbs (e.g., *putea* 'can,' *cuteza* 'dare'), to predicates of necessity (e.g., *opu e* 'need there.is'), and to various adjectives (e.g., *gata* 'ready'), as subject complement, in indefinite relative structures, and in future tense formations with *voi*/*vrea* 'want' and *avea* 'have.'[159] Some representative examples are given in (7.131):[160]

(7.131) a. Aurul poate sparge cetâţi (C.B. I, 369,3)
gold can.3SG destroy.INF cities
'Gold can destroy cities'

b. Opu ĭaste tuturor gata a fĭ (Cod. Vor. 137,14)
need is.3SG all ready INFM be.INF
'Everyone must be ready'

c. Gata semu a lu ucide elu (Cod. Vor. 51,2)
ready are.1PL INFM him.ACC kill.INF him.ACC
'We are ready to kill him'

158 See §4.3.3.2 for another instance of the borrowing of the syntactic requirements of a loanword. Romanian here may have substituted the borrowed *trebuie* for inherited *opu e*, but kept the infinitival syntax of the inherited lexeme.

159 See Diaconescu 1977 for more on the history, and now also Jordan 2009 for a historical overview and an interpretation of the developments within a generative syntactic framework.

160 We follow Meyer-Lübke's abbreviatory schema for Early Modern Romanian citations: Cod. Vor. = Codex Voroneţean (Sbiera 1885), Gaster = Gaster 1891 (a 1642 text), and C.B. = Hasdeu 1880.

d. I era dragŭ a cetí la scripturi (C.B. I, 386,17)
him.DAT was.3SG dear INFM read.INF in scriptures
'To read in the scriptures was dear to him'

e. N' au ce mancá (Matt. 15,32)[161]
NEG have.3PL something eat.INF
'They have nothing to eat'

f. Vreame de voĭŭ dobandí … (Cod. Vor. 64,1)
time when want.1SG get.INF
'When I will get time, … '

g. Nu aveţi a ínnţeleage (Cod. Vor 104,4)
NEG have.2PL INFM understand.INF
'You will not understand'

In this same period, finite clauses with *se* (later *să*) occur in most of these same contexts:

(7.132) a. Nu putea se protivicască-se văntu-lui (Cod. Vor. 87,3)
NEG can.3SG DMS resist.3SG-INTR wind-the
'He cannot resist the wind'

b. Mai lesne e să nu se însoare (Matt. 19.10)[162]
more easy is.3SG DMS NEG INTR marry.3SG
'It is easy that one not marry'

c. In veac-ul ce va să fie (Gaster I,100,26)
in life-the that want.3SG DMS be.SBJV.3SG
'In the life that will be'

However, in the *avea* future construction, with the verb *cuteza* 'dare,' and in the indefinite relative construction, finite-clause variants are not found in Early Modern Romanian. Still, as in New Testament Greek – see (7.121) – and Bulgarian (see §7.7.2.1.1.2.2.2), the two competing structures could appear conjoined:

(7.133) Poate vedea şi să cunoască (Gaster I,114,7)
can.3SG see.INF and DMS recognize.SBJV.3SG
'He can see and recognize'

As the examples in (7.131) above show, short-form infinitives were present in Early Modern Romanian. The exact source of the shortening is controversial but the details are not relevant here.[163] What is important is that long infinitives could serve in complement structures, as in (7.134):

(7.134) Aadu început a zidire biserica (Gaster I,72,38)
have.3SG begun INFM build.INF church
'He has begun to build a church'

161 This example is cited from Gaster's 1890–1892 edition of the Gospel of Matthew.
162 This example comes from Gaster's 1890–1892 edition of the Gospel of Matthew.
163 See Joseph 1983a: 156–157 for some discussion.

Such syntax is parallel to what is seen in Meglenoromanian with the long infinitive, and offers a basis for thinking that the prototype for Meglenoromanian infinitival usage, even if enhanced by contact with Balkan Slavic, would have been similar.

Competition between infinitival syntax and finite syntax for complementation and other uses, as seen in Early Modern Romanian, has continued into modern Romanian, and the parameters defining the competition in earlier times – partly dialectal, partly stylistic, and partly syntactic – have continued as well throughout the transition from the earlier stages to the present, as the discussion in §7.7.2.1.1.1.3.3 makes clear. Indeed, Pană-Dindelegan 2013: 221 states that in "contemporary Romanian, the infinitive-subjunctive replacement process is far from over." It seems that the finite construction, which may be viewed as the Balkan pattern, was winning out at some point, certainly by the early nineteenth century, but one can then see "a tendency to return to infinitive structures, on the Romance pattern" and this Romance pattern "can be noted in texts showing educated usage – especially those belonging to journalistic, scientific, juridical, and administrative styles, but also in other types of scholarly texts" (Pană-Dindelegan 2013: 222). The reference to a "Romance pattern" and the occurrence of this pattern in the Romanian of educated users suggest the ideological underpinnings of the reemergence of the infinitive, with the linguistic elite looking to other Romance languages as models for Romanian usage. Indeed, that is the assessment of Close 1974, who takes the frequency of infinitival use in literary Romanian as the result of conscious imitation of foreign, especially French, models on the part of influential nineteenth-century Romanian writers, including Ion Eliade Rădulescu, Barbu Paris Mumuleanu, Iana Văcărescu, Constantin Aristia, Grigore Alexandrescu, Cezar Boliac, and Constantin Faca. While these authors were working with native material and the infinitive was still a living part of the language, this turn towards Romance models was a rejection of a perceived contact-based Balkan element in favor of syntactic expressions more in line with a puristic Romance ideology that depended on Latin as a legitimator of Romanian identity as connected with Western rather than Eastern Europe. In this way, it is a familiar topos in the recent history of the Balkan standard languages and nation-states similar to that concerning Turkish loanwords as discussed in §4.4, although in this case, the consequences went beyond lexicon and affected syntax. (Cf. Friedman 2007b on infinitival phenomena in separating Croatian from the rest of BCMS.)

7.7.2.1.1.2.4 Albanian The determination of the diachrony of the infinitive in Albanian is perhaps the most complicated of all the Balkan languages because of the lack of crucial hard evidence. Asenova 2002: 145, citing Jokl, Xhuvani 1960: 70, Gabinskij 1973: 310–323, Černjak 1973: 12, and Çabej 1977: 105–246, for instance, declares the question of whether the infinitive in *me* is Common Albanian as *otkrit* '[still] open.'[164] Nonetheless, a few choice nuggets can be found in various

164 Although Asenova mentions Jokl, she does not give a specific citation there, nor is he cited in her bibliography. The definitive collection of Jokl's works is Jokl 2018, but see also other citations of Jokl in our list of references. See also Arapi 2010 on the infinitive in Early Modern Albanian as compared with Romanian.

dialects, and with judicious use of the Comparative Method, as applied both to a comparison of Albanian dialects and to a comparison of Albanian with other evidence from around Indo-European, some reasonable inferences about the prehistory of the infinitive in Albanian can be developed.

Two things seem to be clear. First, the contemporary Tosk infinitival of the type *për të punuar* '(in order) to work' is a relatively recent innovation within Albanian. Second, the ending *-r-* (Tosk) / *-n-* (Geg) that characterizes so many Albanian participials, e.g., *lodhur*/*lodhun* 'tired,' is widely held to derive from an earlier *-(*e/o*)*no-* suffix, an Indo-European suffix that characterizes deverbal nouns and adjectives (participles) in, for instance, Germanic, as in the Gothic infinitive *bairan* 'to carry' (neuter **bheronom*, cf. Sanskrit *bharaṇam* 'act of carrying' < **bherenom*) and participle *bairans* 'having been carried' (masculine **bheronos*, cf. Sanskrit *di-na-s* 'having been divided' [root *dā-*] < **də-no-s*, *pūr-ṇa-s* 'having been filled' [root *pr̄-*] < **pl̥H$_1$-no-s*). The issues that are not clear are just how the forms in **-no-* came to be used as they are, infinitivally, and what the *për të punuar* type has replaced.

Important in this regard is the fact that Geg has a different form for its infinitive, but one still based, generally speaking, on the **-no-* verbal noun/participial. As noted in §7.7.2.1.1.1.4, the Geg infinitive uses a particle *me* with the participle, e.g., *me lodhun* 'to tire,' and this type does have some parallels in Tosk, especially the fixed expression, *do me thënë* 'that is to say,' which is basically a Gegism in Tosk. However, according to Altimari 2011b: 247, who mentions La Piana, Xhuvani, Riza, and Domi as holding this view, the infinitive of the type *me* + participle type is old and was a part of Common Albanian. According to this view, it subsequently receded from Tosk while expanding in Geg (cf. Boretzky 2014; Fiedler 2004). An alternative view, for which Altimari 2011b: 447 mentions Pedersen, Cabej, and Sh. Demiraj, is that the *me* + participle type is a northern (i.e., Geg) innovation, so that any traces of the *me* type in Tosk are due to Geg influence.[165]

What makes the former view more compelling, in Altimari's opinion, is that there are indications in the Tosk of Italy (Arbëresh) of Geg-like infinitives with *me*, both in relatively early texts, such as the eighteenth-century *Codice Chieutino* (1736–1739) by Nicolò Figlia, and in modern-day dialects north of Calabria. The textual evidence includes examples such as (Figlia, p. 51):

(7.135) a. sã dī me berë
how-much knows INFM do.PTCP
'that which he knows (how) to accomplish'

b. je për me o dashurë mire mbi gjith
are for INFM INTR want.PTCP well by all
'you exist in order to be loved by all'[166]

165 Altimari 2011b does not contain any specific citations of the authors mentioned. We can note here that nowadays contact between Geg and Tosk speakers is such that *me* + participle is also used by some Tosk speakers colloquially (VAF field notes 2008). The question of the colloquial use of the *me* infinitive among Tosk speakers in Albania and North Macedonia today is in need of further investigation.

166 *Dua mirë* is an idiom meaning 'love.'

The dialect evidence includes examples from Civita such as the following (where *pir* corresponds to standard Albanian *për* 'for'):

(7.136) a. ky shërbes ësht pir me qeshur
this service is for with laugh.PTCP
'this service is for laughing / is to be laughed at'
b. kjo vest ësht pir me blerë
this robe is for INFM buy.PTCP
'this robe is for buying / is to be bought / should be bought'

Altimari 2011b: 454 concludes from such evidence that the *me* + participle infinitive type "*devait sans doute appartenir au système verbale de l'albanais commun*" ('must undoubtedly belong to the verbal system of Common Albanian'), and he speculates that the *me* + participle type may have replaced an Indo-European infinitive. Since, as noted above, the suffix involved in the Albanian participle contains elements found in the suffix underlying the Germanic infinitive and a Sanskrit verbal noun, it is not unreasonable, following Joseph 1983a: 95, to suppose that the prehistoric Albanian infinitive was like the Germanic infinitive. The addition of the particle *me* can be likened to the process by which an introductory element, often prepositional in origin, came to co-occur with the infinitive in Germanic (West Germanic **tō*, e.g., English *to*, German *zu*; North Germanic **at*, e.g., Old Norse *at*, Danish *at*, Icelandic *að*; East Germanic (Gothic) *du*), even in the earliest attestations, given the widespread, though not obligatory, use of *du* in Gothic.[167] The type seen in (7.136b), with *për* plus the *me* + participle infinitive, would then represent further reinforcement of the infinitive not unlike what is seen in Germanic, with English *for* co-occurring with *to* + infinitive, e.g., *I'm ready* ***for to fade*** ("Mr. Tambourine Man," by Bob Dylan), perhaps originally a meaningful prepositional combination, as the examples in (7.136) suggest.[168]

The *për të* + PTCP type, also seen in Arbëresh dialects (Altimari 2011b), probably has a different history. As suggested in §7.7.2.1.1.4, the ostensible analysis of this infinitive as the preposition *për* governing a nominalized participle is almost assuredly correct from a diachronic perspective. One can find in the Buzuku missal of 1555, the earliest dated Albanian text longer than one line, constructions like *për të lutunit* (12,55) 'for the praying,' i.e., '(in order) to pray,' like a Tosk infinitive (*për të lutur*) but with nominal case and definiteness morphology on the participle, showing that it is treated like a real noun.[169] An analysis that moves away from the Early Modern Albanian formation seen in Buzuku to the

167 And in the Balkans, there is the parallel process in late Classical Greek τοῦ – formally the genitive of the definite article and taken from articular infinitive use – being generalized as an all-purpose infinitival marker in the Koine period and beyond; see Kesselring 1906 and Joseph 1983a: 49–50.

168 The comparisons offered here with Germanic are meant typologically, not to suggest a genetic connection with the developments in Albanian. However, as Hamp 2010 has pointed out, there is evidence for areal connections between Albanoid and Germanic in the Northwest Indo-European contact zone.

169 Such formations in Buzuku show nominal rection, with a genitive object as expected with a noun, not an accusative case as expected with a verb, e.g., *për të liruom të kuatëvet* (196,43–44) 'for the forgiving (*liruom*) of sins (*të kuatëvet*).'

contemporary situation where *për të* is no longer seen as a preposition plus verbal noun (PC + PTCP), but is instead an infinitival marker, would go together with a change in what is still the participle. Whether these changes are linked in some way, whether one depended on the other, or whether they were simultaneous are questions that must remain open. Under this analysis, the presumed de-nominalizing of, e.g., forms like *të lutunit* thus allowed *për të* to replace the *për me* of early historical Tosk (as seen above in (7.136)).

In this account, then, Balkan Tosk presumably gave up the *me* infinitive altogether before reconstituting an infinitive with *për* + *të*. Thus even without documentation showing the pathways by which the infinitive was lost in Tosk such as is available for Greek or Slavic, it is still possible to surmise something about the prehistory of the infinitive in Albanian by making use of the Comparative Method and of dialect materials.

7.7.2.1.1.2.5 Romani As noted in §1.2.3.5, the recorded history of Romani begins in the sixteenth century, but the materials dating from that period are rather limited for the purposes of determining anything about the grammar, as they mostly consist of short lists of words and phrases. However, given that Romani is an Indic language, there is actually much that can be known about its prehistory, thanks to the abundance of information about earlier stages of Indic. It is from such information that the situation with the infinitive prior to the stage referred to as Early Romani (see §7.7.2.1.1.1.5) – the ancestor language that developed in Greek-speaking territory – can be glimpsed. In presenting this picture of Romani prehistory, we draw on the excellent discussion in Beníšek 2010.

Beníšek emphasizes two key facts. The first is that that the Old Indic (OIA) evidence (Vedic Sanskrit and later Classical Sanskrit) as well as that of Middle Indic (MIA), shows that Indic from earliest times had an infinitive, which he defines as a verbal noun that did not undergo nominal inflection and was used in complementation. The second is that Romani underwent convergence with Greek, starting from the earliest contact in Asia Minor and continuing into the period when Romani speakers moved into Southeastern Europe and came into contact with other Balkan languages. That period of convergence led to the situation seen in the present-day Romani dialects both in and beyond the Balkans, namely that there is no native infinitive and finite complementation is the norm (see §7.7.2.1.1.1.5).[170] Putting these two facts together, and with judicious use of the Comparative Method working with data from MIA and Modern Indic (NIA) to show that a new verbal noun with the form **-ibo* arose in Romani based on an OIA gerundive in *-tavya*, Beníšek 2010: 65–66 argues that:

> it cannot be denied that a MIA forerunner of Romani did use an infinitive or a fossilised case form inherited from OIA in complementation. However, as soon as

170 As noted in §7.7.2.1.1.1.5, non-Balkan Romani dialects with infinitives have innovated these forms, reconstituting the category generally under influence from co-territorial languages that have infinitives (Friedman & Joseph 2019; Boretzky 1996).

> the new gerundive-based verbal nouns had been introduced, the language preferred them in complementation. Before its Balkanisation, Proto-Romani must have already employed its own verbal nouns in complementation.

Under this scenario, therefore, Romani started out as a language with a true infinitive – a view espoused by others but not argued for systematically – and then, as Beníšek (p. 80) puts it, due to "intensive contact with spoken Greek in Byzantine Anatolia, being later reinforced by the development of most dialects within the Balkan Convergence Area," the finite-complementation construction developed. Thus, the use of finite complementation "in present-day Romani … cannot be an Indic legacy" (p. 80).

7.7.2.1.1.2.6 West Rumelian Turkish To judge from the comparative evidence of other dialects of Turkish and of other Turkic languages, the more restricted use of nonfinite complementation and the expanded use of finite verbs in complement structures found in West Rumelian Turkish represent an innovation. Johanson 1998: 61, in discussing the general structure of Turkic languages, states that "action nouns, participles, and converbs allow nonfinite realisations of predications as embedded clauses," noting further that "clauses based on paradigmatic verbal nouns, action nouns, are complement clauses, mostly corresponding to English *that* clauses"; moreover, "there is usually a difference between factive and non-factive action clauses," with different suffixes marking the distinction. Based on the evidence of Old Turkic (Erdal 2004: 278–322, 448–494), it is clear that the instances of verbal nominalizations functioning as subordinate clauses noted in §7.7.2.1.1.1.6 represent inherited Turkic syntax.

The innovative finite complement patterns, by contrast, are generally taken to be the result of language contact, showing the effects of influence from Slavic (although Friedman 1982c: 30 says that "it has often been observed that the syntax of Balkan Turkish and Gagauz reflects a great deal of Sl[a]v[ic] influence or has been essentially Slavicized"). He notes further that Albanian is also a likely source of influence on the Turkish of North Macedonia and Kosovo. These WRT finite optative/subjunctive forms, then, can be seen as calques on the *da* and *të* clauses of Slavic and Albanian, respectively. WRT does not innovate a DMS element to match *da* or *të*, but the modality of the DMS clauses is matched by the irrealis modality of the Turkish verbs, just as the finiteness of the Balkan Slavic and Albanian complement clauses is matched by the personal endings on the subordinate verbs; see (7.118) in §7.7.2.1.1.1.6 for examples.

7.7.2.1.1.2.7 Summary Regarding Diachrony Based on the foregoing, it is clear that all of the languages historically show movement away from the use of infinitives, however at different speeds and thus to different extents, and the situation overall for Albanian is more complex; Tosk Albanian has reconstituted an infinitival use of a deverbal nominalization similar to usages found in Macedonian and Aromanian, while Geg Albanian has maintained what may have been a Common

Albanian infinitive rather robustly. But Geg shows other ways in which it does not fit the model of the Balkan core, e.g., in having nasalized vowels and distinctive vowel length. The Tosk usage is much closer to that of Macedonian and Aromanian, as would be expected geolinguistically. Moreover, one cannot discount influence from Italian, with Italy being the main outlet to the West for Albania for centuries as well as the dominant language of Albanian modernization during the fascist period and a language of prestige, the knowledge of which was not lost during the isolation of the Hoxha regime. Thus, in a way, the potential influence of Italian on Albanian has some comparability to the manner in which the influence of French in the nineteenth century seems to have played a role in the way the infinitive developed in Romanian as opposed to Aromanian and Meglenoromanian.

The current state of affairs with contemporary varieties of the Balkan languages, moreover, as argued elsewhere herein (e.g., §5.4.1.1,iv), need not, and indeed demonstrably does not, reflect the historical state of affairs of just a couple of hundred years ago. In fact, given what is known about the historical development of infinitives in each of the languages, as sketched here, there was a greater degree of convergence regarding the absence of an infinitive across most of the Balkans in the eighteenth and nineteenth centuries. From a purely linguistic standpoint, the most striking fact about the infinitival developments in the Balkans, besides the general absence of the form and category, is that similar usages are resistant to replacement across the languages. The verbs 'can' and 'dare,' as well as the future formation, are among those contexts that retain infinitive usage in Balkan Romance, in Greek (not just Medieval Greek but also in the two dialects that still have infinitives, Southern Italy Greek [Griko and Greko] and Pontic Romeyka Greek), and in Balkan Slavic, even if they ultimately shift completely to the finite complementation pattern. This fact is arguably attributable to the modal semantics of these predicates and thus their (perhaps) "universal" preference for a like-subject pattern between main clause subject and subordinate clause subject. Like-subjects can of course be expressed with finite subordination, but inasmuch as infinitives do not require the specification of their own subject (although unlike subjects can be involved), they are conducive to like-subject contexts (see also footnote 127). Moreover, as Joseph 2019d argues (and see also §7.7.1.1 and footnote 96), these are also verbs that form a single event with the action of the syntactically complement verb, so that (a possibly somewhat cross-linguistically universal notion of) event structure may also help explain parallels in infinitival retention.

7.7.2.1.2 Nominal Complementation

It is noted in §7.7.2.1.1 that while the replacement of the infinitive by a finite verb occurred across the Balkans (see also §7.7.2.1.3), other means of infinitival replacement occur, in particular the use of fully nominal forms. That possibility is realized throughout the Balkans, and in some cases contact-induced convergence seems to be involved, albeit not necessarily always.

Thus, for example, as the infinitive was receding and innovative finite complementation was advancing in Medieval Greek, one can find examples such as those in (7.137):[171]

(7.137) a. τὸ τρέξιμο ν' ἀρχίσουν (*Erotokritos* II.1316 (seventeenth century))
the running.NTR.ACC FUT begin.3PL
'They will begin (the) running'
b. ἄρχιζαν τὸ κλάψιμον (*Theseid* Β΄ 78[2] (sixteenth century))
began.3PL the crying.NTR.ACC
'they began (the) crying'

And, Lucas 2013: 45 cites the following example from a folk song ('Charalambis'):

(7.138) ο γέρο Χαραλάμπης δεν θέλει παντρειά
the old Xaralambis NEG wants marriage.F.ACC
'The old Charalambis does not want to marry'

She offers an interesting discussion of how to translate this example and how the use of the nominal form differs from a verbal complement semantically (p. 135):

> It is interesting that in M[odern]G[reek] the performative action of 'marrying' is expressed by a nominal phrase after θέλω. Thus in the Greek folksong 'Charalambis' … [the] phrase with a present subjunctive of the verb παντρεύομαι ('I'm getting married') does not carry the same meaning [as (7.138)]. Rather ο γέρο Χαραλάμπης δεν θέλει να παντρεύεται indicates that Charalambis is pondering the ongoing wedding of someone else, whom he does not want to carry through with the act.

She sees this development as an extension into complementation of the lexical process noted by Mirambel 1966 concerning "an increase in the productivity of verbal nouns"; she cites Joseph 1983a: 44–45 on this:[172]

> One further means of replacement for the infinitive, especially in its uses as a substantivized nominal form with the definite article (the so-called "articular infinitive") was the use of derived abstract verbal nouns. In fact, the diminishing productivity of the infinitive in this use was counter-balanced by an increase in the productivity and consequently the number of abstract verbal nouns in the language. As Mirambel (1966: 175) puts it:
>
> C'est, en compensation, le développement dans le système nominal, des substantifs en *-sis* (*-sē*) et en *-ma* ou *-simo*, qui remplacent les emplois antérieurs des infinitifs "substantivés" accompagnés de l'article: *tò phileîn* "le fait d'aimer", *tò gráphein* "le fait d'écrire", *tò kápnizein* "le fait de fumer", etc. sont remplacés par les substantifs neuters *tò phílēma*, *tò grápsimo*, *tò kápnisma*.[173]

171 We are indebted to Dr. Sandra Lucas of the University of Copenhagen for bringing (7.137b) to our attention.

172 Thumb 1912: 65–66 notes that the Greek "abstract verbal nouns" in -σιμο, e.g., φέρσιμο 'behavior' or ξύσιμο 'scraping,' "often serve as a substitute for the obsolete infinitive."

173 'This is, in compensation, the development in the nominal system of substantives in *-sis* (*-sē*) and in *-ma* or *-simo*, which replace the earlier uses of 'substantivized infinitives' accompanied by the article: *tò phileîn* 'the act of loving,' *tò gráphein* 'the act of writing,' *tò kápnizein* 'the act of smoking,' etc. were replaced by the neuter substantives *tò phílēma*, *tò grápsimo*, *tò kápnisma*.'

What Mirambel is drawing attention to is admittedly not syntax per se, but rather more a lexical matter having to do with one way in which the productive articular infinitive's function is taken over by different derivational patterns more purely nominal in nature.[174] Still, it appears to be one way the language has utilized the nominal system to balance out developments in the verbal system.

Balkan Slavic also shows the verbal noun in infinitive-like uses. Joseph 1983a: 117 gives example (7.139a) of the use of the Macedonian deverbal noun in *-nje* in what amounts to complementation, and Čašule 1989 offers many more such complement-like examples, as well as other infinitival uses such as purpose expressions, as in (7.139bc):

(7.139) a. Ne treba sedenje
NEG must sit.VBLN
'One should not sit here'

b. Sofija izleze stokmena za beganje (p. 279)
Sofia went-out ready.F for escape.VBLN
'Sofia went out ready to escape'

c. Odat na plivanje (p. 280)
go.3PL to swim.VBLN
'They are going swimming'

Čašule (p. 289) sees this "more intensive varied, wide use of the V[erbal]N[oun]" as the noun "becom[ing] a cardinal non-finite means, which in some functions assumes the 'responsibilities' of the infinitive." There is interplay between verbal nouns and infinitives, as reflected in the fact that Indo-European infinitives are all, etymologically, deverbal nouns (see footnote 176). In this regard, note also that Stankiewicz 1986a: 210 says that the Bulgarian *voda za piti* 'water for drinking' construction (see §7.7.2.1.1.1.2.2) might involve infinitives "reinterpreted as verbal nouns," while Joseph 1983a: 125 suggested that *piti* might be influenced by an old verbal noun, *pitьe*.[175] Furthermore, there are some verb-like distinctions, in terms of aspect, that are filled by deverbal nominals in the 'upon (lit., 'one')' construction in Aromanian and southwestern Macedonian discussed in §6.2.2.2.3 and the examples cited there. Such back-and-forth between infinitives and verbal nominalizations, whether one is reinterpreted as the other or not, echoes the historical origins of Indo-European infinitives, for across the whole family, the oldest layers of infinitives reflect case forms of deverbal nouns that have come to be embedded in the verbal system.[176]

The use of verbal nominalizations with 'one,' though not complementation per se, deserves mention here too as it is a type of subordination and involves several of the languages. In particular, as noted in §6.2.2.2.3, in Aromanian, Macedonian, and

174 Though not strictly speaking Balkan Greek, it can be noted that deverbal nominals in complement functions are quite common in Pontic Romeyka, even with its infinitives; see Sitaridou 2014ab.

175 The verbal noun analysis can only work for the small number of verbal nouns in *-t-*; most verbal nouns end in *-n-*, e.g., Blg *klane* 'slaughtering.VBLN' but *vol za klati* 'an ox for slaughtering.'

176 This is especially evident in Vedic Sanskrit, with its seventeen infinitives from datives, locatives, accusatives, etc. of various nominal formations derived from verbal roots, but the same is true for infinitives in all of the languages.

Albanian, the numeral 'one' can be used together with the verbal noun in a temporal sense, meaning 'once/as soon as X VERB-ed' or 'upon VERB-ing.' Sandfeld 1930: 123 gives examples such as the following:[177]

(7.140) a. edno idenje od oranje mažot (Mac)
one go.VBLN from plowing man.DEF
'once/when her husband had returned from plowing … '

b. unã intrare ãn biericã… (Aro)
one enter.VBLN into church
as soon as he had entered into the church … '

c. Një të dalë çifuti edhe iku (Alb)
one PC go.out.PTCP Jew.DEF and left.AOR.3SG
'Once/When the Jew had gone out, he too left'

A similar construction with 'one', but with a gerund, occurs in Skopje Džambaz Romani:

(7.141) Gova vreme, ek phure sovi,
that time one old.PL sleeping.GRD
mora te uštol (Rmi, Džambaz; Bodnárová:2018b)
must DMS arise.3SG.PRS
'In those days, while the old folks were (still) sleeping,
she [the new bride in the household] had to get up'

The parallelism in form and function in these languages virtually guarantees a contact relationship for such constructions; Sandfeld is unsure as to which language is the source of this syntagm, but Markovikj 2007: 165 offers a clear explanation of its origin and spread. He sees it as originating among Aromanian speakers in contact with Macedonian. Since Aromanian does not have superordinate aspect, whereas Macedonian does, he argues that this usage is a means Aromanian speakers innovated by which Macedonian perfective aspect in such temporal expressions could be rendered in their system. Thus the use of 'one' + verbal noun contextually perfectivizes the action being discussed, making reference to a single completed act. This use, he claims, was then calqued (back) into Macedonian, possibly by Macedonian speakers in contact with Aromanians or, more likely, it was transferred into Macedonian by Aromanians using or shifting to Macedonian, from which it spread to more general use in the language.[178] A similar process would have led to its appearance in Albanian, also a language without the superordinate aspect found in Macedonian. The Romani adaptation is undoubtedly also from Macedonian influence, although the use of the gerund in *-i,* which is an archaism otherwise not recorded in the western Balkans (Matras 2002: 160), while not being a verbal noun *sensu stricto*, given the occurrence with *ek* 'one,' is clearly part of the same phenomenon.

177 In this section, we have updated all spellings to current norms.

178 Note that in Macedonian, deverbal nouns are always formed from imperfective verbs, so that a way of expressing perfectivity with the verbal nouns is a convenient addition to the Macedonian repertoire.

Related to this, most likely, is the use of 'one' with a verbal noun in both a dependent clause and in a main clause, without a finite verb at all. Sobolev et al. 2018 give examples of this from Albanian and Aromanian (see also Sobolev 2005a: 73–74):[179]

(7.142) a. Një të thënë burri, një të vajtur. (Alb)
one PC say.PTCP man.DEF one PC leave.PTCP
'After having said (this), the man immediately left'

b. Unã videari, unã irutipsiri. (Aro)
one see.VBLN one love.VBLN
'Fell in love at the first sight'

To the extent that the 'once' construction of (7.142) is a contact-based phenomenon, this multiple use of a nominalization must surely also have a contact origin in one of the languages.

A similarly related usage involves the preposition 'with' plus a verbal noun to give a temporal expression meaning 'when' or 'upon.' Sandfeld 1930: 123–124 notes this as occurring in Albanian and (dialectal) Greek, as in (7.143ab) and in Aromanian along with 'one' as above, in (7.143c):

(7.143) a. me të dëgjuar zën e bilbilit (Alb)
with PC hear.PTCP voice.DEF.ACC PC nightingale.DEF.GEN
'upon hearing the voice of the nightingale … '

b. με το πααίνωμά του σε μίαμ μειάλη βρύση (Grk)
with the go.VBLN his to a big fountain
'upon his arriving at a great fountain … '

c. Cu unã mãngare, crescurã trei coarne (Aro)
with one eat.VBLN grew three horns
'Once he had eaten, three horns grew'

As Sandfeld points out, Macedonian here is rather like Aromanian, using the construction with 'one' given above but with the preposition *od* 'from,' thus:

(7.144) od edno sleguvanje od kočijata …
from one descend.VBLN from car.DEF
'once she had descended from the carriage … '

These uses are not necessarily replacements for older infinitival constructions, though there is an infinitival construction in Greek that is functionally parallel (see §7.7.2.3.1) and one can note the Aromanian verbal nouns are of infinitival origin. Still, these examples show that the verbal nouns can function in ways that mirror infinitival uses in other languages, e.g., Western Romance (e.g., Spanish *al salir* 'on leaving') or German (e.g., *beim Schreiben* 'upon writing').

179 We have adapted Sobolev's transcriptions to the system used herein.

7.7.2.1.3 Finite Complementation

As noted in previous sections, infinitives in the Balkan languages generally gave way to fully finite forms, i.e., forms marked for the person and number of their subject. And, as previously noted, this mode of replacement is significant since there are other conceivable ways in which speakers of a language could compensate for the loss of an infinitive, such as outright nominalizations (see §7.7.2.1.2). But the parallelism seen with the use of finite replacements for the infinitive goes deeper than just the presence of finiteness in complement clauses. As discussed by Sandfeld 1930: 175 and as elaborated in §7.7.2.1.3.1, there are actually two types of finite replacement for the infinitive, indicative and modal, and all of the languages show this bipartite replacement strategy. Thus there are actually three dimensions to the convergence here: the loss of infinitive itself, the replacement of the infinitive by finite verbal means, and the bipartite nature of that finite means.

7.7.2.1.3.1 Indicative versus Modal Complementation

The indicative/modal distinction introduced above requires some further elaboration. On the one hand, through it, as Sandfeld 1930: 175 puts it, "*on distingue nettement entre les propositions à sens plus ou moins final et celles qui n'ont pas ce sens*,"[180] a distinction echoed in Schaller 1975: 102, who uses "*Finalsatz*" ('final sentence') to refer to the modal type (cf. Asenova 2002: 166 and sources cited therein). Alternatively, one could say it pertains to realis and irrealis uses of the earlier infinitive, so that it is essentially the same as the distinction made for Ancient Greek between "declarative" infinitives and "prospective" infinitives (see, e.g., Jannaris 1897: 569, 571). As Rijksbaron 2007: 98 puts it, the declarative infinitive "represents a statement or thought of the subject of the main verb concerning some state of affairs in the 'real' world"; the prospective infinitive, by contrast, represents unrealized states, events that will happen or which are part of an as-yet nonoccurrent state of affairs. In a sense, then, the distinction is between complementation that is associated with complements expressing propositions that can have a truth value, so-called veridical, i.e., indicative, complementation, on the one hand, and complementation (mutatis mutandis, prospective infinitives and their replacements) that is associated with nonveridical modality, i.e., modal complementation, on the other.

These modalities correspond to different uses of infinitives in earlier stages of the languages in question and different types of finite replacements for those infinitives as part of the overall decline of the infinitive described in detail in §7.7.2.1.1. Importantly, in all of the languages, except Judezmo, which entered the Balkans after these changes had taken place, and Turkish,

180 "One can distinguish clearly between propositions with a more or less final [modal] meaning and those that do not have this meaning [indicative]."

Table 7.1 *Balkan indicative versus modal complement marking*

	Indicative (COMP)	Modal (DMS)
Alb	se (që)	të
Aro	cã (trã)	s(i)
Megl	că	si/să
Rmn	că	să
Blg	če	da
Mac	deka (oti, što)	da
Trlk	što ~ da	da
Grk	ότι (πως, που)	να
Rmi	kaj (oti, či [Bugurdži], etc.)	te
WRT[184]	ki, (se)	OPT
Jud	ke	ke

which does not have native subordinating conjunctions, there are different subordinating elements that accompany the finite complement and are associated with the different modalities.[181] The forms in question are shown in Table 7.1.[182]

Examples of these subordinators, showing the indicative/modal distinction, are given in (7.145) from a sampling of the languages, where the (i) sentences are indicatives and the (ii) sentences are modals; the subordinators and verbs are italicized:[183]

(7.145) a. i. Vjarvam *če* našite *šte* *spečeljat* (Blg)
believe-PRS.1SG COMP ours.DEF FUT win.PFV.PRS. 3PL
'We believe that our (team) will win'

ii. Iskam našite *da* *spečeljat*
want.1SG ours.DEF DMS win.PFV.PRS.3PL
'I want our (team) to win'

181 A controversial aspect of the analysis of the DMS is the claim that it can be analyzed as a mood marker, possibly a true affix and thus a piece of inflectional morphology, rather than as a complementizer. In this regard, Lavidas & Drachman 2012: footnote 21 review different positions taken on this issue for Greek. Here the issue of insubordination is also relevant (§6.2.4.2.8, cf. also §6.2.4.3.2).

182 The parenthetical forms are other indicative subordinators; Romani and Macedonian variants are borrowed, the others are native. See §4.2.1.3 and §4.3.3.4 for more on borrowing of subordinators as well as §6.2.4.3.1.2 on the use of the Turkish optative after Rmi *te*. In the case of Torlak, the usual BCMS variation also occurs (cf. Gołąb 1964b; Higgenbotham 1976).

183 Some of the languages, of course, as discussed in §7.7.2.1.1, allow infinitival complementation to different degrees in the modal cases, Geg Albanian being the most liberal in this regard.

184 Although Turkish usually has converb constructions, it also has COMP *ki* (a borrowing from Persian; cf. Kakuk 1960), and in WRT this is sometimes replaced with Alb *se,* e.g., in Gostivar (Jašar-Nasteva 1970).

b. i. Εμείς πιστεύουμε *ότι θα νικήσει* η ομάδα μας (Grk)
we.NOM believe.1PL COMP FUT win.3SG the.NOM team.NOM our
'We believe that our team will win'

ii. Εμείς θέλουμε *να συζητήσουμε* την απόφασή σας
we.NOM want.1PL DMS discuss.1PL the.ACC decision.ACC your
'We want to discuss your decision'

c. i. Mislam *deka* Petar *e* pameten (Mac)
think.1SG COMP Petar is smart
'I think that Petar is smart'

ii. Nie planirame *da odime*
we.NOM plan.1PL DMS go.1PL
'We are planning to go'

d. i. Mi-a spus *că e* supărat (Rmn)
me.DAT-has.3SG told COMP is.3SG angry
'He told me that he was angry'

ii. Vreau *să slabesc*
want.1sg DMS get.thin.1SG
'I want to lose weight'

e. i. Nj-spusirã *cã* tini *murishi* (Aro)
me-told.3PL COMP you died.3SG
'They told me that you had died'

ii. Acãtsarãm *s- cãntãm*
began.1PL DMS sing.1PL
'We began to sing'

f. i. Plaka i tha *se kishte*
old.woman.DEF him.DAT said.3SG COMP had.3SG
bâ gadi gjithshka (Geg Alb, Camaj 1984:247)
made.PTCP ready all
'The old woman said that she had prepared everything'

ii. Mundem *ta mbaronj* (Tosk Alb)
can.1SG.NACT DMS.it.ACC make.1SG
'I can make it'

g. i. Besojnë *se* këto problemë nuk *zgjidhen* lehtë
believe.3PL COMP these problems NEG solve.NACT.3PL easy
'They believe that these problems cannot be solved easily'

ii. Do të vazhdojmë *të ulemi* atje
FUT DMS continue.1PL DMS sit.NACT.1PL here
'We will continue to sit here'

h. i. Džanav *kaj ka den* man maro kiralea (Rmi; Jusuf 1984:32-33)
know.1SG that FUT give.2PL me bread cheese.INS
'I know that you will give me bread with cheese'

ii. Mangava *te džav* khere
want.1SG DMS go.1SG homeward/at.home
'I want to go home'

As these examples indicate, these complement clauses mostly fill object argument slots. Greek is unique in the Balkans in that these finite clauses cannot in

themselves fill initial subject argument slots, as (7.146) shows, with Macedonian and Romanian providing the contrast.[185]

(7.146) a. i. *ότι είναι ένοχος είναι φανερό (Grk)
COMP is guilty is clear.N
ii. Deka e vinoven e jasno (Mac)
COMP is guilty is clear
iii. Că este vinovat este clar. (Rmn)
COMP is guilty is clear
'That he is guilty is clear'

b. i. *να μιλάει ελληνικά κανείς είναι εύκολο (Grk)
DMS speak.3SG Greek someone.NOM is easy.N
ii. Da zboruvaš grčki e lesno (Mac)
DMS speak.2SG Greek is easy.N
'For one/you to speak Greek is easy'

For Greek, some further adjustment is needed if an indicative or a modal clause is to serve as a preverbal subject; in particular, such a clause must be nominalized via use of the definite article, as in (7.147ab), to be compared with the Greek in (7.146ab). This option is not required or available in any of the other languages; note the contrast with the acceptable Albanian and Romanian in (7.147cd) and (7.147ef), respectively.[186] In all of the languages, these clauses can be extraposed, occurring post-verbally; some examples are given in (7.148; note that in 7.148f, 3SG 'is' cannot begin a sentence in Macedonian):

(7.147) a. To ότι είναι ένοχος είναι φανερό (Grk)
DEF.ART.NTR.SG COMP is guilty is clear.NTR
'That he is guilty is clear'

b. To να μιλάει ελληνικά κανείς είναι εύκολο (Grk)
DEF.ART.NTR.SG DMS speak.3SG Greek someone.NOM is easy.NTR
'For one to speak Greek is easy'

c. Se/Që është fajtor është e qartë (Alb)
COMP is guilty is PC clear
'That he is guilty is clear'

d. Të qeshin është e vështirë (Alb)
DMS laugh.3PL be.3SG PC difficult
'For them to laugh is difficult'

e. Că este vinovat este clar. (Rmn)
COMP is guilty is clear
'That he is guilty is clear'

185 The Geg infinitive also offers the possibility of a subject complement occurring preverbally in sentence-initial position, as in:

i. Me i ikë rrezikut nuk âsht ligështí por urti
INFM it.DAT leave risk.DEF.DAT NEG be.3SG cowardice but prudence
'To avoid danger is not cowardice but prudence' (Camaj 1984: 247)

186 While one might think in principle that the languages with postpositive definite articles might allow the article to be associated with the clause and positionable with respect to a clause (e.g., after the first inflectable word of the clause), such is not the case.

f. Să parcăm aici ar fi ilegal. (Rmn)
DMS park.1PL here COND be illegal
'For us to park here would be illegal'

(7.148) a. i. Είναι φανερό ότι είναι ένοχος
is clear.N COMP is guilty (Grk)
'It is clear that he is guilty'

ii. Είναι ευκολό να μιλάει κανείς ελληνικά (Grk)
is easy.N DMS speak.3SG someone.NOM Greek
'For one to speak Greek is easy'

b. i. Është e qartë se/që është fajtor (Alb)
is PC clear COMP is guilty
'It is clear that he is guilty'

ii. Është e vështirë të qeshin (Alb)
is PC difficult DMS laugh.3PL
'It is difficult for them to laugh'

c. Jasno e deka e vinoven (Mac)
clear is COMP is guilty
'That he is guilty is clear'

d. i. Este clar că este vinovat (Rmn)
is clear COMP is guilty
'That he is guilty is clear'

ii. Ar fi illegal să parcăm aici. (Rmn)
COND be illegal DMS park.1PL here
'It would be illegal for us to park here'

The use of the free-standing definite article in Greek, as in (7.147ab), is itself not shared with languages that have postposed definite articles (Albanian, Balkan Romance, Balkan Slavic), but somewhat analogous constructions are possible with an initial emphatic demonstrative pronoun and a complement clause that is rather like an appositive, thus giving a demonstrative pronoun + COMP structure:

(7.149) a. Toa, deka e dobar go znaeme (Mac)
that COMP is good it know.1PL
'That he is good, we know'

b. tova, če … ima naučni dokazatelstva (Blg)
that COMP … has.3SG scientific proof
'That … [e.g., X is true] has scientific proof'

c. atë që po mendoj unë është … (Alb)
that COMP PROG think.1SG I.NOM is …
'That which I am thinking is … '

d. Asta, că este vinovat, ştim sigur (Rmn)
this COMP is.3SG guilty know.1PL surely
'We surely know this, that he is guilty.'

As noted, the definite article nominalization of Greek is unique in the Balkans, although it resembles the Turkish strategy insofar as both languages require nominalizations in this context. The Greek construction, however, is native,

continuing – and extending – the Ancient Greek articular infinitive nominalization. Given the prevalence of extraposition in Indo-European languages (as in Romance and Germanic languages especially), there is no reason to consider a contact-based explanation for this Balkan structural parallel.

The Balkan languages do not allow subordinate clause elements to occur to the left of COMP without some sort of additional marking, while the DMS markers, being analyzable as inflectional elements rather than clause-delimitors per se, do allow such "leakage"; the examples in (7.150) show this pattern, for instance, in Greek (a) and Macedonian (b):

(7.150) a. i. *Εμείς πιστεύουμε η ομάδα μας ότι νίκησε
we.NOM believe.1PL the.NOM team.NOM our COMP won.AOR.3SG
'We believe that our team won.'

ii. Εμείς ελπίζουμε η ομάδα μας να νικήσει
we.NOM hope.1PL the.NOM team.NOM our DMS win.PFV.3SG
'We hope that our team will win.'

b. i. *Našive deka pobedija [go] znaeme
ours.DEF COMP won.3PL it know.1PL
'We know our team won.'

ii. Se nadevame našive da pobedat
INTR hope.1PL ours.DEF DMS win.3PL
'We hope that our team will win.'

Thus the difference between indicative and modal complementation is not just a semantic distinction but it correlates too with syntactic differences in the nature of the subordinator associated with the complement clause; other differences are presented in the next section (§7.7.2.1.3.2).[187]

7.7.2.1.3.2 Subordinate Tense-Mood-Aspect

The possibility of using finite verbs in complement clauses raises the issue of how verbal inflection is realized in such clauses, since finite verbs are typically marked for – or involved in systems of marking for – tense and mood and various sorts of aspect. The morphosyntactic dimensions of inflection are covered in §§6.2.1–6.2.4, and the more syntactic angle is explored here, looking at how inflection is affected in the particular circumstance of a subordinate verb that is governed or licensed by some particular combination of main verb or complementizer(-like) element. There are similarities and differences to be noted among the languages, and contact is at most a marginal phenomenon.

7.7.2.1.3.2.1 Subordinate Tense As far as indicative complements are concerned, there are no tense restrictions on what sorts of verbs can occur; the full range of verb forms that can occur in main clauses can also occur in Balkan indicative complements introduced by the indicative complementizer (Alb *se,* Aro *cã,* Rmn/Megl *că,*

187 See Bara et al. 2005: 292 on Aromanian subordinate constructions with a zero complementizer.

Blg *če*, Mac *deka, oti*, Grk ότι, Rmi *kaj* (also Trk *ki*)): present, future, and past, including perfect and pluperfect.[188] Thus for example, English *We believe that our team is winning, will win, won, would have won, has won,* etc. can be translated into all the Balkan languages with the indicative complementizer and appropriate tense forms (present, future, past, conditional, etc.), with the proviso that in Balkan languages with evidential distinctions, additional nuances can be introduced grammatically rather than lexically, and some languages differ in which tense is expected in a subordinate clause, depending on the main verb. A complete listing of every possibility in every Balkan language is beyond what is needed here.

Given that indicative complements express propositions grounded in realia (or imagined realia in the case of conditionals and futures) and that the controlling main verbs simply allow those propositions to be expressed, this flexibility is unsurprising. Thus any convergence among the languages in regard to tense in these indicative complements is best taken to be a matter of the expressive demands that Balkan languages show as human languages per se, and not a matter where language contact has played a role.

The case is different, however, with modal complements in the Balkan languages, in that there are restrictions on tense in modal subordinate clauses and some developments seem to have been affected by contact between languages. Albanian, Aromanian, Greek, and Romani allow any verbal tense to occur with the DMS when a given semantic intent for a complement demands it.[189] Bulgarian and Romanian do not permit the aorist after the DMS (except, etymologically, in Blg *da ne bi* – cf. 6.2.4.2.8 and footnote 119), but in Macedonian DMS + aorist occurs after *kako da* 'as if.'

The situation with complement-clause tense in Albanian also differs from other Balkan languages in its rule of sequence of tenses (see also §6.2.1.2). The Albanian imperfect tense, marked with the DMS *të*, in the standard and the dialects on which it is based, is expected in a modal complement to a main verb that is past tense, as in (7.151):

(7.151) a. Desha të shkoja / *të shkoj[190]
wanted.1SG.PST DMS go.1SG.IMPF go.1SG.PRS
'I wanted to go'

b. Dua të shkoj / *të shkoja
want.1SG.PRS DMS go.1SG.PRS go.1SG.IMPF
'I want to go'

188 For those languages with evidential distinctions, the semantics of the main verb can influence the choice of verb forms in subordinate clauses (cf. Friedman 2014b: 38).

189 The specifics of these demands differ from language to language. See Mozer 2007 for interesting examples of this flexibility in Greek; for instance, even though complements of 'hope' (ελπίζω) usually look to the future, she notes that it is possible to use the aorist past tense with να when one wants good news about an event known to have occurred in the past, as in ελπίζω η αίτηση να υποβλήθηκε εμπρόθεσμα 'I.hope the application that it.was.submitted on.time.' See §6.2.4.2.8 on modal aorists.

190 A form of (7.151a) with the present subjunctive (*Desha të shkoj*), parallel to what is found in the other Balkan languages, is possible colloquially (Newmark et al. 1982: 80).

This sequence-of-tense phenomenon is not found in other Balkan languages, and so is interesting from the perspective of differences among the languages.[191] Worth noting here is that some Aromanian dialects pattern like Albanian, and Sandfeld 1930: 117–118 notes the following Aromanian example:

(7.152)	Cu	vruta	nu	puteam	s	me	aduneam
	with	beloved	NEG	was.able.1SG.PST	DMS	REFL	met.1SG.PST
	'I was unable to meet with my sweetheart'						

In (7.152), there is a past tense modal complement embedded under a past tense main verb using the DMS; Sandfeld 1930: 117–118 explains it as "*le procédé de l'albanais*" ('the Albanian process') for sequence-of-tense manifesting itself in Aromanian. Calquing of the Albanian pattern by Aromanian speakers would seem to be the mechanism for the transfer of this pattern across the languages.

7.7.2.1.3.2.2 Subordinate Mood In all of the Balkan languages, modal complement clauses, besides showing verb forms that occur freely anywhere, also show some verb forms that either cannot occur, or do not freely occur, in main clauses. Each language realizes this restriction of verb forms to subordinate clauses in a different way, so that subordinate mood shows some analytic complications that subordinate tense and subordinate aspect do not share. The restrictions are at least superficially similar to the occurrence of special "subjunctive" verb forms in languages like French and German that are associated with subordination, such that a consideration of subordinate mood in the Balkans has the potential to be an important part of complementation in general terms. However, it is unlikely that there is anything Balkanologically significant about mood in subordinate clauses, other than that coming from the standpoint of interesting, but not contact-related, phenomena that allow for cross-language comparisons relevant for a linguistics of the Balkans, as opposed to Balkan linguistics. Still, there is some value to surveying briefly the issues, as they shed light on important facts that might not otherwise be discussed.

One issue, already mentioned in §7.7.2.1.3.1, is where the modality in modal complement clauses resides. That is, it is possible to view the DMS elements themselves as mood markers and thus as carrying the inflectional value of [MOOD]. This is especially so when there is no morphological difference between the verbs that can occur in modal subordinate clauses and those that can occur in indicative subordinate clauses, as in Greek, Balkan Slavic, and in some dialects, Romani. The distinction between long and short present tense forms in Romani is discussed in §6.2.4.1.5. Albanian and Balkan Romance have, in a limited way, mood forms that are morphologically distinct and are traditionally referred to as

191 Asenova 2002: 260 refers to an "absence of 'agreement of tenses' in the Balkan languages," but in fact all her examples have to do with taxis (see §6.2.3), rather than sequence of tenses. At issue in all the examples is the use of a simple past where standard Bulgarian would use a pluperfect of the type 'be.IMPF' + LPT.

subjunctive; these normally occur in a subordinate clause with a DMS.[192] Examples from Albanian showing this restricted distribution are given in (7.153ab), with (7.153c) showing the form found in indicative contexts, and parallel examples, in other than 2/3SG forms, where there is no difference between forms in subjunctive contexts and in indicative contexts, are given in (7.153de); the difference between (7.153abc) and (7.153de) is noteworthy and instructive:

(7.153) a. Duam të shko-sh / shko-jë në shkollë nesër.
want.1PL DMS go-SBJV.2SG go-SBJV.3SG to school tomorrow
'We want you / him to go to school tomorrow'

b. *Shko- sh / *Shko-jë në shkollë.
go-SBJV.2SG go-SBJV.3SG

c. Shko-n në shkollë.
go-IND.2/3SG
'You go / (S)he goes to school.'

d. Duam të shko-jnë / shko-ni në shkollë nesër.
want.1PL DMS go-3PL go-2PL to school tomorrow
'We want them / you(all) to go to school tomorrow'

e. Shko-jnë / shko-ni në shkollë.
go-IND.3PL go-IND.2PL to school
'They / You(all) go to school.'

In these languages, one could view [MOOD] as realized inflectionally on the verb itself, though that realization would be vacuous (zero) for most cells within the verbal paradigm.

In Greek and Balkan Slavic, there are no paradigmatic cells in which a morphological distinction is made between an indicative and a subjunctive mood, as shown in (7.154) for Greek and in (7.155) for Macedonian:

(7.154) a. θέλω να τρέχω /τρέχεις /τρέχει /τρέχουμε /τρέχετε /τρέχουν όλη τη
want.1SG DMS run.1SG / 2SG / 3SG / 1PL / 2PL / 3PL all the
μέρα
day
'I want that I / you / (s)he / we / you(all) / they run all day'

192 The Albanian native term is *lidhore* (from *lidh* 'tie, connect'), and the Romanian is *conjunctiv*. See §6.2.4 on the morphosyntax of mood in general in the Balkans, and §6.2.1.1.4 on some contact-related aspects of the Albanian subjunctive paradigm. For Albanian, formally distinct subjunctives occur only in the second and third person singular except for 'have' and 'be,' which also have distinct first person singular and third person plural forms; in Romanian distinct subjunctive endings occur only in the third person singular and plural and in the present of 'be' and 'have,' which have suppletive stems. Meglenoromanian is like Romanian in that distinct subjunctives are limited to the third person present, but this limitation also applies to 'be' and 'have.' In Aromanian there are special subjunctive forms only for the 3SP of *hiu* 'be,' *am* 'have,' and *ştiu* 'know.' In Romani, the opposition of an indicative (ending in *-a*) versus a subjunctive (ending in -Ø) occurs in all persons and with all verbs except *sijum* 'be' in the dialects that make the distinction; see §6.2.1.1.1. In colloquial usage, the Albanian DMS *të* can be omitted in future tense forms, e.g., *do shkosh* as a variant of *do të shkosh* for 'you will go,' while in Romani the use of the subjunctive after *ka* 'FUT' is normal.

b. τρέχω / τρέχεις / τρέχει / τρέχουμε / τρέχετε / τρέχουν όλη τη μέρα
run.1SG / 2SG /3SG / 1PL / 2PL / 3PL all the day
'I.am. / you.are. / (s)he.is. / we.are. / you(all).are. / they.are.running all day'

(7.155) a. sakam da trčam /trčaš /trča /trčame /trčete /trčaat cel den
want.1SG DMS run.1SG / 2SG / 3SG / 1PL / 2PL / 3PL all day
'I want that I/you/he/we/you-all/they run all day'

b. trčam /trčaš /trča /trčame /trčete /trčaat cel den
run.1SG / 2SG / 3SG / 1PL / 2PL / 3PL all day
'I.am. / you.are. / (s)he.is. / we.are. / you(all).are. / they.are.running all day'

Thus in Greek and Balkan Slavic, positing a subjunctive versus indicative distinction realized on the verb itself has no formal justification, so the DMS itself is a candidate for carrying the modal value. Moreover, there is an overt reason in Greek at least – Balkan Slavic, like the rest of Slavic, having lost the inherited Indo-European modal negator in *m-*, does not have this morphological feature – to associate [MOOD] with the DMS itself, because the combination of the DMS plus a verb is negated differently from the verb in indicative contexts. As discussed and illustrated in §7.6, the negation marker μη is used with να (and with the hortative ας, thus also a modal marker) whereas the negator in indicative contexts is δεν. That Albanian and Romani have a similarly distributed negation distinction of *mos*/*ma* respectively with DMS *të/te* and *nuk* or *s'/na*, respectively, in indicative contexts seems to support the distinction made on the basis of morphology for a special subordinate mood.[193] In Balkan Slavic, the negator *ne* comes between the DMS *da* (and hortative *neka*) and the main verb, and if there are other clitics in the phrase they all follow *ne*. Also, in Romani, the collocation *te na* (DMS NEG) is possible (see §§6.2.4.3.2, 7.4.1.2.1, 7.4.1.2.3, 7.6.3.2, and 7.6.2, as well as examples 7.87c, 7.155, and 7.263d).

There is one way, however, in which the facts of Greek and Balkan Slavic as presented above need further elaboration, namely with regard to the interaction of subordinate mood and superordinate aspect, which is also discussed in §6.2.2.3. In particular, the verbs given in (7.154) are imperfective aspect, and for imperfectives, the verb form that occurs with the DMS can occur in all indicative contexts, including as a bare element in main clauses with no conjunction or particle with it. By contrast, present perfective forms in all three languages are more or less required to occur with some sort of particle or subordinator, although the details vary from language to language. Thus, for example, forms like Grk τρέξω (1SG), etc. from the verb 'run' (perfective stem τρεξ- versus imperfective stem τρεχ-), are limited to occurrence with the DMS να, FUT/COND θα, or HORT ας, as well as certain circumstantial (i.e., adverbial) or pronominal subordinators (typically indefinites).[194] Similar restrictions

193 As discussed in §7.6, the distribution is not quite as clean as presented here and may be better viewed as indicative negation versus modal negation rather than indicative versus subjunctive per se. But in the context of a concern for subordinate mood, the statement here is useful.

194 This fact led Joseph 2012 to liken the present perfective in Greek to the "conjunct" forms of Old Irish, which generally never occur unaccompanied by a particle or conjunction of some sort.

apply to Macedonian, but the circumstantial and pronominal subordinators cannot occur without a modal particle. Bulgarian is like Greek in terms of subordinators, but like Torlak together with the rest of BCMS, bare present perfectives are possible, albeit rare and restricted. Examples in (7.156), some of which reprise examples from §6.2.2.3, are illustrative:

(7.156) a. Θέλω να τρέξεις γρήγορα / ας τρέξουμε γρήγορα (Grk)
want.1SG DMS run.PFV.2SG. fast HORT run.PFV.1PL fast
'I want you to run fast' 'Let's run fast!'

a'. Iskam da go napraviš / neka go napravi (Blg)

a''. Sakam da go napraviš / neka go napravi (Mac)
want.PRS.1SG DMS it.ACC do.PFV.PRS.2SG / JUS it do.PFV.PRS.3SG
'I want you to do it' 'Let him do it!'

b. Θα τρέξεις γρήγορα αύριο (Grk)
FUT run.PFV.2SG fast tomorrow
'You will run fast tomorrow'

b'. Šte go napraviš utre (Blg)

b''. Kje go napraviš utre (Mac)
FUT it.ACC do.PFV.PRS.2SG tomorrow
'You will do it tomorrow.'

c. Όταν έρθεις, φέρε τον φαιλόνη (= (6.82)) (Grk)
when come.PFV.PRS.2SG bring.PFV.IMPV.2SG the.ACC cloak.ACC

c'. Koga dojdeš, donesi nametaloto … (= (6.82)) (Blg)
when come.PFV.PRS.2SG bring.PFV.IMPV the.cloak …

c''. Koga kje dojdeš, donesi ja nametkata ((= 6.82)) (Mac)
when FUTcome.PFV.PRS.2SGbring.PFV.IMPV it.ACC.F the.cloak
'When you come, bring the cloak'

d. *ανάψω ένα τσιγάρο … (Grk)
light.PFV.PRS.1SG a cigarette

d.' Zapalja cigara … (cf. (6.81)) (Blg)

d.'' *Zapalam cigara … (Mac)
light.PFV.PRS.1SG cigarette …
'I'll light a cigarette … ' (iterative context, see (6.81))

Thus, one might want to claim that the forms like τρέξεις are subjunctives, carrying inflectional marking (of zero) for [MOOD], as they are usually banned from occurrence alone in main clauses, just like the Albanian and Balkan Romance overtly marked subjunctive forms (except, of course, when the latter do in fact occur independently, cf. §6.2.4.3.1.2). Here, though, for Greek, the negation evidence argues to the contrary, for perfectives not associated with the DMS (or ας) negate like indicatives, with δεν, as in (7.157). Moreover, a similar argument can be made for Macedonian perfective presents, which can occur with the negator and no other modifying element in a negative interrogative, as in (7.158) – the only "insubordinate" use for this form (cf. Kramer 1986):

(7.157) a. Δεν θα τρέξεις αύριο / *μη θα τρέξεις / *θα μη τρέξεις
NEG FUT run.PFV.2SG tomorrow MNEG MNEG
'You will not run tomorrow'

b. Όταν δεν τρέξεις γρήγορα, όλα δεν πάνε καλά / *όταν μη τρέξεις …
when NEG run.PFV.2SG fast all.N.PL NEG go.3 PL well
'When you do not run fast, everything does not go well'

(7.158) Zošto ne sedneš?
why NEG sit.PFV.PRS.2SG
'Why don't you sit?'

Thus even though present perfectives in Greek and Balkan Slavic resemble subjunctives in their distribution, being restricted (or, in Bulgarian, mostly restricted) to subordinate contexts or obligatory cooccurrence with a supporting element of some sort (e.g., FUT θα*, kje, šte* or, for Greek and Bulgarian, subordinating conjunctions like όταν*, koga* 'when'), it is clear that they are not carriers of inflection for [MOOD] in themselves, but rather that the DMS να*, da* is the determiner of subjunctive mood and is thus a mood marker.[195] Greek therefore, in a certain sense, has a special subordinate mood, in the form of the clauses headed by the DMS να (or ας), and in the same sense, Macedonian does, too, with the eight subordinators identified by Kramer 1986 (see §6.2.2.3). For Bulgarian, the situation is more like that in Albanian and Romanian insofar as independent occurrence of perfective presents is a possibility, albeit highly restricted. A basic difference between Greek and Balkan Slavic on the one hand, and Albanian and Balkan Romance on the other, is the fact that the latter have morphologically distinct subjunctives of the auxiliaries 'be' and 'have' that can form analytic past tenses, whereas in Greek and Balkan Slavic, there being no morphologically distinct subjunctives as such, subjunctivity in the equivalent expressions is carried entirely by the DMS or other appropriate particle (see §6.2.2.3, §6.2.4).

The fact that Greek present perfective forms need to co-occur with some supporting element is shared with Macedonian and, to a large extent, Bulgarian.[196] This striking convergence is just one of several ways in which details of the verbal system of Macedonian and, in some respects Bulgarian, match that of Greek. Another is the placement of weak object pronouns with respect to the verb, a feature shared by Greek, Macedonian, Albanian, and Balkan Romance, and, again in some respects but not entirely, with Bulgarian. Whether the aspectual parallels are due to language contact is not clear, but it is the case that an Ancient Greek form equivalent to a Modern Greek present perfective form, namely an aorist (perfective) subjunctive, e.g., 2SG τρίψῃς (root τριβ- 'rub'), a subjunctive form

195 For Greek, if να determines subjunctive mood, then examples like να υποβλήθηκε 'that it be submitted' (see footnote 189), where an aorist tense form occurs with να, would suggest that Greek has a past subjunctive as well as a present subjunctive; whether this is an acceptable consequence we consider to be a question for linguists specifically focusing on the analysis of Greek to resolve. The DMS with aorist is impossible in Bulgarian and restricted to *kako da* 'as if' in Macedonian (see §6.2.4.2.8).

196 Bulgarian and BCMS have only limited uses of bare perfective presents, but they do occur. See the discussion around (7.156–158) for a consideration of Macedonian in this regard.

morphologically distinct from an indicative (cf. present (imperfective) indicative τρίβεις, subjunctive τρίβῃς, and aorist (perfective) indicative ἔτριψας), could occur unsupported in a main clause, for instance with a deliberative value ('Should you rub?'), so that the restriction on independent present perfectives in Greek postdates Ancient Greek. This fact gives some plausibility to an assumption of contact with Slavic as playing a role in the emergence of this restriction in Greek, but more study of the use of perfective forms in Medieval Greek is needed to settle the issue.

Subordinate mood, therefore, while interesting in that it is analytically complex in the individual languages, intersecting in some with aspect, does not offer much in the way of clearly contact-induced convergent phenomena.

7.7.2.1.3.2.3 Aspect in Subordination With regard to aspect in subordinate clauses, there is relatively little to say of Balkanological interest, since aspect itself, as a grammatical category, is realized differently across the languages (see §6.2.2). However, in the two language groups that have superordinate aspect, i.e., Greek and Balkan Slavic, there is one point of similarity that is linguistically interesting in its own right, but also interesting at least from the standpoint of the comparative syntax of the Balkans and possibly also from the standpoint of comparative Balkan syntax. It is instructive, moreover, as to the value of taking account of the full range of data, both comparative and historical, that the Balkan languages offer.

In particular, in both Greek and Balkan Slavic, the modal complement embedded under the verb 'begin' must be imperfective aspect, as shown in (7.158a) and (7.158b), respectively:

(7.159) a. Počnav da kažuvam / *kažam (Mac)
began.1SG DMS say.IPFV.PRS.1SG / say.PFV.PRS.1SG
'I began to say'

b. άρχισα να μιλάω / *μιλήσω (Grk)
began.1SG DMS speak.IPFV.PRS.1SG / speak.PFV.PRS.1SG
'I began to speak'

What makes this convergent detail of subordinate aspectual syntax linguistically interesting is that one can think of the beginning of an action as a momentary event, for which perfective aspect might be considered appropriate; yet, both Greek and Balkan Slavic (and the rest of Slavic, see below) seem to view the beginning of an action as the unfolding of an event of some duration, and thus use the imperfective. And, it is noteworthy that these languages do not just allow imperfective, but require it. The historical record shows that this is not an inheritance in each branch from some constraint that could be attributed to Proto-Indo-European, because Koine Greek of the Septuagint allows, and may even favor, perfective aspect for its complement, i.e., aorist infinitives (so Lavidas & Drachman 2012: footnote 24). Thus a requirement of imperfective aspect in such complements is a development within the history of Postclassical Greek. Koine Greek of the New Testament

strongly favors imperfective aspect, since out of the ninety instances of complementation with ἄρχομαι 'begin,' eighty-nine have an imperfective complement infinitive (i.e., one based on the present stem); there are no instances of a perfective infinitive (i.e., based on the aorist stem), and a single instance of a perfect infinitive. Later on, however, this preference seems to have become a grammatical requirement, and given its possible oddness from a logical standpoint, as noted above, it is of a sort that one might think to explain in one of the languages via contact. If it were contact related, it would presumably be the result of Slavic influence on Greek, since the comparative Slavic evidence speaks against a claim that Greek influenced Slavic; in particular, in Russian, as (7.160) indicates, the same requirement for imperfective in the complement of 'begin' holds, so the constraint is not restricted within Slavic just to the Balkans:

(7.160)	Ivan	načal	pisat'	/ *napisat'	knigu.
	Ivan	started	write.IPFV.INF	/write.PFV.INF	book.ACC
	'Ivan started to write a book'				

A fuller study of the aspect of complements of ἄρχομαι in later Postclassical, and even in Byzantine and Medieval Greek, is needed to settle the issue, since Slavic influence would not have been possible before the sixth century. Should it turn out not to be a matter of contact, nonetheless, the convergence described here can stand as a useful methodological exercise in how to use the historical and the comparative record in deciding the origin of parallels among Balkan languages.

7.7.2.1.3.3 Composite Finite Subordinators

One parallel among the Balkan languages in connection with finite complementation is the occurrence of finite subordinators – subordinating conjunctions that govern a finite verb – that are composite formations, showing combinations of various independent elements, some univerbated into a single word and some simply showing co-occurrence, creating a bipartite marking for subordination. In some instances there are exact matches between languages, e.g., Albanian *megjithëse* 'although,' Aromanian *cu tute cã* 'idem,' and Greek μολονοτι 'idem,' where all three are composed of the preposition 'with' (Alb *me*, Aro *cu*, Grk με), the word for 'all' (Alb *gjithë*, Aro *tute*, Grk όλον), and the indicative complementizer (Alb *se*, Aro *cã*, Grk ότι). Or, the matching may be in terms of gross structure, as with 'before,' which consists of an adverbial for 'before, ago' (Alb *para*, Grk πριν, Mac *pred*, Blg *predi*, Rmi *anglal*) with a subordinator, though it is the indicative subordinator *se* followed by the DMS *të* in Albanian but just the modal subordinator (DMS) να, *da, te* in the other languages (Friedman 1985a). There are also some differences in regard to these composites. For instance, while Albanian *po* and Aromanian *cara* in the meaning 'if' can form bipartite subordinators, co-occurring with the DMS *të* and *s'*, respectively, the Greek equivalent (ε)άν and BSl *ako*'if' cannot co-occur with the DMS να and *da,* respectively (i.e., *(ε)άν να έρθει /*ako da dojde 'if he-comes'). However, Greek 'before' can also omit the DMS altogether

Table 7.2 *'in order that' in the Balkans*

Alb	për të
Blg	za da
Mac	za da
Grk	για να
EMR	pentru să (see also footnote 244)
Rmi	dži te

(πριν έρθει / πριν να έρθει 'before he comes') whereas the other constructions mentioned above cannot omit the DMS (and *se* for Albanian). One composite form that is widespread is the combination of the preposition 'for' with the DMS introducing purpose expressions, thus 'in order that' (see also §7.7.2.3.2) (see Table 7.2).[197]

Modern Romanian has a slightly different form of this composite subordinator, with 'for' (*pentru*) plus *ca* 'as' and the modal subordinator *să*; Aromanian differs more significantly from this Balkan pattern, having simply *ca să*, and Meglenoromanian has *că si* and other constructions (Atanasov 2002: 276).

As with so many convergent structures in the Balkans, one needs to be careful about concluding that contact is involved, as elsewhere in Indo-European one can find such composite subordination marking, e.g., Middle English has *before that*, and modern American English dialects have *because that*. The existence of forms of this sort outside of the Balkans means that such composition may simply be a natural strategy for creating, or extending, subordinators. Still, chronology is important when it can be established. In cases where there is matching of details, as with *megjithëse*/μολονότι/*cu tute cã*, it is hard to see how contact between speakers/languages could not be involved (see also §4.3.3.4).

7.7.2.1.4 Causes of Infinitival Developments

Given the convergence with regard to the developments with infinitival complementation in the Balkans outlined in the preceding sections (§§7.7.2.1.1–7.7.2.1.3), it is natural to raise the question of what their causes were. Language contact has been implicated in these developments by virtually all observers, though in different ways.

The earliest account, Leake 1814, assumed that Slavic was the driver of contact-induced change in the Balkans. Other early accounts, as well as some later ones

197 Romani uses just the native DMS *te* in this meaning, although some dialects in contact with Balkan languages facultatively use a borrowed element as well, e.g., Agía Varvára *ja te* (*ja* from Grk για; Igla 1996: 182), Bugurdži *či te* (*či* from Kosovo Geg complementizer *či*, StGeg *qi*, StAlb *që*; Boretzky 1993: 99), etc.

(e.g., Kopitar 1829; Miklosich 1862; Weigand 1925; Solta 1980) held that a prehistoric substratum was responsible. Besides being an untestable hypothesis, it runs into serious chronological problems, since the loss of the infinitive and the rise of finite complementation were accomplished over a long period of time within recorded history. Greek is the language in the Balkans with the longest documented history, and it is clear (see §7.7.2.1.1.2.1) that there was a stage of the language, namely Classical Greek, with a well-developed infinitival system. A similar observation can be made, though at later periods of time, for the respective predecessors to Balkan Slavic, namely Old Church Slavonic (see §7.7.2.1.1.2.2) and Balkan Romance, namely Vulgar Latin (see §7.7.2.1.1.2.3).

The viewpoint aired next was that of Sandfeld 1930: 177–178, and it proved to be most influential. Reacting against Weigand 1925: xi, Sandfeld proposed that Greek, as an adstratal language of considerable prestige in the Balkans, was responsible for the demise of the infinitive in all the languages. He cited the general south-to-north decreasing strength/extent of infinitive-loss in support of his hypothesis, as well as the early loss seen in Greek, beginning in the Koine era before there was contact with the other languages, and the prestige accorded to Greek as a language of learning and religion.

By 1970, however, Stankiewicz 1986a: 207 was able to write regarding the loss of the infinitive that "it is now commonly agreed that this 'Balkan' feature cannot be ascribed to the influence of only one language (such as Greek)." Rather, Stankiewicz claimed, it "was probably the result of interaction between various languages, among which a special role must have been played by Arumanian as a mediator among the Balkan vernaculars." He goes on, however, to effectively deny the role of language contact, at least for Balkan Slavic, suggesting ultimately that "the restriction and ultimate loss of the infinitive must be viewed primarily as an internal development which has gradually led to the elimination of some of its functions but which has left other functions intact, specifically the use of the infinitive with modal verbs (including the auxiliary of the future, **xъtjǫ*)." Joseph 1983a: 179–212, in essence countering the second part of Stankiewicz's stance while taking the recognition of interaction between various languages as a starting point, develops an account of causation for infinitive-loss in the Balkans and the rise of finite replacements that – without necessarily ascribing a central role to Aromanian – builds on the dynamics of contact in a multilingual setting.[198]

That account, adopted here in more or less the same form, starts with the observation that finite and nonfinite complementation co-existed in all of the languages in the earliest attested stages so that structurally different but functionally parallel variants were present in each of the languages. From such a starting point, some undeniable language-internal developments are evident, for instance the retreat of some uses of the infinitive in New Testament Greek, as sketched in

198 We say "in essence" because Joseph did not specifically cite Stankiewicz nor was he aware of the work, owing to its appearing originally in an Italian Slavistic journal in 1973. But the essence of both parts of Stankiewicz's position was precisely consistent with the approach Joseph took, in the first instance, and argued against, in the second instance.

§7.7.2.1.1.2.1 as part of the ebb and flow of the use of various verbal forms. It is significant that in each language there were finite verb forms that were, or came to be, homophonous with the forms that infinitives (eventually) took; for instance, by regular sound changes, Classical Greek λύσει 'he-will-loosen' (3SG future) and λύσῃ '(that) he-loosen' (3SG subjunctive) came in Postclassical times to both have the same pronunciation as the Classical future infinitive λύσειν 'to loosen,' and the form of the Postclassical remade aorist infinitive, all converging on [lísi]. The relevance of such mergers becomes clear below. Cf. also the homonymy of Bulgarian infinitive remnants with the 2/3SG aorist (with which, etymologically, it shares a stem).

It is known that in contact situations, analytic constructions and overt morphological marking are generally favored, as they promote communicative efficacy.[199] The same holds for the redundancy offered by multiple marking for person and number, for instance on a main verb and on a finite complement standing in for an infinitive and with full encoding of the features of the subject.[200] The ideal environment for the realization of such favoring would have been the multilingual contact environment in the west central Balkans, taking in what is now northern Greece, North Macedonia, and Albania, especially during the Ottoman period when speakers of Albanian, Aromanian, Greek, Macedonian, and Romani, i.e., Balkan Albanian, Balkan Romance, Balkan Greek, Balkan Slavic, and Balkan Indic would all have been interacting with one another, using one another's languages, most likely somewhat imperfectly, through the filter of their respective first languages. In such a milieu, the selection of subordinate-clause variants with finite forms would have been favored as speakers of the different languages with some command of the other languages would have accommodated to their interlocutors by selecting the variant that provided the greatest amount of information in itself. Moreover, the possibility that true infinitives could be (mis)interpreted as finite forms – given the convergence in form between infinitives and some finite form in each language – would have aided the selection of finite forms and would have contributed to the rise of finite complementation at the expense of the infinitive. In this way, the demise of the infinitive would have continued and would have been extended through a mix of language contact and some language-internal developments. Still, without the contact among speakers of different languages, it is hard to see how some homophony alone would have led to the loss of the infinitive to the extent experienced by the Balkan languages.

Locating the key impetus for the loss of the infinitive in the Ottoman period makes sense for several reasons. First, it is during this period that there is the most fluidity to language contact and, until the nineteenth century, a lack of both modern standard languages and the literacy that goes with them. Moreover, speakers of many different languages were interacting, so that one can thereby motivate how it is that so many of

199 Such is the case with pidgin languages, for instance, which tend to be highly analytic, or, for that matter, with many creole languages.

200 Or alternatively, in pidgin and creole languages, by multiple expression of subject pronouns.

the languages were affected. Also, the Ottoman period squares with what is known about the chronology of infinitival developments, for in Greek and in Balkan Slavic, at least, where there is the greatest depth of historical documentation, it can be seen that the complete loss of the infinitive comes only in or after the sixteenth century.

In such an environment, social factors must surely have been especially relevant in the spread of innovative finite subordination.[201] As noted in §7.7.2.1.1.2.2.3, Stankiewicz 1986a draws attention to the fact that mid-twentieth century Serbo-Croatian offers a laboratory, as it were, for observing the processes behind the replacement of the infinitive in earlier times, and indeed, social differences, e.g., regarding level of education, can be seen to play a role in modern variation in BCMS (Joseph 1983a: 131–148, referred to there as Serbo-Croatian).[202] The relevance of social factors becomes especially clear when one considers the situation with infinitives in Judezmo.

Balkan Judezmo preserves the Ibero-Romance infinitive, but there is some reduction in the use of infinitive in favor of finite complementation, in the form of subjunctive mood verbs, parallel to usage of co-territorial Balkan languages. For instance, the Judezmo use in modal questions, e.g., 'When might we come to get you?,' of the bare subjunctive by itself, with only the subordinating element *ke* and no controlling main verb, as in (7.161a), mirrors Balkan clauses, e.g., in Greek (7.161b) and Macedonian (7.161c), with a DMS only with the modal verb and similarly no controlling matrix verbal element:

(7.161) a. Kwando ke te vengamoz a tomar? (Jud)
when that you.ACC we.come.SBJV to take.INF

b. Πότε να ρθούμε να σε πάρουμε? (Grk)
when DMS we.come DMS you.ACC we.take

c. Koga da ti dojdeme da te zememe? (Mac)
when DMS you.DAT we.come DMS you.ACC we.take
'When might we come to get you?'

As a control against which to measure the "Balkan-ness" of (7.161a), one can consider (7.162), the equivalent sentence in Modern Peninsular Spanish, where a controlling verb (*quieres*) is needed to introduce the subjunctive of 'come':

(7.162) ¿Cuándo quieres que vengamos a recogerte?
when you.want that we.come to take.INF.you. ACC
'When do you want us to come to get you?'

Moreover, North African (i.e., non-Balkan) Judezmo is like Modern Peninsular Spanish in this regard, requiring a controlling verb for the subjunctive.

201 Many linguists argue that social factors are always relevant in the spread of an innovation, a position we support.

202 Given that variation involving finite and nonfinite forms is on-going, this observation still holds for present-day BCMS (see B. Belić 2005). The post-1991 break-up of the former Serbo-Croatian has seen normativistic pressures to favor the infinitive over *da*-clauses in Croatian (cf. Friedman 2007b).

As for the importance of understanding social considerations with the infinitive, based on developments in Judezmo, it must first be noted that although Judezmo aligns with its Balkan neighbors in certain uses of a finite subjunctive mood (as in (7.161)), nonetheless the earlier Ibero-Romance infinitive remains in Judezmo; Thessaloniki Judezmo in the early twentieth century, for instance, had infinitives occurring as complements to adjectives and verbs, as in (7.163):

(7.163) a. Tienes una vos mui buena para kantar
have.2SG a voice very good for sing.INF
'You-have a voice (that is) very good for singing'

b. Ke pueda fazer
what can.3SG.SBJV do.INF
'What might he be able to do?'

Furthermore, both surviving Greek Judezmo speakers of Thessaloniki, who also speak Greek natively, and surviving speakers in Skopje and Bitola, who also speak Macedonian, use infinitives, as in (7.164):[203]

(7.164) ¿Puede recontar historía? (Roza Kamhi, Skopje 2008; VAF field notes)
can.3SG tell.INF story
'Can I tell a story?'

Thus in Balkan Judezmo, we have both an innovative tendency to use finite subjunctives that calque typically Balkan constructions and a conservative retention of earlier infinitives; this mix of infinitival and noninfinitival usage in Balkan Judezmo is documented by Dobreva 2021 (see also Dobreva 2016), working with textual and secondary sources spanning the late nineteenth century into the twenty-first century. The infinitive is retained even in situations in which these speakers are (now, at least) bilingual in infinitive-less standard Balkan languages (Greek, Macedonian, or Bulgarian, as the case may be). It can be noted too that there are social associations, here of a religious nature, with noninfinitival languages: Standard Greek, Bulgarian, and Macedonian are all associated with Orthodox Christianity, despite the fact that that they all have native speakers who are Muslim (and, at this point, also Jewish).[204]

Thus the persistence of the use of infinitives in at least some Balkan Judezmo varieties most likely reflects a lesser degree of contact between Jews and non-Jews in the Balkans than among the non-Jewish speakers of various languages in the region. The fact that there is some finite subjunctive use that parallels non-Jewish

203 Note that (7.164) has impersonal active *puede* (as opposed to the ostensible reflexive form *se puede* of Modern Peninsular Spanish), on the model of Macedonian *može* 'it.can' (3SG).

204 The complexities of language and identity are beyond consideration here. Suffice it to say that all three languages have Muslim speakers. For Greek, they still identify the language as Greek (e.g., https://www.unhcr.org/en-us/news/latest/2018/2/5a8ed3314/refugee-family-renews-century-old-ties-crete.html); for Macedonian, their ethnic identity sometimes conflicts with the fact of Macedonian mother tongue; and for Bulgarian, the Rhodopian dialect of the majority of Muslim speakers is promoted as a separate Pomak language in Greece, although in Bulgaria there are also Christian speakers of the same dialects who consider their language to be Bulgarian. For historical reasons, Albanian avoided the association of language with religion (see Friedman 2016).

usage shows that there was some degree of contact at a high enough level to result in some convergence, but the relative social isolation of Judezmo speakers in Balkan cities, combined with their relatively late entrance into the Balkans, would have prevented Jewish speakers from fully converging linguistically with their non-Jewish neighbors. Relevant here too is the evidence of the nineteenth-century anecdotal tales in Cepenkov's 1972a collection of Macedonian materials: out of 155 such tales, twenty-four have code-switches into Turkish, four into Greek, three into Albanian, two into Aromanian, and one into Romani, but none into Judezmo; that is, representatives of all the other languages code-switch into the language associated with their ethnicity in these tales, but when Jews code-switch from the Macedonian matrix of the narrative, the switch is into Turkish, another indication of the marginalization of Judezmo with respect to the Balkan linguistic social hierarchy. While it is true that non-Jewish merchants in the bazaars often had some knowledge of Judezmo (Bunis 2011, cited in §4.4.3), multilingualism tended to be unidirectional (Friedman 2000d). Thus the persistence of the use of infinitives in at least some Balkan Judezmo varieties seems to be an important reflection of a lesser degree of contact between Jews and non-Jews in the Balkans than among the non-Jewish speakers of various languages in the region.

It is impossible, therefore, to separate the developments with the infinitive in Judezmo from their social circumstances in the Balkans. By extension, it can be argued that social factors involving freer contact between different groups of speakers, the antithesis of the more restricted contact involving Judezmo, must have played a role in the spread, and ultimate triumph, of innovative finite verbal usage in the various languages of mainstream Balkan society in cities, towns, and ultimately villages across the region.[205]

Like Judezmo, Romani, too, was involved in mostly one-way multilingualism. By contrast, though, Romani was in intimate contact with Greek for hundreds of years beginning no later than the eleventh century or so. Thus the absence of infinitives in Romani is understandable, from a chronological standpoint, if not a social one. And while we cannot know what Proto-Romani looked like at the point of first contact with Greek, Romani became as dependent as Greek and Macedonian on its analytic subjunctive to perform infinitival functions.[206] The overall developments with the infinitive in the Balkans, therefore, were caused by a complex interaction of particular language-internal facts with the external impetus of contact among speakers of the different languages. These developments are thus a case of multiple causation, as argued for by Joseph 1983a: chapter 7 (see also Joseph 1983c), with several different factors combining to bring about the observed changes.

205 See §3.2.2.10 on the "gravity" model of diffusion as applied to the Balkan infinitival developments.

206 Some Romani dialects outside the Balkans have begun developing new infinitival constructions in contact with European languages that do have infinitives (Friedman & Joseph 2019; Boretzky 1996), and some dialects in eastern Bulgaria have borrowed Turkish infinitives as such under the influence of Turkish (see §7.7.2.1.1.1.5).

7.7.2.1.5 The Fall of Nonfinite Subordination and Rise of Finite Subordination as Balkanisms

As §7.7.2.1.1 and §7.7.2.1.3 make clear, developments involving restrictions on the use of infinitival subordinate clauses, both complements and adjuncts, and the concomitant increase in the use of finite subordinate clauses occurring in all of the Balkan languages, are one of the most widely distributed syntactic features in the region. Moreover, the previous discussion demonstrates that there are actually three logically independent but nonetheless convergent strands of development here that are relevant: the loss of infinitive itself (taking in both infinitival forms and the morphosyntactic category of infinitive), the replacement of the infinitive by finite means of expression, and the occurrence of two distinct types of finite means of replacement, i.e., both indicative and modal. That all three of these developments are found in all the Balkan languages is telling and would seem to demand a contact-based explanation.[207] In addition, the account given in §7.7.2.1.4 for the causes of the loss of the infinitive and the spread of finite complementation depends on language contact. Thus it would seem to be uncontroversial to consider – as do scholars from Kopitar to Sandfeld and beyond, and as does every handbook of Balkan linguistics that there is (see §§2.2, 2.3, 2.4, passim) – that these features are Balkanisms.

Nonetheless, there are some countercurrents to this trend of treating these developments as Balkanisms. Hock 1991: 495–497 (see also Hock 1988), for instance, draws attention to the fact that retreat of the infinitive in favor of finite complementation occurs in certain constructions in Romance and Germanic languages. In particular, he notes that while Latin used infinitives in like-subject (cf. (7.165a)) and unlike-subject (cf. (7.165b)) constructions with verbs of volition:

(7.165) a. Volo fugire
want.1SG flee.INF
'I want to flee'

b. Volo te fugire
want.1SG you.ACC flee.INF
'I want you to flee'

in French, the infinitive is used only in the like-subject condition, with a finite subjunctive clause used in the unlike-subject condition, as in (7.166):

(7.166) a. Je veux partir
I want.1SG leave.INF
'I want to leave'

b. Je veux que tu partes
I want.1SG that you leave.2SG.SBJV
'I want you to leave'

207 Sandfeld 1930: 175 is very insistent on this point, saying "*ce qui est remarquable, ce n'est pas tant le refoulement de l'infinitif que le fait que toutes les langues en question l'ont remplacé exactement de la meme façon*" ('what is remarkable is not so much the retreat of the infinitive as the fact that all the languages in question replaced it in exactly the same manner').

A similar pattern can be found in German. For Hock, such limited retreat of the infinitive suggests that the Balkan developments must be viewed as just part of a larger European areal feature, in fact part of a larger European sprachbund (see §3.4.1.3 on this) and, therefore, that there is nothing particularly Balkan about the infinitival developments. A different sort of counter to the widely held view is given by Sh. Demiraj 1969, who argues that in Albanian, the loss of the infinitive can be accounted for on purely language-internal grounds, without any contact with speakers of other languages needed.

Nonetheless, despite such considerations, and despite some differences in the way in which, and especially the extent to which, infinitive-loss is realized in the different languages, as indicated by the facts in the preceding sections and as emphasized in Joseph 1983a (see also Joseph 1983c), there is good reason to see the replacement of the infinitive as a Balkanism. Therefore, by way of concluding this section, it is useful to bring in two types of evidence that support the interpretation of the infinitival developments as a Balkan-particular phenomenon: geography and chronology.

As for geography, for the most part, in each case, the more peripheral the language or dialect, the less likely it is to participate fully in the loss of the infinitive and its replacement by finite subordination, and the more centrally located a language or dialect is in the Balkans, the more likely it is to show infinitive loss. Particularly telling from a geographic standpoint are the comparisons in Table 7.3, where [+infinitival]

Table 7.3 *Geography of Balkan infinitive-loss*

[+infinitival]	[–infinitival]
Romeyka Greek (eastern Turkey)	mainland Greek
Southern Italy Greek	mainland Greek
Arbëresh Tosk Albanian	mainland Tosk Albanian
Geg Albanian	most of Tosk Albanian
West South Slavic (BCMS, Slovene)	East South Slavic (Mac/Blg)
Bulgarian	Macedonian
Maleševo-Pirin, Lower Vardar Mac	the rest of eastern & all of western Mac
Northern BCMS	Southern BCMS
Non-Torlak (N/W) BCMS	Torlak (Southeastern) BCMS
Western European Romance[208]	Eastern (non-Italo-)Romance
Iberian Hispanic	Balkan Judezmo
Istro-Romanian	Balkan Romance (Aro, Megl, Rmn)
Romanian	Aromanian/Meglenoromanian
Northern Romanian (Maramureş, Crişana)	Southern Romanian
East Rumelian & Anatolian Turkish	West Rumelian Turkish
Modern Indic (e.g., Hindi)	Romani
non-Balkan Romani	Balkan Romani

208 This invented term is meant to encompass all of Western Romance plus the Eastern Romance language branch known as Italian or Italo-Romance.

means that the infinitive is alive or remained alive longer in the language or dialect to some (not insignificant) degree and [-infinitival] means that there essentially is no infinitive.

The generalization emerging from such comparisons is that the more deeply embedded geographically a language or dialect is in the Balkans, the weaker its category of infinitive is; by the same token, the more peripheral the language or dialect, the more robust the infinitive is. Admittedly, there are some exceptions to this generalization, e.g., Cypriot Greek is geographically peripheral but lacks the infinitive to the same degree as mainland Greek. While there may be other forces at work in such cases that can help to explain them as exceptions to the generalization,[209] even with them, it nonetheless holds true in the vast majority of cases and thus provides support for the notion that there is something characteristically Balkan about the loss of the infinitive and its replacement by finite forms.[210]

Peripherality in Table 7.3 is geographic in nature, but there can be chronological peripherality as well, and along this dimension, too, a characteristic Balkan angle is to be recognized to the infinitival developments. That is, although the ultimate loss of the infinitive in some of the languages is late, or has not yet occurred – Romanian, for instance, preserves the infinitive as an option even in contemporary usage – it can be localized temporally in the sixteenth or seventeenth century for the languages that lack it most fully, especially Greek, Macedonian, and Tosk Albanian prior to the reconstitution of the *për të punuar* infinitival type. Thus the relative robustness of the infinitive in Balkan Judezmo, as compared to its linguistic neighbors in the Balkans, especially Greek and Macedonian, while certainly in part having a social basis, as discussed in §7.7.2.1.4, may also in part be due to chronology, a point made also by Dobreva 2021: 29, 34 (see also Joseph 2021). That is, the entry of Jews from the Iberian Peninsula into the Balkans, in the sixteenth century and afterwards, came at the end of the most intense waves of infinitive-loss in the region. One can speculate, then, that being peripheral to the temporal period most associated with strong infinitive-loss may have played a role in the survival of the Judezmo infinitive to the degree it has in contemporary usage.

The alternative to recognizing a Balkan aspect to language contact leading to the loss of the infinitive and the rise of finite subordination is to consider the parallel developments in neighboring languages as nothing more than a huge coincidence. Such a coincidentalist stance is a most unsatisfying position, though one that is neither disprovable nor provable. However, there is a control against which to measure the coincidence in that with Torlak BCMS, the divergence from the rest of BCMS comes exactly in the dialect zone that is most in contact with Macedonian

209 For instance, contact between Greeks of the mainland and Cypriot Greeks – note that the Cypriots have a separate word for mainlander Greek (καλαμαριστικά 'pen-pushers'), suggesting on-going exposure to Balkan Greek – may have helped to impose infinitivelessness on the Cypriots.

210 Sandfeld 1930: 175 makes a similar point about geography, though he took the infinitival/noninfinitival gradience to be essentially north-south within the Balkans – note that he did not utilize information about Romani or peripheral Albanian and Greek – and interpreted it as showing that infinitive-loss spread from Greek northwards.

and Bulgarian in which the infinitive is absent, or nearly so, a fact even recognized by A. Belić 1905: 473.[211]

Putting geography and chronology together and taking them into consideration along with the linguistic details of what happens to nonfinite complementation in the various languages provides some enlightenment here, as together they render even more improbable the coincidentalist interpretation. It may be concluded, therefore, that all the developments associated with the infinitive – the loss of the infinitive, the prevalence of finite subordination, and the occurrence of two types of finite subordination – constitute a Balkanism, a contact-induced convergence among the various languages.

7.7.2.2 Adjectival (Relative) Clauses

Adjectival clauses, i.e., relative clauses, serve as modifiers to nominals and are the clausal functional equivalent of lexical adjectives. The process involved in their formation can be referred to as relativization. In the Balkans, one finds both a relativization strategy involving a full clause with a complementizer and one involving a reduced clause realized with participles. Both show some developments of interest from a Balkanological standpoint.

The two strategies come together in a change seen in West Rumelian Turkish (Friedman 1982c, 2002a, 2006c) and Gagauz (Menz 1999: 84–91).[212] In these languages, use of the typical Turkish reduced-clause strategy is reduced or given up in favor of a full-clause strategy that mirrors the Slavic and Albanian type of relative clause using a subordinating conjunction. Participial and gerundial constructions are replaced by subordinate clauses introduced by question words reinterpreted as subordinators and calqued from Slavic using native material, e.g., *ne* 'what' (cf. Slv *što* 'idem'), *ani* 'which' (StTrk *hangi*), *kaçan* (older [*h*]*açan*) 'when,' as well as borrowings such as Albanian *se* 'which,' etc. Example (7.167) from WRT (Yusuf 1977: 65, cited in Friedman 2006c: 39) is given here with Standard Turkish and Macedonian translations to illustrate the convergence of WRT with Balkan Slavic, as well as the Albanian where the subordinator (*që*) is not interpretable as a question word:

(7.167) a. Cetir o çitabi ne verdım Sana (WRT)
bring.IMPV that book.ACC what gave.AOR.1SG you.DAT

b. Sana verdiğim kitabı getir (StTrk)
you.DAT give.PTCP.1SG book.ACC bring.IMPV

211 Here it is worth remembering that at the time A. Belić was writing, these dialects were spoken in Ottoman Turkey, and both Serbia and Bulgaria claimed the territory and therefore the dialects. A. Belić was involved in those territorial claims by supporting the arguments that they (and therefore the territory on which they were spoken) belonged to Serbian, but he acknowledged both Macedonian and Bulgarian influence and treated Macedonian as a language distinct from Bulgarian.

212 See also Jašar-Nasteva 1970, Matras 2003/04, 2009 and Matras & Tufan 2007 on this development in the Gostivar and Ahmed 2005 on the Ohrid-Prespa regions, as well as in other Macedonian WRT dialects.

c. Donesi ja knigata što ti ja dadov (Mac)
bring.IMPV it.ACC book.DEF what you.DAT it.ACC gave.AOR.1SG

d. Sjelle librin që ta dhashë (Alb)
bring.IMPV.it.ACC book.DEF.ACC COMP you.DAT.it.ACC gave.AOR.1SG
'Bring the book that I gave you'

Thus Slavic (and/or Albanian, where appropriate) syntax, with a relative word in complementizer position in the relative clause and a finite verb, is the model upon which is based the new relative clause strategy in the Turkish of the western Balkans (and Gagauz in contact with Bulgarian, and, later, Romanian and Russian).

Besides this change in relative and adverbial clause structure, there are two main developments with adjectival clauses that have a broader distribution in the Balkans. These are a change in the form of the complementizer, discussed in §7.7.2.2.1, and a variety of changes involving participles, discussed in §7.7.2.2.2.

7.7.2.2.1 Invariant Relative Marker

The earlier stages of all the Indo-European Balkan languages showed a full-clause strategy of relativization involving an inflected relative pronoun filling the complementizer position in the (relative) clause modifying a nominal. This pronoun was marked for case, much like English *who*/*whom*/*whose*, but also showed agreement in gender and number with the modified nominal head of the relative clause. In later stages of these languages, however, an invariant (absolute) relative clause marker occurs as an alternative, and generally preferred, strategy – especially in colloquial registers. This change in the form of the relativizer-complementizer entails both a lexical dimension and a syntactic dimension.

From a lexical standpoint, one of the forms of the invariant relativizer is interesting, as it is based on locative adverbials, in some instances the interrogative word for 'where.'[213] Sandfeld 1930: 107 notes that not only does Greek που, from Ancient Greek ποῦ 'where?,' come to be used as a relativizer, but the same holds as well for Blg *deto*, from CoSl *kŭde 'where?' (cf. OCS *kъde*, Blg *kăde* (> *gde, de*); Mac *kade* is the source for the complementizer *deka*), and for Alb *tek* 'at.' In Romani, *kaj* is likewise interrogative 'where' and COMP 'that.' Sandfeld cites the following examples:[214]

(7.168) tova gdeto mi kazvate ne e istina (Blg)
that where.REL me.DAT say.2PL NEG is true
'that which you tell me is not true'

(7.169) ti tek je një shpesë (Alb)
you where are.2SG a bird
'you who are a bird'

213 See §4.3.3.4 for a fuller consideration of the lexical side of contact effects on complementizers.
214 The orthography has been modernized; *gdeto* is an older variant of *deto*.

(7.170) η βούλλα που την είχε κρυμμένη στη γλώσσα του (Grk)
the ring where it had.3SG hidden on.the tongue his
'the ring which he had hidden on his tongue'

Inversely, but relatedly, in Aromanian in Greece, *ţi*, originally 'that-which; which, who,' has come to have the meaning 'where?' also, calqued on the polysemy of Greek που. Similarly, Albanian *që* 'that' is probably a borrowing from the Latin relative pronoun *qui* 'who,' *quid* 'what,' or *quod* 'which' (Çabej 2002: 438), and so likely was originally a relativizer. Significantly, it has come to have, in Sandfeld's words, "*tous les sens de gr. ποῦ*" ('all the meanings of Greek ποῦ'), including the locative sense.[215]

These inverse developments show that language contact is involved to some extent here, through the mechanism of calquing. But for the original impetus behind the change of a locative to a relative word in the first place, one may have to look ultimately to nothing more than a natural semantic equivalence of "where" to "in/on which" and an extension thence to a general relativizer; English *where* shows the first step here (cf. *a place where I could lay my head*) and has been generalized as a complementizer, though not specifically relative in nature (cf. *I see where John's been declared the winner*). Sandfeld sums this view up with the observation "*il est vrai que de pareilles constructions relatives se trouvent aussi ailleurs et sont fréquentes un peu partout dans le langage populaire*."[216] It is worth noting, however, that for Slavic this development is limited to Balkan Slavic.

From a syntactic standpoint, the use of a universal invariant relativizer has several consequences. For one thing, it means that the gender distinctions that were essential in relativization in earlier stages of the languages are not indicated when the invariant form is used. And, in colloquial Macedonian, this development has been extended even into the loss of the animate/inanimate distinction in the case-marked relative pronoun, so that *koj* in the meaning 'which [one]' as opposed to 'who' (the former with all types of nouns, the latter with animates) is used where *kogo* 'whom' would be expected in oblique function.

Moreover, with an invariant relative complementizer, as opposed to a case-marked relative pronoun, there is no evidence – in syntactic theories that tolerate movement of elements – of syntactic movement as part of the relativization process; a case-marked pronoun would show case marking associated with a position within the relative clause, hence would also show movement away from that to placement in a complementizer position; by contrast, an invariant relative marker would show no such evidence. The fact that Greek που, Albanian *që*, and Balkan Slavic *deka/deto* can be used to introduce (certain kinds of) complement

215 The word *që* can refer to location in time in expressions meaning 'from X (onward)', e.g., *që moti* 'ages ago' (lit. *që* time), *që tani* 'henceforth' (lit. *që* now), but it can involve space as well as time, e.g., *që nga Janina gjer në Shkodër* 'from Ioánnina to Scutari [= Shkodra –VAF/BDJ]' (Mann 1948: s.v.). Thus this is an instance of isosemy (see §4.3.10) involving a grammatical meaning. In StAlb, *tek* has a primary meaning of 'at'; for more on *tek*, see §7.9.4 and footnote 29 in Chapter 6.

216 'It is true that similar relative constructions are found elsewhere as well and are frequent pretty much everywhere in colloquial language.'

clauses, generally factive ones, suggests that the invariant relativizer is not always a relative or other type of pronoun per se but is just a regular (clause-introducing) complementizer that occurs in complementizer position, introducing a subordinate clause. This aspect of the syntax of Balkan relativization is likely to be independent in each language, i.e., simply a consequence of the particular configuration of facts associated with relativization, and of the analytic demands of particular syntactic theories. That is, any language, Balkan or otherwise, with a similar set of facts would be subject to the same sort of analysis; there is nothing contact-related about a nonmovement analysis of such relative clause structures.[217]

Example (7.170) above shows another syntactic consequence of the invariant relativization strategy. The emergence of an invariant relativizer ties in with object doubling in that a weak object pronoun, such as την in (7.170), which would be translated with the same sort of resumptive pronoun in all the Balkan languages, picks up – resumes, as it were – the reference of the modified nominal head of the relative clause and the invariant relativizer.

Such pronouns also can occur in the relative clause to mark the grammatical relation associated with the relativizer, or to mark pragmatic functions such as enhanced topicality; Sandfeld (1930: 107) gives (7.170) above from Greek as well as the examples (7.168) and (7.169) respectively from SW Macedonian and Albanian, in which the element *"relatif est souvent repris par un pronom personnel*" ('the relativizer is often resumed by a personal pronoun'). We add to those the further examples in (7.171) and (7.172):

(7.171) Našle edna čupa, kade ja plete majka mu[218] (Mac)
found.3PL one girl where.REL her.ACC wove.3SG mother her.DAT
'They found a girl whose mother had plaited her hair'

(7.172) vendit tek e kish ajo ndër urdhër të saj (Alb)
country.DAT where it.ACC had she under orders PC hers
'(to the) country that she had under her command'

Moreover, for relative clauses that target some grammatical relations, the weak object resumptive pronoun is sometimes required, but the various languages differ significantly from one another, as they do with object reduplication in main clauses (on which see §7.5.1). Examples (7.173)–(7.177) from Kallulli 2008: 243–244, with their Macedonian, Bulgarian (cf. Rudin 1994), Greek, and Romanian equivalents, highlight some of these differences. In Albanian, with the universal relativizer (*që*) the resumptive pronoun in the relative clause is optional (7.173a) unless the relative clause comes before the main one, in which case the resumptive pronoun is disallowed (7.174a). In Macedonian, however, the pronoun is required in both contexts (7.173b, 7.174b), while

217 Such is the case for English relative clauses headed by *that*, for instance; see Stahlke 1976 and van der Auwera 1985 for discussion of English *that* as a complementizer and not a pronoun, and, more relevant to the Balkans, van der Auwera & Kučanda 1985 on Serbo-Croatian *što* 'that' in relative clauses.

218 The relevant SW Macedonian dialects neutralize the gender opposition in 3SG clitics, a calque on Albanian and Aromanian; see §6.1.3.1.

in Bulgarian it is optional, albeit preferred, as the construction is perceived as colloquial (7.173c, 7.174c). However, if the inflected relativizer is used, and the head is indefinite, Albanian and Macedonian require reduplication (7.175ab), but it is disallowed in Bulgarian (7.175c). If the head is definite, Albanian requires the universal relativizer (7.176a) and forbids a resumptive pronoun as well as not permitting the inflected relativizer with or without the pronoun (7.177a). Macedonian and Bulgarian both permit either relativizer with the difference that Macedonian requires the resumptive pronoun in both instances (7.176b, 7.177b), while Bulgarian merely permits it with the universal relativizer (7.176c) and forbids it with the inflected one (7.177c). In Greek, the resumptive pronoun is possible within the relative clause regardless of whether the relativizer is inflected or uninflected, though a definite head does not permit the pronoun with the inflected relativizer (7.177d),[219] while in Romanian, a pronoun is required with the uninflected relativizer (7.173e) –(7.177e).[220]

(7.173) a. Ky ështe libri që (e) solli Ana (Alb)
this.M is book.M.DEF COMP it.ACC brought.AOR.3SG A. (Mac)

b. Ova e knigata što ja donese Ana
this.N is book.F.DEF COMP it.F.ACC brought.AOR.3SG A.

c. Tova e knigata deto (ja) donese Ana (Blg)
this.N is book.F.DEF COMP it.F.ACC brought.AOR.3SG A.

d. Αυτό είναι το βιβλίο που (το) έφερε η Άννα (Grk)
this.N is the.N book.N COMP it brought.AOR.3SG the A.

e. Aceasta este cartea pe care Anna a adus-o (Rmn)
this is book.DEF DOM COMP A. has brought-it.F.ACC
'This is the book that Anna brought'

(7.174) a. Libri që (*e) solli Ana është ky (Alb)
book.M.DEF COMP it.ACC brought.AOR.3SG A. is this.F

b. Knigata što ja donese Ana e ovaa. (Mac)
book.M.DEF COMP it.ACC brought.AOR.3SG A. is this.F

c. Knigata deto (ja) donese Ana e tazi. (Blg)
'book.M.DEF COMP it.ACC brought.AOR.3SG A. is this.F

219 See Joseph 1983e for some discussion of the resumptive pronoun relativization strategy in Greek, as well as other strategies. The resumptive pronoun is structurally possible, which is the focus here, though its presence or absence depends in part on particular contexts and intonations, and can vary by speaker. See also Friedman 2008c on relativization in the Balkan languages in general within the context of object reduplication (on which see §7.5.1).

220 In our data, the object marker *pe* is required with the invariant relativizer *care*, though there are speakers who relativize without the *pe*. Gheorghe 2013: 490 gives examples of relativization in what he calls "spoken Romanian," i.e., colloquial usage, that have *care* without *pe* and with a resumptive pronoun, e.g.,:

i. mobilă potrivită pentru garsoniera$_i$ care$_i$ o$_i$ au
furniture suitable for studio COMP it.ACC have.3PL
'suitable furniture for the studio that they own' (lit., ' … that they.have it')

d. To βιβλίο που (το) έφερε η Αννα είναι αυτό. (Grk)
the.N book.N COMP it.N.ACC brought.AOR.3SG the A. is this.N

e. Cartea pe care Anna a adus-o este aceasta (Rmn)
book.DEF OBJ.MRKR COMP A. has brought-it.F.ACC is this
'The book that Anna brought is this [one]'

(7.175) a. Lexova një libër të cilin *(e) mora në (Alb)
read.AOR.1SG one book PC which.DEF.ACC it.ACC got.AOR.1SG at
bibliotekë
library

b. Pročitav edna kniga koja ja zedov od bibliotekata (Mac)
read.AOR.1SG one book which.F it.F got.AOR.1SG from library.DEF

c. Pročetox edna kniga kojato (*ja) vzex ot bibliotekata (Blg)
read.AOR.1SG one` book which.F it.F got.AOR.1SG from library.DEF

d. Διάβασα ένα βιβλίο το οποίο (το) πήρα από τη βιβλιοθήκη (Grk)
read.1SG.PST a.N book.N the.N which.N it.N took.1SG from the library

e. Am citit o carte pe care am primit-o de la bibliotecă (Rmn)
have.1SG read a book DOM COMP have.1SG received-it from at library
'I read a book which I got from the library'

(7.176) a. Lexova librin që (*e) mora në bibliotekë (Alb)
read.AOR.1SG book.DEF.ACC COMP it.ACC got.AOR.1SG at library

b. Ja pročitav knigata što ja zedov od (Mac)
it.ACC read.AOR.1SG book COMP it.F. ACC got.AOR.1SG from
bibliotekata
library.DEF

c. Pročetox knigata deto (ja) vzex ot bibliotekata (Blg)
read.AOR.1SG book.DEF COMP it.F.ACC got.AOR.1SG from library.DEF

d. Διάβασα το βιβλίο που (το) πήρα από τη βιβλιοθήκη (Grk)
read.1SG.PST the.N book.N COMP it.N took.1SG from the library

e. Am citit cartea pe care am primit-o de la (Rmn)
have.1SG read book.DEF DOM COMP have.1SG received-it from at
bibliotecă
library
'I read the book that I got from the library'

(7.177) a. *Lexova librin të cilin (e) mora (Alb)
read.1SG.PST book.DEF.ACC PC which.DEF.ACC it.ACC got.1SG
në bibliotekë
at library

b. Ja pročitav knigata koja ja zedov od (Mac)
it.F.ACC read.AOR.1SG book.DEF which.F it.F.ACC got.AOR. 1SG
bibliotekata
from library.DEF

c. Pročetox knigata kojato (*ja) vzex ot bibliotekata. (Blg)
read.AOR.1SG book.DEF which.F it.F.ACC got.AOR.1SG from library.DEF

d. Διάβασα το βιβλίο το οποίο (*το) πήρα από τη βιβλιοθήκη (Grk)
read.1SG.PST the.N book.N the.N which.N it.N took.1SG from the library

e. Am citit cartea pe care am primit-o de la bibliotecă (Rmn)
have.1SG read book.DEF DOM COMP have.1SG received-it from at library
'I read the book which I got at/from the library'

Relevant here too is the Greek requirement for a resumptive pronoun with possessives, shown in (7.178a) and the Macedonian treatment of possessed as definites in (7.178bc). In the case of Greek, a resumptive pronoun referring to the possessor in the relative clause is required but the DO is not reduplicated, whereas in Macedonian, the possessed DO requires reduplication by virtue of being possessed, but in (7.178b), *čija* 'whose' obviates the need for DEF marking on 'book' and leaves the marking of the possessor as optional, while in (7.178c), the relativizer requires both possessor and possessum to be reduplicated. Today, modern Macedonian speakers consider *čija* bookish, *komu* is normative, *na kogo* is an easternism that has come to dominate among some younger speakers, and *na koe* is the most innovating construction and favored by the youngest generation.

(7.178) a. το αγόρι που χάσαμε το βιβλίο του / *Ø
the boy REL lost.1PL the book his
'the boy whose book we lost'

b. momčeto čija kniga [mu] ja izgubivme
boy.DEF whose.F book him.dat it.ACC.F we.lost
'the boy whose book we lost'

c. momčeto komu / na kogo / na koe mu ja
boy.DEF who.DAT / of who.ACC / of which.N him.DAT it.ACC.F
zgubivme knigata
we.lost book.DEF
'the boy whose book we lost'

Although pronouns in the Balkan languages show case distinctions, even in the languages that have lost case altogether elsewhere, the emergence of an invariant relativizer is consistent with the movement in these languages in the direction of more analyticity and less inflection in the nominal system. Thus there is likely to be a contact-related aspect to these developments, inasmuch as the move towards analytic structures is connected to the paramount importance of fulfilling communicative needs in contact situations.

7.7.2.2.2 Participles

The ancestors of the modern Balkan Indo-European languages for which we have ancient or medieval attestations (Greek, Balkan Romance, Balkan Slavic, Romani) all arrived in the Balkans with rich systems of participles inherited from Indo-European, although the simplification of the Romani participial system may have begun in India or West Asia. The Modern Indic languages of South Asia all have significantly richer participial systems than Romani (Masica 1991: 324–325). Moreover, it is clear from the internally reconstructible evidence of Albanian that its ancestral language, too, must have had a similarly rich system of Indo-European origin. The other systems were all more or less eliminated during the medieval and modern periods. The most important remnants are to be found in verbal adverbs (see §7.7.2.3.2) and the descendants of past passive

participles (which sometimes become deverbal adjectives, i.e., without voice restrictions). Also, Slavic has remnants of the resultative participle in *-l,* and Romani has a perfective participle (from the Old Indic participle in *-ta-,* Matras 2002: 151), both of which are important in forming modern past tenses. Reductions are found throughout Slavic and Romance, so the fact of reduction itself is not specifically Balkan, although some usages, i.e., results of the tendency to reduction, may demonstrate specifically Balkan convergences (see §6.2.3 on the perfect, §7.7.2.1.1 on infinitives, §7.7.2.2 on relative clauses, and §7.7.2.3 on adverbial clauses).

Moreover, given Romani's relative lack of participles vis-à-vis the rest of Indic, as well as the obvious influence of Greek seen in the borrowing of the Greek mediopassive participle marker -μενο- (Matras 2002: 160), it is fair to say that Romani developments are consistent with the Balkans rather than with South Asia, where the vigor of participles may have a South Asian areal component (Masica 1976). We can also note here that in the case of Macedonian, the limitation of the Common Slavic resultative participle in *-l* to a nonparticipial formant of analytic tenses is not shared with Bulgarian, which still permits that participle in attributive uses. Likewise, the Macedonian reflex of the Common Slavic past passive participle – formed only from transitive verbs in OCS – is now a pure verbal adjective that can be formed from intransitive as well as transitive verbs. This is also true in southeastern Bulgarian dialects in contact with various other languages, but not in Standard Bulgarian and the dialects on which it is based. In the case of Macedonian, contact with Aromanian and Albanian, which have this same type of phenomenon, is relevant (see §6.2.3). Most of Bulgarian, however, patterns with Greek in not having productive verbal adjective formation of this type (PPP morphology) with intransitive verbs. We can also note, in passing, that the participles were artificially introduced into literary Russian from the South Slavic recensions of Church Slavonic – colloquial Russian having also lost the system except for certain remnants – and from literary Russian they were re-introduced to literary Bulgarian in the nineteenth century.[221] Another relevant point is the situation with WRT and Gagauz, where the original complex participial (and, more generally, nonfinite) system of Turkish (and other Turkic languages) has been replaced to varying degrees by subordinator plus finite verb (see §7.7.2.1.1.1.6, §7.7.2.1.1.2.6, and §7.7.2.3.1.3). Thus, while the loss or simplification or transformation of complex inherited participial systems is a feature of all the Balkan languages, the shared innovations are better treated in the developments of specific types of subordinate clauses such as relative and adverbial.

221 Native Russian remnants are generally also adverbial or nominalizations, but these are clearly developments independent of the participial system introduced from Church Slavonic. North Russian dialects, however, can also form verbal adjectives from intransitive verbs, a development that is independent of, but parallel to, Macedonian.

7.7.2.3 Adverbial Clauses

Adverbial clauses serve a variety of functions, mostly modifying main clauses along typical dimensions for adverbs, covering various temporal relations, including causality, circumstance, means, and conditionality. A few specific uses, forms, and constructions that fall under this rubric are of Balkanological interest. One such construction is the use of the DMS elements (Alb *të*, Grk να, BSl *da*, Rmi *te,* BRo *să, si, s'*), in conditional ('if') clauses; this use, however, is treated in §6.2.4.2.1, inasmuch as it almost prototypically reflects the modality of these elements. Here we discuss certain temporal or circumstantial constructions, the use of verbal adverbs, and the form of purpose clauses.[222]

7.7.2.3.1 Verbal Adverbs (Gerunds)

Among the nonfinite forms in the various Balkan languages are some that are based on or descended from participles but that lack the agreement and inflectional properties of participles per se. They function in ways that differ from participial modifiers, modifying verbal clauses or sentences rather than nominal clauses. Inasmuch as typical participles are verbal adjectives, these sententially modifying forms can be considered verbal adverbs, a term that is used in Macedonian (*glagolski prilog*). In the Romanian grammatical tradition, the term *gerunziu* is used, from Latin *gerundium* (from *gerō* 'bear, carry [out]'), which is also used in Turkish (*gerundium*).[223] Turkish also has the neologism *ulaç*, which, like the Albanian term *përcjellore* comes from a native root meaning 'accompany.' Bulgarian uses the Russian borrowing *deepričastie* which is usually translated either 'gerund' or 'verbal adverb,' although etymologically it means 'active participle' as does the Greek term ενεργητική μετοχή. The Bulgarian/Russian and Greek terms have a diachronic reference. The terms *gerund* and *gerundive*, while used for various nonfinite forms in the grammatical descriptions of a variety of languages, frequently mean 'verbal adverb' in descriptions of Romance, Slavic, and Turkic languages. In this section we use the term *gerund* to mean specifically 'verbal adverb,' including analytic constructions such as standard Albanian *duke* + participle (see §7.7.2.3.1.1). We use the term *gerundive* both as an adjective form of *gerund* and to refer to gerund-like items and markers. In addition to their modifying function, gerunds also serve a clause-linking function referred to earlier as that associated in some languages with converbs (cf. §7.7). Moreover, like the descendants of past passive participles, gerunds in some Balkan languages can be used in analytic constructions that can be treated as paradigmatic or quasi-paradigmatic. In the cases of Balkan Slavic, Balkan Romance, Albanian, and Romani, there are a few phenomena associated with these forms that are either definitely or possibly of

222 See (7.123) in §7.7.2.1.1.2.1 for the discussion of the Greek temporal infinitive which is, strictly speaking, a type of adverbial clause. On the typologically parallel development in Slovene, see Mihevc-Gabrovec 1973 and Friedman & Joseph 2019.

223 Romanian dictionaries gloss *gerunziu* as 'indefinite/present participle.'

Table 7.4 *Balkan gerund suffixes*

Mac	-jkji, e.g., *gledajkji* '(while) seeing'
Blg	-jki, e.g., *četejki* '(while) reading'
Grk	-οντας, e.g., πηγαίνοντας '(while) going'
Aro	-nda/-ndalui, e.g., *durñinda*/*durñindalui* '(while) sleeping' (south only)
Rmn	-nd, e.g., *alergând* '(while) running'
Megl	-ndurlea/-ndăra, e.g., *căntăndurlea / căntǫndăra* '(while) singing' (rare)
Rmi	-indo[s/j/Ø/r], e.g., *phirindo*[*s/j/Ø/r*] 'while walking' and *-i,* e.g., *phuči-phuči* 'after much asking'
Trk	a variety of constructions[224]

diffusional origin. In other instances, similarities contribute to an appearance of parallel structuring across the languages regardless of source.

7.7.2.3.1.1 Use and Emergence

Gerunds as understood here normally express a temporal (especially simultaneous), causal, concessive, or conditional – and subordinated – relation between one verb (the gerund), representing a reduced clause, and another verb, representing a main clause; it is in this way that the gerund functions as a kind of clause-linking subordination.[225] Table 7.4 lists the native suffixes in languages that have gerunds.[226]

Albanian has a periphrastic construction using a participle preceded by a particle whose shape in the current standard and most of Tosk is *duke* and in the former standard and most of Geg (also Arbëresh and Dukat, Vlora region) *tue*, usually glossed 'while,' etc.,' e.g., Tosk *duke punuar*, Geg *tue punue*

224 The rich system of Turkish adverbial suffixes is beyond the scope of this section, since Turkish uses these where the Indo-European Balkan languages would have a variety of subordinators. See Lewis 1967: 174–194 and Göksel & Kerslake 2005: 461–485 for details. Turkish is mentioned in this section only when specific contact phenomena are or might be involved. Of particular note is the replacement in WRT of gerundial forms – as with participial forms – by subordinate clauses calqued from Balkan Slavic and Albanian (cf. Friedman 1982c, 2006c; Matras 2003/2004).

225 In Romanian, Albanian, and some dialects of Romani, however, the gerund can also function as part of a main clause with 'be' as the auxiliary or copula. We return to this point below.

226 Balkan Slavic shows a huge amount of variation in the form of the suffix used here, as Hellgren 1996 documents. See now also Majer 2021. We give standard language forms, but we note that the Standard Bulgarian form is an importation from those Balkan Slavic dialects that became standard Macedonian, cf. Beaulieux 1950: 194, and footnote 230 below. For our purposes, the standard forms generally suffice. The variations have to do with specific reflexes of Common Slavic **tj* (*št*, *šč*, *č*, *ć*, *ḱj*, etc.) and various petrified case endings. For Romani, the shapes of the form in *-indo-* are likewise quite varied by dialect (see Boretzky & Igla 1994: 411, 2004: 1.190–191), while the gerund in *-i* may have been influenced by Greek (Matras 2002: 160). The Albanian analytic gerund marker also shows tremendous dialectal variation (see footnote 227 below, and Gjinari 2007: Map 128). This leaves Greek and Romance as relatively more uniform, although, as noted in the text, the gerund is absent or obsolescent in SDBR, and some speakers of Meglenoromanian have borrowed the suffix from Macedonian (see §6.2.1.1.3).

'while working.'[227] The adverbial marker can be compared to Bulgarian *kato* plus finite form, which in eastern Bulgaria is the only native gerundive construction.[228] The Albanian gerund, unlike that of most Balkan languages, can also be used in most dialects to form a progressive quasi-paradigm, e.g., Tosk *jam duke dalë* 'I am going out,' which is in competition with *po dal* 'idem.' These two types of progressive compete across most of the Albanian diasystem. To the extent that in a given dialect one or the other construction is the only one occurring, it is generally the case that *po* is southern Tosk while *tue* is eastern Geg, but the progressive with gerund is the sole construction in Muzhakat (Grk Mouzakéïka), a Çam area, and progressive *po* is the sole form in Repa (northeastern Kosovo), so the characterizations 'Geg' and 'Tosk' in this case are simply general tendencies (see also §6.2.2.4).

In the southern Vlax (Džambaz, Gurbet) and Balkan II (Bugurdži = Burgudži) Romani dialects of Kosovo and western North Macedonia, Albanian *tuj* is borrowed and attached to finite forms, e.g., *tuj džav / džas / džal* 'while going 1SG / 2SG / 3SG,' with the person of the finite verb coreferential to the person of the main verb, but not influenced by its tense, as in (7.179):

(7.179) Tuj džav o dromal, araćhum me phrales (Boretzky 1993: 83)
GRDM go.1SG.PRS the road.ADV found.1SG.AOR my.OBL brother.ACC
'Going along the road, I found my brother'

The Albanian gerundive progressive can also be calqued into these dialects of Romani, e.g., *sjom tuj džav* 'I am going' (Boretzky & Igla 1994: 411).

Synthetic gerunds are, for the most part, relatively rare. For Bulgarian, Hauge 1999: 135 writes that they "are more frequent in bookish style."[229] More revealing is Stojanov 1983: 382, where it is made clear that the form was imported into standard Bulgarian via nineteenth-century literary works from what he calls "southwest Bulgarian dialects," i.e., what the rest of the world recognizes as Macedonian, where even today verbal adverbs (gerunds) are alive and in common usage.[230] In the eastern Balkan Slavic dialects that formed the basis of the Bulgarian standard, the verbal adverb has survived only in a few lexicalized forms (Stojkov 1968: 164). In Macedonian, the gerund is formed only from imperfective verbs with the exception of the lexicalized *bidejkji* 'because' ('while being'; cf. English *being as how*). In southwestern Macedonian dialects in contact with Aromanian, e.g., Ohrid-Struga, gerunds are extremely rare

227 As mentioned in footnote 226, the Albanian verbal adverb marker is subject to considerable dialectological phonological variation: south of the Shkumbî river (and including Arbëresh and Arvanitika) we also have *dyke, tuke/tukë, tyke, dikue*, *te*, *tue*, *ture, tuk me, duk me*; north of the Shkumbî we also have *tuo, tu, tuj, tye,* and *tyj* (Sh. Demiraj 1985: 971; Gjinari 2007: Map 128).

228 Unlike Albanian *duke*, however, Bulgarian *kato* has a wide range of uses and lexical meanings, e.g., 'like, as, since, when, as soon as, how, as if, etc.' and it can be preceded by the preposition *sled* 'after' or the adverb *tăj* 'thus,' it can be preceded or followed by the relativizer *če* 'that,' and it can modify both NPs and VPs. Thus, while *kato* + finite verb is the only native eastern Bulgarian equivalent of the gerund, *kato* is not a dedicated gerund marker in the sense that *duke* (etc.) is for Albanian.

229 See also Leafgren 2019 on Bulgarian gerunds. 230 Cf. also footnote 226.

(Markovikj 2007: 162–163). The gerund is alive in BCMS (including Torlak; A. Belić 1905: 581ff.), where there is a present gerund from imperfective verbs and a past gerund (from the old past active participle) for perfective verbs. In Balkan Slavic, the past active participle survives only in the lexicalized adjective *bivš* 'former' (lit., 'having been'). We can thus say for Balkan Slavic that the gerund was lost as a grammatical category in the eastern part of the larger diasystem, and restricted in the western part. A similar restriction has occurred in Albanian, where there is no synthetic gerund and past gerunds of the type *duke pasë bërë/duke qenë bërë* 'having done/been done (become)' are found mostly in old authors from the north (Sh. Demiraj 1985: 971).

Similarly, there is an uneven distribution of gerunds in Balkan Romance. They occur in Romanian, where, as in Albanian, they can participate in the formation of analytic finite forms. Unlike Albanian, however, these are not progressives, and in fact the use of the gerund with various tenses of *fi* to form progressives was considered a regionalism and, according to Nicula 2013: 248, it has disappeared entirely from modern usage. The present presumptive, however, e.g., *o fi mergând*, etc. 'supposedly/apparently he is going, went, etc.,' is a living paradigm in Romanian (see §6.2.5.7 on the presumptive mood).[231] In Aromanian, however gerunds occur only in the south (i.e., Greece) but not in the north (i.e., Albania and North Macedonia), as noted by Capidan 1932: 481–482, and confirmed by Markovikj 2007: 162 for Ohrid-Struga and Gołąb 1984a: 106 for Kruševo. Even in Greece, Aromanian gerunds are not particularly robust. Bara et al. 2005: 219 write that in Turia (Grk Krania) the gerund was obtained only through elicitation, e.g., *kãtãndalu* 'singing,' and the only naturally occurring example was the lexicalized *imnãndalui* 'on foot' (lit., 'while walking').[232] For Meglenoromanian, Atanasov 2002: 234–235 notes that its occurrence is very rare in modern usage. It is replaced by *cu* 'with' plus infinitive, a construction of a type also found in Albanian, Aromanian and mutatis mutandis in Macedonian (see below). The Meglenoromanian dialect of Tsărnarekă (Grk Kárpi) has borrowed the Macedonian gerund marker *-ájkji*, e.g., *nirdzeájḱi* 'going' (Caragiu-Marioțeanu et al. 1977: 209; Atanasov 2002: 234–235). In Greek, gerunds can be heard on occasion in informal conversation, but they are rare.

An additional Albanian form with a similar meaning, characterized as "ingressive or inchoative" by Camaj 1984: 165, and called the "participial absolute" by Newmark et al. 1982: 100, consists of the verbal noun (i.e., the participle preceded by the particle of concord) preceded by *me* 'with' or *një* 'one, a' or *me një* (cf. Sh.

231 Unlike the gerunds of the other Balkan languages, the Romanian gerund can behave and inflect like a participle, as seen in the following examples from Nicula 2013: 246: *o rană sângerând de două ore* 'a wound [that.has.been.] bleeding for two hours' and *o rană sângerândă* 'a bleeding [F] wound.'

232 Caragiu-Marioțeanu 1968:122–124, 160–161; Nevaci 2006: 175–176; and Vrabie 2000: 65 all discuss gerunds in Aromanian without specifying dialectal distribution. Nevaci 2006: 175 notes that, unlike in Romanian, the Aromanian gerund cannot occur under negation.

Demiraj 1985: 979).[233] This construction has parallels in Aromanian and (SW) Macedonian (cf. Markovikj 2007: 164); see also §7.7.2.1.2.

Thus we can argue that gerunds are present in the Balkan languages in the sense wherein 'language' denotes a generalization for a varied set of registers and dialects. In fact, evidence from dialect descriptions, where there would not be significant influence from a normative literary standard, offer evidence of the reality of gerunds and gerundive constructions, although the synthetic forms are arguably in retreat.

As to their origins, the gerunds in each language are tied historically to earlier forms and systems of verbal nominals, i.e., verbal substantives (gerunds) and verbal adjectives (participles), and they represent forms that have been functionally specialized and, in the case of the adjectives, have generally come to be invariant, no longer showing agreement. The form of the Balkan Romance gerund most likely "continues the ablative form of the Latin gerund in *-ndo*" (Nicula 2013: 245), whereas participles are the source of the Albanian, Greek, Balkan Slavic, and Romani forms. Greek -οντας continues the Ancient Greek active participles in -ο/αντ- (Mirambel 1960; Manolessou 2005; Gorton 2013), the Romani *-(i)nd-* likewise continues the Old Indic present active participle in *-nt-* (Matras 2002: 160), and the Slavic forms continue the related present active participle in **-ontj-*.[234] In all of these languages, the gerunds constitute a small remnant of what was once a much larger system of participial forms, as discussed in §7.7.2.2.2. For Greek, Manolessou 2005 documents the occurrence of both the invariance and the adverbial usage in earlier Greek prior to any period of contact with the ancestors of the remaining Balkan languages (except, perhaps Albanoid). The participle in OCS occurred commonly in dative absolute constructions that offer in some instances adverbial subordination similar to what is found in later Balkan Slavic, although later dialectal forms frequently appear to be from locatives.

The Albanian periphrastic form is built on a participial base, but the participle is the single generalized one in *-n-* (> *-r-* in Tosk)*/-t-/-m* and not a descendant of the old present participle. Moreover, the periphrasis gives it the appearance of being a relatively recent formation. For the origin of the marker *tue/duke*, the following summary of the proposed historical explanations is based on Sh. Demiraj 1985: 971–979, which can be consulted for the detailed bibliography. Pedersen proposed a combination of *[kë]tu* 'here'+ *ke* = *ku* 'where.' Çabej proposes *tuk* 'where' + *e* 'it.ACC.'[235] Sh. Demiraj 1985: 972 cites his own earlier proposal of a preposition *tu[k]* + *e* 'it.ACC' with an unarticulated verbal noun, although he qualifies his acceptance of Çabej's proposal of the second element "*me reserve*" ('with reservations'). Demiraj goes on to argue convincingly that the initial *t-* is original. He further adduces arguments for the relationship of *[kë]tu* with *tuk, tek, te*, all with a meaning of '[there] where' as well as 'here.' As to the final *-e* in *tuke, tue*, etc., Demiraj cites

233 This *me* is also the marker of the Geg infinitive (cf. §7.7.2.1.1.1.4).

234 See footnote 226 on the variety of realizations.

235 The explanation of the variants is of Albanological or general linguistic significance, but does not affect the Balkanological arguments that concern us here.

the expression *tue e tutje* 'here and there' in Buzuku as well as the fact that the suffix *-e* must be of relatively recent origin, since otherwise, as an unaccented final element, it would have disappeared or given some other result. We are thus left with a relationship of the type *te* : *tek* = *tu* : *tuk* = *tue* : *tuke*. Demiraj's primary question is whether to consider the historical source to be adverbial 'where' or prepositional 'at,' and, on the basis of the occurrence of other prepositions with the participle or verbal noun in various adverbial and other constructions (e.g., *me* 'with,' *pa* 'without,' *për* 'for'), Demiraj favors a prepositional explanation. On the 'where' side, however, are some Balkan parallels to be discussed below.

The Romani gerund in *-i* is of obscure origin, but Matras 2002: 160 suggests it may be related to the loan-verb adaptor *-i*, which in turn may be from the Greek 3SG present marker. The combination of the original gerund and this *-i* accounts for the shape *-indoj* found in Bugurdži/Burgudži. In many or most Romani dialects of the Balkans, the gerund is usually *-indor*, with a few frozen lexical items in *-es* or athematic nouns in *-os*.

All in all, we can say that there is a general movement in the direction of invariance and adverbial clausal modification in all of these languages, and the culmination of the emergence of the present-day forms with their particular syntax came in the period of Balkan contact in the second millennium CE. To be sure, the processes could be described as internally motivated for each individual language, and indeed comparison with non-Balkan Slavic and Romance provides arguments for this viewpoint.[236] One could also argue for a typological explanation.[237] Nonetheless, the development of various analytic constructions also points to some specifically Balkan, i.e., contact-induced, developments.

First, it is important to recognize that, just as there are finite replacements for the infinitive in the Balkans (see §7.7.2.1 in general, and §7.7.2.3.3 on purpose clauses), so too the nonfinite gerunds have finite parallels introduced by subordinating conjunctions (complementizers). As seen above in the Romani example cited in (7.179), the Albanian gerund marker can be borrowed and used with a finite verb form. In Greek, the gerundial phrase κοιτάζοντας προσεχτικά '(while) looking carefully' is equivalent to a full finite subordinate clause ενώ κοίταζα προσεχτικά 'while I.was.looking carefully.' In Bulgarian, the gerund in *-ejki/-ajki* was purposefully added to the otherwise eastern-Bulgarian-based standard, which dialectal base has forms in *-št-*, etc. The shape that was adopted in the standard language was that of Pirin Macedonia and the southern Dupnica region just north of there (Stojkov 1968: 164). It was promoted by writers from eastern Bulgaria for ideological

236 For instance, the French construction with *en* 'in' plus an invariant present participle, e.g., *en passant* 'in passing,' *en chantant* '(while) singing,' parallels the Balkan Romance gerunds in use, and Russian gerunds in *-ja*, e.g., *čitaja* '(while) reading,' even if not formed with cognate material, do show a functional parallel to the Balkan Slavic forms.

237 Stump 1985, for example, has demonstrated that the interpretation of absolutes in English, constructions that are similar in some ways to the Balkan gerunds, is pragmatically determined. This in turn suggests the possibility of a tendency in human language as such rather than a matter of transmission or diffusion.

reasons, as it was not native to their dialects. (cf. Leafgren 2019). In the written language, analytic constructions with *kato* + finite verb are most common, and the imperfect is also frequently used in such functions (Leafgren 2019).

Similar finite paraphrases can be found in any of the Balkan languages, and, as noted above, in some dialects of the languages in question, the gerund is rare or nonexistent and has receded in favor of finite or other, nongerundial, alternatives (§7.7.2.1.2). This development is in keeping with the general trend in the Balkans away from the use of nonfinite forms in the direction of finite verbal constructions.

Moreover, Sandfeld 1930: 108 mentions a particular convergence involving a finite equivalent to the gerunds, and this convergence is also connected to analytic gerunds themselves. He notes that in Greek, Balkan Romance, and Albanian, the collocation of 'there' plus 'where' with a finite verb (sometimes preceded by the DMS) can be used as the equivalent of the gerund:

(7.180) a. acolo iu imna / acolo iu să minduia (Aro)
there where walks there where DMS thinks
'while he was walking' / 'while he was thinking'

b. di undi ąș ąmna drumu (Megl)
from where REFL walks road
'while he was going along the road'

c. atje tek flinte (Alb)
there where sleep.IMPF.3SG
'while he was sleeping'

d. εκεί που πήγαιναν στο δρόμο (Grk)
there where go.IMPF.3PL on.the road
'while they were going along the road'

We can add to Sandfeld's observation by noting that in addition to these languages we also have locative adverbial constructions with finite verbs in Balkan Slavic, as in (7.181–7.182):

(7.181) Ene go dedo kaj ide, torba mu polna so mečinja (Mac)
here him gran'pa where go.3SG bag him.DAT full with bears.DIM
'Here's gran'pa coming with his sack full of bear cubs'

(7.182) Go videle kaj što odi na kino (Mac)
him saw.3PL where COMP goes.3SG to movies
'They saw him going to the movies.'

Etymologically, Bulgarian *kato* corresponds exactly to colloquial Macedonian *kaj što* insofar as *ka* is 'where' and *-to* is the relativizing particle. In Skopje Bugurdži, a similar usage of *kaj* 'where' occurs, e.g., (cf. also §6.2.2.4):

(7.183) Kaj ikalel o sastər, pa čivlas xari pošik
as took.out.3SG the iron and.then throw.IMPF.3SG little sand
'As he takes out the iron, he would add a little sand.' (Bodnárová et al. 2018a)

The borrowing of the Macedonian gerund marker into Meglenoromanian and the borrowing of the Albanian gerund marker into Romani cited above are two obvious

examples of contact-induced change connected with gerunds. Likewise, gerund constructions using the verbal noun are so clearly parallel among SDBR, Albanian, and Macedonian that convergent calquing is certainly a plausible explanation, e.g., Megl *cu lăgári* 'running' (lit., 'with running'), Mac *so trčanje* '(with) running,' Alb *me të vrapur* 'idem.' See §6.2.2.2.3 for additional discussion.

7.7.2.3.1.2 Coreference Relations

One detail about the syntax and semantics of Balkan gerunds is the question of control or coreferentiality. The Greek -οντας gerund is noteworthy, Balkanologically speaking, as it seems to be involved in an effect via language contact on a co-territorial variety of Albanian. More specifically, in Greek, the default and by far the most common and generally preferred reading of -οντας forms treats an understood subject of the gerund as being controlled by, i.e., coreferent with, the subject of the main clause.[238] That is, the slot in the syntax and/or semantics for the interpretation of the subject of the gerund is controlled by the main clause subject.[239] How this "subject control" is to be modeled formally is not a concern here, but the key fact is that the gerund is generally interpreted as having an actor/agent who is coreferent with the actor/agent (subject) in the main clause.[240]

In Albanian, however, the understood subject of the *duke* + participle construction that parallels the gerunds (see §7.7.2.3.2.1) can be either the subject of the main clause or an object, as the examples in (7.184) from Newmark et al. 1982: 97–98 show:

(7.184) a. Njeriu po afrohej duke ecur
man.NOM.DEF PROG approached.3SG GRDM walk.PTCP
'The man was approaching by walking'

b. Duke kaluar përpara Mapos, i ra ndërmend të blinte
GRDM pass.PTCP by Mapo her.DAT fell to.mind DMS bought.3SG
një suvenir
a souvenir
'Passing by the Mapo [a Tirana newspaper], it occurred to her to buy a souvenir'

c. Pashë në rrugë një tufë kalorësish duke ikur
saw.1SG on road a bunch riders.ABL GRDM go.PTCP
'I saw a bunch of riders going on the road' (i.e., 'as they were going … ')

The interpretation of the subject of the *duke* form is controlled by a subject in (7.184a) but by an object in the other sentences, an indirect object in (7.184b) and a direct object in (7.184c).

238 Tsimpli 2000 offers a few cases where the -οντας form is controlled by a nonsubject, but she considers these to be pragmatically determined and certainly somewhat unusual.

239 This is true for the gerunds in other Balkan languages as well; see, e.g., §6.2.2.2.3 on Macedonian.

240 That is, one could posit an empty "PRO" subject with the verbal adverb that is indexed as coreferent with the main clause subject, though there are other possibilities, including considering the verbal adverb to be subjectless and the interpretation more a matter of pragmatics than syntax per se. See Tsimpli 2000 for discussion.

By contrast, in Arvanitika, the Albanian dialect in Greece in heavy contact with Greek for centuries, as noted by Tsitsipis 1981, the interpretation of the subject of the gerund parallels Greek, being restricted to only subject control; there are thus no object-controlled instances to cite, and Tsitsipis 1981, and Joseph 1992d following him, give the example in (7.185):[241]

(7.185) Nani duke marr miruðja kujljtova
Now GRDM take.PTCP smell remembered.1SG
'Now I remember (it) by smelling'

Admittedly, an object-control interpretation would be pragmatically unusual in (7.185), as the understood object ('it') is not likely to be taking a smell of something, but such an interpretation apparently is systematically unavailable even when pragmatically viable, according to Tsitsipis.

The difference in the default interpretation of these gerunds between the variable subject or object control in standard Albanian and the subject-only control in Arvanitika is striking, all the more so since Arvanitika has moved in the direction of its contact language, Greek. What we see here appears therefore to be a matter of a shift in the syntax/semantics of control induced by language contact; the likely mechanism here is transfer from Greek as the second, but dominant, language for many of these Arvanitika speakers onto the control possibilities with parallel structures in their first language, in essence a calquing of the Greek pattern into Arvanitika.

The Arvanitika restriction is all the more interesting as there is some fluidity across the Balkans in the control of the understood subject of gerunds. For Macedonian, for instance, we can note that while co-referentiality with the subject is both the norm and the standard, in reality noncoreferential gerunds occur:

(7.186) a. Odejkji po patot, mi padna čantata
go.VBL.ADV along road.DEF me.DAT fell purse.DEF
'Going along the road, my purse dropped'

b. Vrakjajkji se doma, Vi poželuvame "srekjen pat"
return.VBL.ADV INTR home you.DAT wish.1PL "bon voyage"
'Returning home (i.e., as you return home), we wish you "bon voyage"'

The Bulgarian analytic gerund is likewise not necessarily co-referential with the subject, e.g.,:

(7.187) a. Kato se sprja trenăt, toj skoči na zemjata
as INTR stopped train.DEF he jumped to ground.DEF
'The train having stopped, he jumped to the ground.'

b. I sega mi se stjaga sărceto, kato si gi pripomnja
and now me.DAT INTR stopped heart.DEF as INTR.DAT them remember.1SG
'Even now, remembering them (= as I remember them), my heart contracts.'

241 Examples are hard to come by, since the use of the verbal adverb has diminished considerably in Arvanitika, a likely victim of the severe endangerment the language has been experiencing for the past half-century or more.

For Romanian, gerunds that are not coreferential with the subject are, as in Albanian (but not, e.g., French or Italian) part of the norm. Thus example (7.188) is ambiguous:

(7.188) Andreea îl vede margând spre şcoală
A. him sees walk.GRD to school
'Andreea sees him while s/he is walking to school' (Nicula 2013: 251)

7.7.2.3.1.3 Finite Adverbial Clauses in Balkan Turkic

As with adjectival clauses, so too with adverbial clauses, where the Balkan tendency is to use a subordinator and a finite verb, e.g., Mac *koga*, Blg *kogato*, Alb *kur*, Rmn *cînd*, Aro *cãndu,* Megl *cãn,* Rmi *kana*, Trk/Gag *(h)açan*, as shown in (7.189):[242]

(7.189) Zerzele açan idi, ben uydum (WRT, Skopje, 2008 VAF)
Zemjotresot koga beše jas spiev (Mac)
Zerzele kur ishte unë flija (Alb)
earthquake when was.3SG I sleep.PST.1SG
Zelzele olunca ben uyuyordum (StTrk)
earthquake become.GRD I sleep.IMPF.1SG
'When the earthquake happened, I was asleep'

The general loss of participles or their transformation into ordinary adjectives or de-verbal adverbs is part of this tendency. Here Balkan Turkic provides a striking example of contact-induced change related to that discussed in §7.7.2.2. In WRT and the dialects of Gagauz with heavy Bulgarian contact, the usual adverbial use of participles gives way to analytic constructions with subordinators of interrogative origin, as in example (7.190):

(7.190) Bän açan bu papaz şkolasındaydım
I when this priest school.3POSS.LOC.PCOP.1SG
ozaman gelirdim evä (Gag, Menz 1999: 120)
that.time come.GPRS.PCOP.1SG home.DAT
'When I was at the seminary, I would always come home'

7.7.2.3.2 Purpose Clauses

Purpose clauses make an appearance at various points elsewhere in this overall discussion, tied to presentations about particular language facts or particular constructs of relevance. They figure prominently in the discussion of the infinitival developments since infinitives were involved in the expression of purpose in all the relevant languages – see §7.7.2.1.1.1.4 for mention of

242 In general the interrogative/subordinating distinction is absent, but Blg *-to* is an exception in this respect. Macedonian can use the (etymologically) related *što* 'what' in a similar function, but in the colloquial language it is not obligatory, and in the standard language (and the dialects on which it is based) the rules are subject to variation. In StTrk *haçan* 'what time, when' is archaic, *kaçan* is provincial, and *ne zaman* is standard. Both WRT and Gagauz have *açan.*

purpose expressions in Albanian, §7.7.2.1.1.2.1 for Greek, §7.7.2.1.1.2.2 for Balkan Slavic, and §7.7.2.1.1.2.3 for Balkan Romance (with §7.7.2.1.1.2.3.1 for Aromanian specifically), while §7.7.2.1.2 considers them in the context of nominalizations, and §7.7.2.1.3.3 treats them under the rubric of composite subordinators. For the purpose of this section, it suffices to offer a few notes on some aspects of the discussion in the previous sections, as a full treatment here would be redundant.

First, the recurring mention of purpose expressions in the sections on the infinitive is hardly accidental; a key use for the infinitive across the Indo-European languages and probably also in Proto-Indo-European – to the extent that it can be said to have had infinitives distinct from verbal nouns (see §§7.7.2.1.1.2.4 and 7.7.2.1.2, and footnote 243) – was the expression of purpose (cf. Jeffers 1975). Thus the infinitives of purpose in the Balkans, whether occurring alone or introduced by a prepositional element of some sort, represent inheritances from earlier stages.[243] Moreover, Haspelmath 1989 has argued that purposives are a cross-linguistically frequent, and thus natural, source of infinitives, so that the use of infinitives in the expression of purpose in pre-modern stages of the Balkan languages is entirely to be expected on typological grounds. And, in that regard, it can be noted that purpose infinitives remain in use in the present day in Albanian, both Geg and Tosk, Romanian, and even (from Deffner 1878: 223; Sitaridou 2007, 2014a,) Pontic Romeyka Greek, as shown respectively by the examples in (7.191):

(7.191) a. Gjergji shkoi përjashta me mësue (Alb, Geg)
Gj. went abroad INFM study.PTCP
'Gjergj went abroad (in order) to study'

b. Shkova në Petrograd për të raportuar mbi gjendjën (Alb, Tosk)
went.1SG to Petrograd for PC report.PTCP about situation
'I went to Petrograd (in order) to report on the situation'

c. Studiu pentru a înţelege mai bine (Rmn)
study.1SG for INFM understand.INF more well
'I study in order to understand better'

d. Xtes ti nixta elepenete parpatesinete? (Pontic Grk)
yesterday the.ACC night.ACC see.PFV.PST.2PL walk.AOR.INF.2PL
'Last night could you see (in order) to walk?'

Second, some developments with purpose infinitives in the individual languages serve as an important reminder that the processes leading to the loss of the infinitive were neither monolithic nor unidirectional. In New

243 This statement holds, even if the grammatical traditions of some of the languages, especially Slavic and Romance (Latin), recognize a category of "supine" distinct from "infinitive," used, e.g., in purpose expressions. Such supine forms diachronically are just specialized case forms of verbal nouns (e.g., accusative of a deverbal noun in *-*tu*- in the case of the Slavic supine), like the forms labeled as "infinitive." Note that Common Slavic *-*tŭ* for the supine is from PIE *-*tum*, the same suffix and ending found in the only Classical Sanskrit infinitive, forms in -*tum* (one of the several infinitives found in Vedic Sanskrit).

Testament Greek, for instance, as noted in §7.7.2.1.1.2.1, purpose expressions were one area where the infinitive expanded in use, counter to the more general decline in infinitival use evident even at that early point in Postclassical Greek (Blass & Debrunner 1961).

Third, as mentioned in §7.7.2.1.3.3 (and see Table 7.2), purpose expressions in the Balkans come to have a composite subordinating marker, consisting of the preposition 'for' plus the DMS appropriate for each language, as in (7.192), with the first person singular of the verb 'write,' thus 'in order that I write / in order for me to write':

(7.192) a. për të shkruaj (Alb)
b. za da piša (Blg)
c. για να γράψω (Grk)
d. za da pišam (Mac)
f. dži te hramonav (Rmi)
g. pentru să scriu (EMR)[244]

In each case in (7.192), the formation is synchronically somewhat anomalous in having a DMS-headed verb ostensibly governed by a preposition, a word that in principle looks for a noun phrase, not a verb, as its complement. These constructions represent finite equivalents for the earlier, and in some cases still synchronically available, purpose infinitive, thus representing a path along which the infinitive could have retreated. Here we can also mention the Albanian use of COMP *që* + DMS (*të*), which is calqued into the Macedonian of Albania as *što/ščo* + DMS (*da)* (Makartsev & Wahlström 2018).

7.8 Diverse Clause Types

Although many different types of clauses are treated in the preceding sections, there are further types that do not fit into the rubrics above but that nonetheless deserve mention since they show some degree of parallelism across two or more of the Balkan languages. As with all such cases of superficial convergence, however, the mere fact of convergence is not enough to guarantee a language contact explanation. Each such case is taken up in turn in the various sub-sections that follow.

244 Early Modern Romanian is given here because Modern Romanian uses *pentru* with *ca* 'as' followed by the subjunctive with the DMS *să*, or *pentru* with the infinitive (thus, *pentru ca să scriu* 'in order that I write' (lit., 'for as DMS I.write') or *pentru a scrie* 'in order to write' (lit., 'for INFM write.INF'), and so with a somewhat different syntax from the older, more Balkan, construction in (7.192). Aromanian differs here in having simply *ca să* (lit.,'as DMS') for 'in order that.'

7.8.1 Verbless Sentences

Although the textbook definition of a clause typically makes reference to noun phrases and verb phrases, languages do have well-formed meaningful complete utterances that do not necessarily contain an overt verb, and the Balkan languages are no exception. In this section, we survey some Balkan verbless sentences, assessing their Balkanological interest and paying close attention to any contact-related developments.

7.8.1.1 Exclamatory Sentences

Various of the Balkan languages show exclamatory utterances that converge in form. One such convergent form consists of a so-called *wh*-word followed by an NP, such as:

(7.193) a. Τι λεβέντης! (Grk)
what brave.NOM
'What a brave guy (he is)!'

b. Što junak! (Mac)
what hero
'What a hero (he is)!'

c. Sa i bukur muaji i majit (Alb)
how PC beautiful month.DEF PC May.GEN.DEF
'How beautiful (is) the month of May!'

d. Tsi mushuteatsã! (Aro, Cuvata 2009: 3)
what beauty
'How beautiful!, What [a] beauty!'

Another such construction consists of an adverbial plus an NP, as in (7.194):

(7.194) a. Κάτω η χούντα (Grk)
down the the junta.NOM
'Down with the junta!'

b. Dolu Tito[245] (Mac)
downwards Tito
'Down with Tito!'

c. Dolu Živkov (Blg)
downward Živkov
'Down with Zhivkov!'

245 This is a line from a political joke: Tito woke up to find *Dolu Tito* written in urine in the snow in front of his house. The secret police (UDBA) were called in and discovered the urine was that of Aleksandar Ranković (Tito's repressive head of internal affairs who fell in 1966), but the handwriting was Jovanka's (Tito's wife; in the 1970s one could make jokes about Jovanka, but not about Tito). It is interesting to note that at exactly the same time that joke was circulating in Yugoslavia (1973–1974), there was one in the United States with the same topos, but with Nixon instead of Tito, Kissinger instead of Ranković, and the handwriting as Pat's. Needless to say, this vignette deserves unpacking, but our point here is simply linguistic.

d. Poshtë diktatori Enver Hoxha (Alb)
down dictator.DEF E. H.
'Down with the dictator Enver Hoxha!'

e. Tele o diktatori / tele e diktatoreja (Rmi)
downward DEF.NOM dictator / downward DEF.OBL dictator.INS
'Down with the dictator'

f. Jos dictatorul![246] (Rmn)
downward dictator.DEF
'Down with the dictator'

g. Diktatör aşağı (Trk)
dictator down
'Down with the dictator'

Such utterances converge also in lacking an overt verb, and in the languages with case marking, like Greek, Albanian, and Romanian, in showing nominative case, as in (7.194a) with η χούντα (versus accusative τη χούντα) or (7.194d) with *diktatori* (versus accusative *diktatorin*). Here Romani shows vacillation between nominative (*o diktatori*) and instrumental (*e diktatoreja*) marking. However, even with various convergent details in the surface forms, there is probably nothing Balkanologically interesting in these elliptical constructions. Languages outside of the Balkans show something similar; compare the English translations given in (7.194), where the verb can be suppressed, *What a guy!*, *Down with the junta!*, and so on. Such non-Balkan parallels, coupled with the fact that exclamations are part of the natural repertoire of numerous, perhaps all, languages, and that the spelling-out of arguments with understood, elliptical, verbs also is not unusual, this Balkan sentence-type very likely was arrived at independently in each language.

There is, however, one verbless exclamatory sentence-type in the Balkans for which a contact explanation is attractive, namely that involving the modal negator, as discussed in detail in §7.6.2. In particular, one-word prohibitive utterances, equivalent in value to the English exclamatory prohibitive *Don't!*, occur in Albanian, Greek, and Romani with their respective modal negators and in Balkan Slavic with its prohibitive negator; all of these utterances can be translated as 'Don't!' Albanian *mos* and the Balkan Slavic negators *nemoj* and *nedej* also permit overt person marking, 2PL *-ni* for Albanian, 2PL *-te* for Slavic. The Turkish modal particle *sakın*, which is etymologically the imperative of a verb meaning 'beware, take care,' has a closely related meaning of 'don't!,' e.g., *Sakın, sevgilim sakın, beni unutma* 'Don't, my dear, don't, don't forget me' (lit., 'don't my.dear don't me. ACC forget.NEG.IMPV'):

(7.195) a. Mos(ni)! (Alb)
b. Μη! (Grk)
c. Ma! (Rmi)
d. Nemoj(te)! (Mac & BCMS)
e. Nim! (Lower Vardar Mac, a reduction of *nemoj*(*te*))

246 The name *Ceauşescu* was understood in 1989.

f. Nedej(te)! (Blg)
g. Sakın! (Trk, borrowed into Alb *sakën,* Aro *sakăn,* Blg *sakăn,* Mac *sakan*, Megl *săcăn*)[247]

It is argued in §7.6.2 that this convergence, an innovative extension of the use of these negators as prohibitive markers with verbs, spread within the Balkans by contact amongst speakers of these languages, and may have been instigated in the first place by a cross-language analogy, i.e., a sort of calquing, based on a Balkan Romance, probably Aromanian, model.[248]

We can also mention here as a typological parallel the use of BSl *Neka!* 'let' as a one-word verbless sentence expressing some sort of concession.

7.8.1.2 Presentational Sentences with Weak Object Pronouns

There are three types of verbless sentences that, in the Balkans, can usually take weak object pronouns. Of these, two can be grouped together as presentational (on the third type, which is interrogative, see §7.8.1.3). The function of the presentationals is to present to the hearer something for seeing/consideration or taking/undergoing. From a structural standpoint, they consist of a head word, glossed here as *presentational particle* (PPCL) – an element sometimes clearly related to demonstrative elements – plus a noun phrase object, either a pronoun or a noun, which can be called the pivot nominal. The two types can be distinguished by whether the object is accusative/nominative, which we call *visual,* or dative, which we call *receptive*, although there is some overlap between the two types of PPCL, as seen below. Consider here the differentiated ambiguity in the English phrase *Here you are!*, which can be uttered either upon discovery of a location/site or upon offering an item, but not for conclusive utterances, versus *There you are!* which can be used for location/site or conclusive utterances, but whose use in a dative function is at best attenuated. The various PPCLs can occur independently as exclamations, but our interest here is in the similarities evident when they take pronominal objects, *sensu lato*.[249]

247 Despite its verbal origin, *sakın* functions in Turkish and the languages that borrowed it as a prohibitive particle. The Turkish negative imperative is formed with the unstressed suffix *-mA*.

248 A possible contributory factor in this innovation was an inherited use of PIE $*meH_1$ in elliptical prohibitions without a verb. This usage is found in Ancient Greek, e.g., μή μοι σύ 'None of that to me!' (lit., 'μή to.me you.NOM') and in Modern Greek, e.g., μη τα χέρια σου έξω 'Don't put your hands outside!' (lit., 'μη the hands of.you outside'), in Albanian, e.g., *si mos më keq* 'in a lamentable state' (lit., 'how [might] *mos* worse [happen]!'), in Romani, e.g., *ma maj* 'No more!, Enough!' (lit., '*ma* more!), and, importantly, outside of the Balkans in Sanskrit, e.g., *mā śabdam* 'Not a word!' (lit., '*mā* (< $*meH_1$) word'). This geographic distribution guarantees that this pattern is of PIE age, so that it would have been a preexisting pattern that could reinforce the innovation of the use of $*meH_1$ without a verb. See Joseph 2002c for discussion. The Tocharian A usage adduced by Hackstein 2020: 26 (see footnote 77 above), exemplified by Toch. A6b1 *mar täṣ!* 'Not that, don't!,' may be this rather elliptical construction.

249 See §4.3.4.3.2 for the lexical dimension, via borrowing, to the similarities among such particles in the Balkans.

In East Slavic and Polish, as well as non-Alpine South Slavic, there are visual PPCLs of pronominal origin with the meanings 'Look! Here is! There is! Behold!, etc.'[250] A striking feature of these visual PPCLs is that they take accusative objects in Balkan Slavic, (7.196a), genitive objects in BCMS, (7.196c), and nominatives in East Slavic and Polish (7.196bd). Alpine Slavic (Slovene), Czech, Slovak, and Upper and Lower Sorbian, do not have reflexes of these items (see also Schaller 1970):[251]

(7.196) a. Eto go (Blg)
Ete go (Mac)
PPCL him.ACC.WK
'There he is!'

b. Oto on (Pol)
Vot on (Russ)
Vos' yon (Bel)
On vin (Ukr)
PPCL he.NOM
'There he is!'

c. Eto ti Mare (BCMS)
PPCL you.DAT Mary.GEN
'Here's Mary for you!'

d. Oto artysta (Pol)
PPCL artist.NOM
'There's an artist (for you)!'

Aromanian *ia/iatu* and Romanian *iată/iacătă* have similar functions and also take weak accusative objects. Papahagi 1974: s.v. identifies *ia* as a Turkism, itself a borrowing into Turkish from Arabic. Cioranescu 1958–1966: s.v. proposes a native etymology for the Romanian, but Greek για is a possible source too. Çabej 1996: s.

250 The nuances of the differences in deixis between interpretations of the various pronominal stems (*s, v* 'proximal' *t* neutral, and *n* distal') in the various languages are not of concern here.

251 Here it is worth noting that the South Slavic languages with the visual PPCL use the deictic element *e-* while the North Slavic languages with that same PPCL use the element *o-*. Moreover, *e* in Russian and Belarusian was assigned a purely deictic function for the masculine singular neutral deictic, which did not happen in Ukrainian and the rest of Slavic, where the languages have no vocalic particle preceding the deictic. What emerges from this is a complex picture, suggesting a tendency in non-West Slavic, i.e., the period when *dl* simplified to *l* and certain reflexes of the so-called third palatalization were shared by the Slavic dialects that became, respectively, South and East Slavic. The interesting data of the Novgorod birchbark letters complicate the boundaries of East and West, but do not alter the basic picture concerning North and South. The dialects that became South Slavic used *e-*, while those that became North Slavic used *o-*. To this can be added the fact that the Slovene-Czech-Slovak-Sorbian region did not share in the apparent innovation. It is arguable that the visual PPCL + DEM began during the Common Slavic period, with variation between *e-* and *o-*. The Slovene-Czech-Slovak-Sorbian region was, at one time, a continuum that included Pannonian Slavic. The continuum was broken by the pincers of the German (Frankish) and Magyar invasions, from west and east, respectively, from the late tenth to the early eleventh centuries, which destroyed the evidence of Pannonian Slavic except for its survival via loan elements in Hungarian (e.g., M. Greenberg 2004). Finally, we can note that the PPCL has, for some speakers, entered colloquial Slovene usage from BCMS. The native Sln PPCL is *lej*. The majority of authors suppose that it comes from *glej!* 'look.IMPV!' Note that Čakavian and Lower Sorbian have the same particle. There has been speculation that the source is the Latin demonstrative *ille*, but this suggestion has not gained wide support (Bezlaj 1982: s.v.).

v.*ja* discusses the same Greek origin as proposed for Albanian (see below). Such a proposal begs the question of why Balkan Romance uses an accusative while Albanian uses a nominative, as in (7.197). Moreover, in Albanian, use of *ja* as a PPCL is limited to the third person. Otherwise it introduces full clauses, as does Greek γιά (see below). On the other hand, Macedonian visual PPCLs must take either a short pronoun or a reduplicated pronoun – like verbs – rather than simply a strong pronoun as do prepositions. These facts are illustrated in (7.197):

(7.197) a. Ia- l (Aro)
Iată- l (Rmn)
PPCL him.ACC
'There he/it is!'

b. Ja ai (Alb)
PPCL him.NOM
'There he/it is!'

c. *Ja unë (Alb)
PPCL I.NOM
*Eve mene (Mac)
PPCL me.ACC.STR
'Here I am!'

Albanian presentational *ja*, with variants *jate, javo, javua, javota,* are identified by Meyer 1891: s.v. as being ultimately from Turkish *ya*, which, when used presentationally would take the nominative. Çabej 1996: s.v., however, citing Hatzidakis 1916, derives ModGrk γιά as in ModGrk γιά νά δούμε (= Alb *ja të shohim* 'hey, let's see') from the AGrk exclamation εἶα 'on! up! away!' He derives *jate* from *ja te(k) është* 'PPCL where it.is,' and Gjirokastër Tosk *jato* from the addition of the Greek neuter weak object pronoun τό. One problem with the Greek etymology, however, is that Grk γιὰ is not presentational but rather vocative and functions much more like the polysemous Turkish *ya* when it means 'Oh!, Hey!' rather than 'Behold!'

Before turning to Greek, we can note that Romani has the native presentational particles *eke, ake* (PROX, DIST), which, in the first two persons combine with personal pronouns and in the third with a copular form inflected for gender, e.g., *ake-tan-o* 'there he is! behold him! (lit., 'there-is-M').[252] In Macedonian Arli (Topaanli), the fact that the Macedonian 1SG.ACC weak form *me* is homonymous with the Romani 1SG.NOM *me* 'I' probably influences the fact that the presentational for 'here I am' can be *ake me* (1SG.NOM; cf. Mac *evo me* 1SG.ACC) or *ake man* (ACC).

Finally, Greek and Balkan Romance share a development in which the PPCL is *na* (Grk νά) + ACC, as in (7.198):

(7.198) a. Νά τον (Grk)
PPCL him.ACC.SG.WK
'Here he is'

252 In Kalderaš, 1st and 2nd person endings can be attached. A variant with *-l-* (*-talo,* etc.) also exists. See Boretzky & Igla 1991: 405–406; Boretzky & Igla 2004: 1.91.

b. Νά τον μουσακά (Grk)
PPCL DEF.ACC.SG moussaka.ACC.SG
'Here's the moussaka'

c. Na-l (Aro)
PPCL-him.ACC
'There he is'

Albanian also has *na* + NOM, but it is either receptive or interjectional and is discussed below. In all of Slavic, *na* is a receptive PPCL taking the dative – see (7.200acd) below – rather than a visual PPCL as in in (7.196) above. The visual PPCL use of *na* also occurs in southern Italian (Papahagi 1974: s.v.) as the equivalent of standard Italian *ecco* 'behold.' Here the influence of Griko is probably significant.

Besides the formations of the type νά + ACC, Greek, like Albanian, can also have the nominative as the pivot nominal in conjunction with the visual PPCL, and this is even favored by some speakers, as in (7.199):

(7.199) Νά ο Γιάννης (Grk)
PPCL the.NOM.M John
'Here's John!'

We now come to the receptive PPCL and receptive uses of the visual PPCL. The word *na* occurs in the receptive function in most of Slavic and in Albanian and Turkish.[253] It has a dative presentational meaning of the type 'Here! Take it! Here/There you go!,' and pragmatically it can have a broad range of nuances from polite to rude, depending on language and context. Contextually, Greek and Balkan Romance can also use *na* (νά) in this function, although in those languages, this is a variant meaning rather than a basic meaning (see (7.201–7.203) below) and involves the interface between ethical dative and indirect object dative. In Balkan Slavic, including here BCMS, as well as Balkan Romance, a weak dative pronoun can be used, but not in Slovene. East Slavic, like Turkish, has only strong datives. Albanian, however, uses a nominative. The PPCL *na* is absent from Romani, where (the etymologically distinct) *na* is the primary negator, and lexical *le* 'take. IMPV.2SG' is used for such meanings. Like the visual PPCL, receptive *na* can take weak pronominal objects in various Balkan languages, but not in Slovene, and a dative pronoun elsewhere in Slavic as in (7.200):

(7.200) a. Na ti, daj mi (Mac, Blg, BCMS)
PPCL you.DAT.WK give.IMPV.2SG me.DAT.WK
'I'll scratch your back if you'll scratch mine'
(lit., 'Here's for you, [now] give to me').[254]

253 Thus, for example, it appears to have been current in Upper and Lower Sorbian in the nineteenth century but has since become obsolete.

254 In Bulgarian, the order can also be *daj mi, na ti*. The expression *na ti, Vranke, kokalče, daj mi, Vranke, železno zabče* 'Here you [are], Little.Crow, [a little] little.bone, give me, Little.Crow, [an] iron tooth' is also used in Bulgaria, and, *mutatis mutandis*, throughout the Balkans and beyond, when a child loses a baby tooth. The tooth is thrown on the roof and the child asks for a strong replacement.

b. Na librin tënd ti. Merr-e! (Alb)
PPCL book.ACC.DEF your.2SG.ACC.POSS you.NOM take.IMPV.2SG-it
'Here's your book, you. Take it!'

c. Na (*ti) knjigo (Sln)
PPCL you.DAT.WK book.ACC
'Here's/Take the book'

d. Na tebe (knjigu) (Russ)
Na tabe (Bel)
Na tobi (Ukr)
Na sana (Trk)
PPCL you.DAT (book)
'Here you are! [offering], Take that!, So much for you!, etc.'

On the other hand, the visual particles cited in (7.200) can also be used by themselves or with a dative pronoun to mean 'Here, take!' as well as various pragmatic extensions that express surprise, interrogation, disbelief, dismissal, or confirmation, depending on context. In Greek, this use of νὰ is equivalent to ορίστε 'here you go [IMPV.2PL],' but Greek νά cannot occur with a bare dative (genitive in form) of the Slavic type seen in (7.200a) above. The examples in (7.201)–(7.203) are illustrative:

(7.201) a. Eve ti! (Mac)
b. Na- ts (Aro)
c. Eve ti go (Mac)
d. Iată- vă- l (Aro)
e. Iată- vi- l (Rmn)
f. Na- vă- l (Aro)
g. Νὰ σου το (Grk)
PPCL you.DAT it.ACC
'Here you are [offering, not discovering], here you go, etc.'

(7.202) Eto ti pătja (Blg)
PPCL you.DAT road.DEF.ACC
'Go hit the road' (lit., 'there is the road for you')

(7.203) Na- ți- o [bună] (Rmn)
PPCL you.DAT INDEF.F [good.F.INDEF]
'Here's a fine howdy do!, Not at all!, You don't say!, Seriously?, Well, I never! Well I declare! Of course!' (cf. Bantaş et al. 1981: 353; cf. the nuances of Eng *Will you look at that!*)

Here it is worth noting that in North Slavic, the presentational always takes a nominative object, but a 2SG ethical dative is possible, as in (7.204):

(7.204) Vot tebe knjiga (Rus)
PPCL you.DAT book.NOM
'Here's a book for you!'

Owing to the nature of the expression, native Russian speakers find 2PL strange, and non-second person wrong.

Finally, in our survey of noninterrogative PPCL usage, we can note that at least in Slavic, visual and receptive PPCLs can co-occur, as in (7.205):

(7.205) a. Eve ti go, na (Mac)
b. Evo ti ga na (BCMS)
PPCL you.DAT it PPCL
'Here you go, take it'
c. Vot tebe (i) na (Russ)
d. Vos' tebe (i) na (Bel)
PPCL you.DAT and PPCL
e. Ot tobi na! (Ukr)
PPCL you.DAT PPCL
'Well I declare!, Well I never!, Well of all things!, etc.'

Under an analysis advanced by Joseph 1981, 1994a for Greek νά, the PPCL *na* can be treated as a verb, specifically an imperative, in sentences with an accusative (Grk, BRo, Alb) or dative (in general when ethical dative is included) pivot nominal. The fact that in some languages, *na* can host the marker of verbal plural imperatives, e.g., *nate* (dialectal Grk, Rus, Ukr, Cz, Sln), Bel *nace,* Pol *nać(e)*, Lower/Upper Sorbian *naśo/naće* (PL), *natej/nataj* (DU), and Alb *nani,* is evidence in favor of this analysis. Both Greek νά and the Slavic and Balkan Romance visual PPCLs take clitic objects the way a verb would. While this analytic step would mean that these presentationals are not, strictly speaking, verbless constructions, their discussion is still relevant here.

Presentational particles can be used independently (Νά!, *Ete!,* etc.), but if they occur with a pronominal element, the pronominal must follow, as is the case with imperatives, for instance (cf. examples 7.198ab) *τον μουσακά νά and *τον νά are impossible in Greek, and the same holds in the other languages, e.g., *me eve (Mac), *kniga eto (Blg), etc.[255] When *na* is used independently (except in Greek), however, it can follow a visual PPCL + object, as in (7.205).

In Greek, the presentational element νά does not occur in the outlying Romeyka dialect of Pontic still spoken in Turkey in the Trabzon area (Ioanna Sitaridou, p.c., 2010), and neither does the presentational construction with a nominal, a geographic distributional fact which also points to a Balkan locus for the accusative grammar of PPCLs.

The use of the nominative with PPCLs in Albanian and Greek may or may not be related. If Albanian *ja* is from Turkish, a nominative is unsurprising. The nominative with *na* in its receptive meaning, however, is strange and may show Greek influence, a possibility that requires some detailed discussion and

255 As suggested by the discussion in §7.4.1.2.2.2.1 concerning examples (7.27) and especially (7.28), it can be noted that for Greek, at least, the word order facts are explicable if PPCLs are imperatives, as imperatives are typically utterance-initial and when weak object pronouns are involved, they always follow their imperatival host. This same argument works for Slavic and Romance but not for Albanian or Turkish.

adds additional dimensions to this particular clause-type that are of Balkanological interest.

In Greek, the nominative could have its source in a reanalysis of neuter nouns controlled by the presentational particle, since, on the one hand, neuter nouns are ambiguous between nominative and accusative and on the other, the semantics of a presentational sentence allow an interpretation of the pivot noun as a nominative.[256] That is, a presentational sentence not only presents and brings into view an entity but it also demonstrates and thus establishes the existence of said entity. In Balkan languages, as discussed more fully in §7.8.2.2.6, the accusative case is found on the pivot nominal in existential sentences, as in (7.206), suggesting it is treated as an object (at some level of analysis); importantly, though, in other existential constructions, with different existential predicates, the pivot can be in the nominative case, treated as if it were a subject, as in (7.207):

(7.206) a. Έχει τρεις τόνους σ' αυτή τη γλώσσα (Grk)
has.3SG three tones.ACC in this the language
'There are three tones in this language'

b. ¿Hay un libro aqui? Si, lo hay (Sp)
there-is a book here yes it.ACC.SG there.is
'Is there a book here? Yes, there is.'

(7.207) a. Υπάρχουν τρεις τόνοι σ' αυτή τη γλώσσα (Grk)
exist.3PL three tones.NOM in this the language
'There are three tones in this language'

b. Na stole jest' kniga (Rus)
on table is.3SG book.NOM
'There is a book on the table'

Thus, the semantic affinity of existentials with nominative and existentials with accusative could well have aided in the (abductive) reanalysis of neuter accusatives as nominatives in a Greek presentational such as (7.208):

(7.208) Νά το παιδί μου (Grk)
PPCL DEF.ACC/NOM.N child.ACC/NOM.N me.GEN
'Here's my little boy!' (said by a mother as she wakes her child for school)

It is assumed here that accusative was the earlier syntax of at least the νά presentational (justified further below), so that (re-)analysis of the pivot nominal as a nominative would have allowed for an innovative overtly and unambiguously marked nominative such as ο Γιάννης, as in (7.199) to occur.

Moreover, Greek has further innovated a special set of weak nominative pronominal forms, answering pronominally to the weak accusative pronoun just as nominative nouns answer to the accusatives in this construction; only the masculine

256 This account is based on Joseph 1981, 1994a, To appear a, though its basic outline draws on suggestions by earlier linguists, e.g., Thumb 1912.

Table 7.5 *Weak nominative pronouns in Greek*

M.SG	τος	F.SG	τι	N.SG	το
M.PL	τι	F.PL	τες	N.PL	τα

singular is distinct from the accusative (and for some speakers, the feminine plural as well) (Table 7.5).

These innovative Greek forms are mostly restricted to the presentational construction and a corresponding locative interrogative construction with πουν 'where is/are?,' etymologically a combination of πού 'where' with a reduced form of είναι 'is, are.'[257] Moreover, they appear to have developed through an analogy between the νά + accusative construction and the νά + nominative construction with the relationship between strong pronominal forms and weak pronominal forms and accusatives and nominatives being pivotal:

(7.209)

νά αυτόν	:	νά τον	::	νά αυτός	: νά	X, X => τος	
ACC.STR		ACC.WK		NOM.STR			NOM.WK

What makes these forms Balkanologically interesting, from the point of view of noting key divergences in the midst of convergent phenomena, is that, as argued by Joseph 2002f, this pronominal type appears in Greek, and, probably independently, in Romani.[258] Only Greek and Romani have an overt and transparent morphological relationship between strong and weak third-person oblique pronouns (e.g., 3SG.M Grk αυτόν/τον, Rmi *ole/le*) and, for Greek, also between strong nominative and accusative forms (e.g., 3SG.M αυτός/αύτόν) that could lead to the extension of those relationships into nominative forms; such relationships are missing from all of Slavic (a situation going back to Common Slavic), while in Albanian and Balkan Romance as well as Romani, the relations of weak to strong and nominative to accusative are simply less transparent, as seen in Table 7.6.

257 See §7.8.1.3 for a further parallel involving a related construction. As discussed in Joseph To appear a, the occasional instances of τος outside of the presentational and interrogative contexts mostly involve the verb 'be' and bear on the origin of these forms and particularly of πουν (inasmuch as it incorporates a form of 'be' in it), though some instances in verbless sentences, e.g., κάτω τος 'He's down,' suggest movement towards wider use of these pronouns.

258 Matras 2002 has some discussion of weak subject pronouns in Romani dialects, both in the Balkans and beyond, and he notes the presence of third-person subject "clitics," based on a stem *l-*, in these dialects. He considers these subject clitics to be an old feature of Romani. In some dialects, they are restricted to occurring with presentationals, existentials, and interrogatives, a distribution rather like that found with the Greek weak nominative forms like τος, raising a natural question as to the relationship, if any, between the Greek developments and the Romani phenomenon. In this case, though, the Romani dialects with these restricted forms include some in the Near East, so that these forms most likely arose within Romani in a pre-European stage. Greek would then still be alone as having innovated such forms within the Balkans. Moreover, even if contact with Romani was involved in the Greek innovation, since all of the Balkan languages were involved in some contact with Romani, it is fair to ask why no other Balkan language reacted in the same way as Greek did to contact with Romani. The language-particular array of relations among the demonstratives provides that answer. The exact dating of the emergence of τος is not entirely clear, but it seems clearly to be a Medieval Greek innovation; for discussion, see Joseph To appear a.

Table 7.6 *Strong versus weak third-person pronouns in the Balkans*

	Case/Number/Gdr	Strong	Weak
BCMS	NOM.SG.M	*on*	—
	ACC.SG.M	*njega*	*ga*
Blg/Mac	NOM.SG.M	*toj*	—
	ACC.SG.M	*nego*	*go*
Alb	NOM.SG.M	*ai / ky*	—
	ACC.SG.M	*(a)të / këtë*	*e*
Rmn	NOM.SG.M	*el*	—
	ACC.SG.M	*el*	*îl / l*
Rmi(Arli)	NOM.SG.M	*ov*	*-l*
	ACC.SG.M	*ole(s)*	*le(s)*
vs.			
Grk	NOM.SG.M	αυτός	(τος)
	ACC.SG.M	αυτόν	τον

Thus, language-particular details in Greek and Romani drove them along their own path of development, so that they came to be differentiated from the other languages that showed a different language-particular array of facts and relationships.

The element νά thus figures in the Greek realization of this Balkan presentational pattern, with weak pronominal forms, mostly accusative – the most Balkan type – but, along with Albanian and Romani with regard to PPCLS innovatively nominative as well, serving as presentational arguments. The etymology of νά is disputed and probably has a contact basis, and it may provide an example of the syntactic properties of a lexical item being borrowed along with the item itself.[259] The nativist proposal for a Greek-internal source for νά offered by Hatzidakis 1905: II.100, 400, and endorsed by Andriotis 1983: s.v. and Babiniotis 2010: s.v., derives it from an adverbialization (with the productive adverbial suffix *-a*) of a particle ἠνί extracted out of the sequence of two Ancient Greek deictic elements, each meaning 'behold, look!,' ἦν and ἴδε, this latter being formally the aorist imperative of 'see.'[260] The sequences of the two forms do occur, but there is no independent evidence for ἠνί nor any reason why it would be treated as an

259 See §5.5.3 (and especially footnote 186)) regarding Trk *de* and its postpositive syntax being replicated in the languages that borrowed it; so also with other Turkish postpositional elements such as *gibi* 'like,' borrowed as a postposition in Ottoman-era Edirne Greek, and *karşı* 'opposite,' borrowed as a postposition in Aromanian (as *carshi*; see §4.3.3.2).

260 We can note in passing attempts by Christidis 1985, 1987, 1990 to link the deictic νά with the modal subordinator να, which is indeed of native origin, deriving from the earlier Greek subordinating conjunction/complementizer ἵνα 'so that.' Joseph 1981 draws attention to some key points of difference between νά and να, similarities which Christidis downplays or interprets differently.

adverbial; importantly, it takes arguments and thus seems more like a predicate than a modifier.[261] Thus a contact-based etymology (i.e., as a borrowing), should be considered. The most obvious possible source is Slavic *na,* although a nativist version of an Albanian etymology can also be proposed. As noted above, the receptive PPCL *na*, meaning roughly 'take!; here!,' occurs throughout Slavic with the syntax of presentational elements noted above in (7.201–205), so a South Slavic source for νά, in and of itself, and further for νά τον, becomes attractive. The same argument applies to Balkan Romance *na*. Alternatively, one could look to Albanian for a source, with νά (and thus maybe Balkan Romance *na*) being from a zero-grade imperative **nm̥* from the root *nem 'take' (as in Germanic, cf. Goth *niman*), which is taken as a cognate of Slavic *na* (Vasmer 1953–1958: s.v.; Skok 1972: s.v.; BER V: 1996: s.v).[262] For Albanian, one could posit the incorporation of the pronoun *e* as direct object, so that **në* + *e*, either directly or as the contraction product *na*, but with an understood object, would give the basis for Greek νά τον, with Greek's own pronoun pressed into service.[263] Çabej 2002: s.v., however, does not propose any such etymology and points only to the fact that *na* is found throughout both Slavic and the Balkan languages. A Balkan contact-related etymology is consistent with the absence of PPCL *ná* from Pontic Romeyka Greek of the area around Trabzon, as noted above. From the perspective of Balkan linguistics, what is relevant here is that under either the Slavic or the (admittedly less likely) Albanian scenario, contact is responsible both for the form of νά/*na* in Greek and Balkan Romance, respectively, and for the syntax that gave the starting point for the Greek-internal innovations leading to nominative forms.[264]

A nonpresentational verbless sentence with an object pronoun occurs in a couple of the languages, as illustrated by Macedonian *Neka go* 'let him [do X].' Such usage does not occur in Bulgarian, but there is a possible parallel in Greek, in, for instance, Ασ' τον 'let him be; go ahead!' and similar constructions with an object and a DMS phrase, e.g., Ασ' τον να μιλήσει 'let him [DMS] speak!'; admittedly the Greek form here could be analyzed synchronically as a reduced form of άσε, an imperative of αφήνω 'let, allow,' which (see §6.2.4.3.3, footnote 339) is the diachronic source of hortative ας.

261 And, as Joseph 1981: 141 remarked, this is "a very complicated etymology for what appears to be a very simple word."

262 Compare zero-grades of roots in imperatives in Ancient Greek, e.g., ἴ-θι 'go!' (root **H₁ei-),* and Sanskrit, e.g., *kṛ-dhi* 'do!' (root **k^{w}er-*); one has to assume that the imperatival ending **-dhi* did not survive into Albanian.

263 Thus possibly also with the contraction with the pronoun seen in combination with the subordinating *të* which ends up as *ta* with a following accusative weak object pronoun (i.e., *të* + *e* => *ta*).

264 The borrowing of other deictic-like particles in the Balkans, such as Alb *xa* 'here you are!' (phonetically [dza]) as the likely source of Greek τζα, whether pronounced [dza] or [tsa], provides some support for the suggestion here that Greek νά, etc. could have a contact source. Given that this particle also occurs in Macedonian, the routes of transmission could be complex. See §4.3.4.3 on such loanwords.

7.8.1.3 WHERE Questions with Accusative

Besides the sentences, presentational and otherwise, of §7.8.1.2, some of the Balkan languages also show interrogatives with a locative sense, i.e., questions with 'where,' that occur without a verb, and with just the question word and a focalized nominal (the "pivot," as in §7.8.1.2) as the subject. Of interest from a Balkanological standpoint is the fact that in both Bulgarian and Greek, there is a locative interrogative construction with 'where' and a pronoun as the pivot where the pronoun, even though ostensibly the subject of the question, occurs in the accusative case form in both languages, and, in Bulgarian but not in Greek, a proper noun could also be used in the accusative in such a construction:

(7.210) a. Kăde go (be)? (Blg)
where him.ACC VOC

b. Πούν τον ([don]) (Grk)
where.is he.ACC
'Where is he?'

c. De Baj Ganja? (Blg; Konstantinov 1895)[265]
where B. G.ACC
'Where is Bai Ganyo?'

d. *πούν τον Μπάϊ Γκάνιον (Grk)
where DEF.ACC B. G.
'Where is Bai Ganyo?'

e. Kăde go (toja) Ivan? Eto go Ivan! (Blg)
where him.ACC (DEM) I. PPCL him.ACC I.
'Where is Ivan? There is Ivan!' (standard peek-a-boo in Bulgarian)

In (7.210ace), *go*/*Baj Ganja* are the object forms of the pronoun/proper name, cf. *Az go znam* 'I know him' (lit., 'I him know'), and in (b), τον corresponds to the object form, as in τον ξέρω 'I know him,' occurring with a voiced [d] due to the preceding nasal (see §5.4.4.1). Diachronically, the nasal in πούν represents a reduced form of the verb 'to be' (synchronically είναι, earlier ένι) but from a synchronic standpoint it is nothing more than a piece added to the question word πού in this construction.

Given the oddity of an objective form of a pronoun occurring as what should be a subject of the verb 'be,' one might suppose that this construction originated in just one of the languages and entered the other as a calque (via cross-language "copying"). However, the fact that Bulgarian has the construction with proper names while Greek does not is a complicating factor here. For Greek, the occurrence of an object form like τον has a potential internal explanation.

Just as it appears that a proportional analogy involving the νά construction gave rise to the innovative nominative weak object pronoun in Greek, as suggested in §7.8.1.2 (νά αυτόν : νά τον :: νά αυτός : νά X, X => τος), so too can a similar analogy explain the accusative with πούν. That is, one can speculate that once Greek innovated a nominative weak object pronoun with νά, that type could spread

265 Today, educated Bulgarian speakers consider such usage archaic or dialectal.

to the semantically similar locative interrogative predicate, πούν 'where is?'. Then, since Greek had both nominative and accusative weak object pronouns with νά, and nominative weak object pronouns with πούν, a simple proportional analogy gets the accusative innovatively occurring with πούν, as outlined in (7.211):

(7.211)

νά τος:	νά τον	::	πούν τος	:	πούν X, X => τον
NOM.WK	ACC.WK		NOM.WK		ACC.WK

However, since Bulgarian lacks a weak subject pronoun (see §7.8.1.2), this explanation has no relevance for Bulgarian. One could suppose that the type of (7.210) is an analogical creation within Bulgarian based on the semantically similar presentational construction (*eto go*, discussed in §7.8.1.2), but the expansion to (or occurrence with) proper names is unexplained. This raises the possibility that the construction originated in Greek, where there is the plausible language-internal source outlined in (7.211), and that its occurrence in Bulgarian is a calque on the Greek syntax. The difference between Greek and Bulgarian with regard to proper names would have to be attributed to a Bulgarian-internal extension from one type of nominal (i.e., a pronoun) to another type (i.e., a proper noun) or a copula-less use of *de* with a subject proper noun.[266] While admittedly speculative, such an account – drawing on both language contact and language-internal developments – explains the facts and provides a natural basis for both the parallels and the differences between Greek and Bulgarian that are evident here.

7.8.2 Subjectless Sentences

Just as the previous section examined sentence-types that are noncanonical in lacking a verb, one of the two parts that traditionally define a sentence, so too are there types that are noncanonical in lacking the other part, specifically an overt subject. The Balkan languages show structural convergences in this domain as well and these need to be considered here. The parallels involving the absence of a subject fall into two main classes: sentences involving the absence of an explicit pronoun in the presence of morphological person marking (so-called pro-drop, itself an anglocentric term) and impersonal sentences, this latter subsuming a number of different individual types.

7.8.2.1 Absence of Explicit Subject Pronoun (Pro-drop)

All of the Balkan languages are so-called pro-drop or null subject languages (cf. Perlmutter 1970; Jaeggli & Safir 1989), a characterization based on the assumption that explicit pronominal or nominal subjects are "normal" as opposed to the fact that in many languages a well-formed sentence need not

266 A fact that might be relevant here is the possibility of accusative in at least one verb-less sentence-type in OCS, namely *tako mi bogy* 'By the gods!,' a variant with accusative of the same phrase with nominative (seen in (7.36d)); *bogy* is accusative plural matching the Greek original (μὰ τοὺς θεούς 'By the gods!,' with accusative τοὺς θεούς).

have an overt nominal, i.e., neither a noun nor a pronoun, as the subject, so that the verb (marked for the subject, and in some languages much more) by itself can determine a sentence. Thus, while a pronominal (or, contextually, nominal) subject can occur overtly, especially under conditions of emphasis or focus, that nominal can also be omitted with no effect on the viability, i.e., the syntactic acceptability, of the sentence, as in (7.212) and (7.213), where the sentences in the former mean 'We [are the ones who] laugh' while those in the latter mean simply 'we laugh':[267]

(7.212)	a. Εμείς γελάμε	(Grk)
	b. Ne qeshim	(Alb)
	c. Nie se smeeme	(Mac)
	d. Nie se smeem	(Blg)
	e. Noi rîdem	(Rmn)
	f. Noi arãdim	(Aro)
	g. Amen hasasa/hasaja	(Rmi)

(7.213)	a. Γελάμε	(Grk)
	b. Qeshim	(Alb)
	c. Se smeeme	(Mac)
	d. Smeem se	(Blg)
	e. Rîdem	(Rmn)
	f. Arãdim	(Aro)
	g. Hasasa/Hasaja	(Rmi)

The absence of subject pronouns is found in numerous languages, non-Indo-European as well as Indo-European. Moreover it is unlikely that this is Balkanologically significant. Pro-drop is almost certainly a common inheritance from Proto-Indo-European, since all of the branches of Indo-European, at one stage or another in their history, at least in some class of sentences, show this feature.[268] Thus language contact is irrelevant to the occurrence of this feature in any of the Balkan languages.

Nonetheless, pro-drop is characteristic of the Balkan languages and enhances the sense of structural parallelism among them as far as syntax is concerned. Moreover it is connected with the general clausal structure of Balkan languages; as argued in §7.4.1.2 above, the central element in the Balkan clause is the verbal complex, and the nonobligatory nature of the expression of subjects outside of the verbal complex, especially since the verb encodes the subject in the person/number endings, gives that

267 Balkan Slavic requires the INTR marker *se* with this particular verb.

268 The languages that are generally non-pro-drop languages, such as Modern French or Modern English, have developed this characteristic within their attested history. French is non-pro-drop while Latin was a pro-drop language. One could therefore argue for a syntactic characterization of such languages as pro-insert rather than pro-drop to better reflect their histories as opposed to an Anglocentric bias. In Hittite, the presence of pro was conditioned by verb type: the weak subject pronouns are found in sentences with a subclass of intransitive verbs, specifically, with unaccusative verbs, so such clauses could be considered pro-insert if PIE is assumed to have been pro-drop.

much more importance to the verb and its various modifying elements. The prevalence of pro-drop also has an important consequence for sentential word order in that it means that verb-initial structures in these languages are frequent and thus lack the marked status that they may have in other languages.[269]

7.8.2.2 Impersonal Constructions[270]

So-called impersonal constructions are traditionally taken to be those that in a certain way lack an overt subject or specification of a subject, or that lack a fully referential subject. This somewhat vague characterization requires some deconstruction. Referentiality is important because there are constructions with overt but nonreferential subjects, most usually referred to as "expletive" subjects, that typically are subsumed under the rubric of "impersonals." For instance, *it* in English expressions like *It is raining* is overt and is a subject,[271] but does not refer to any sort of specific entity in the real world or, for that matter, in any possible world. The qualification "in a certain way" is necessary because, typically, impersonals appear to be subjectless, lacking even an expletive, as in Macedonian *Vrne* '[it] rains (lit., 'rains') but they differ systematically from pro-drop (null-subject) sentences (see §7.8.2.1) in a key way. In the pro-drop construction, in languages that distinguish on the verb such categories as person and number of the subject, the subject is specified and identified by that marking on the verb; by contrast, in an impersonal construction, the verb is typically in the third person, and most usually third person singular, and thus without any overt specification of the subject.[272]

Among the Balkan languages, there are convergences involving a number of impersonal sentence-types, and it is useful to differentiate among them, as the extent and source of convergence differs from type to type. Some may represent inherited impersonal constructions, and thus are not Balkanisms; others, however, do seem to represent convergence resulting from language contact and may therefore be Balkanisms in the sense in which we use the term here.

Drawing on, though adapting somewhat, Guentchéva's 2010 excellent exposition and classification of Balkan impersonals,[273] we recognize here six distinct main types of impersonal constructions in the sense just developed that can be discussed and evaluated areally with regard to contact-induced syntactic

269 See also §7.8.1.2 on subject pronouns in certain presentational and interrogative constructions in Greek.

270 In this section we draw heavily on Friedman & Joseph 2018, where the matter of Balkan impersonals is approached from a somewhat different angle but offering the same basic classification and data.

271 The *it* in *It is raining* behaves like a subject in that it is copied in tag questions (*It is raining, isn't it*), can be raised with verbs like *seem* (*It seems to be raining*), and so on.

272 There are other possible scenarios regarding features and properties of the subjects of impersonals that we cannot address, especially the suggestion of Perlmutter & Moore 2002 that there can be an expletive that controls agreement but has no phonological content; discussing such theoretical issues is beyond the scope of the presentation here.

273 Bauer 2000 offers an overview of some of the relevant (pre-)historical background for the classical languages.

convergence; to these we add some functionally distinctive subcategories within a few of these types. Further, there is also a seventh type of construction, namely the use of an imperative in a nonimperative function, that, although qualitatively different from the six "classic" types, can nonetheless be considered here since the verb itself does not give evidence of an overtly specified subject. This "narrative imperative" behaves like a nonfinite form in that the interpretation of its subject depends on syntactic and pragmatic context, but is arguably impersonal, lacking a morphological specification of a subject, despite the way it is interpreted.

The six "classic" impersonal types to be treated here, along with relevant subcategories are the following:

(a) atmospheric and natural phenomena
(b) experiencer constructions
(c) impersonal constructions with corresponding personal forms
(d) impersonal passives
    (i) speaking
    (ii) generalized activity
    (iii) potential
(e) impersonal modals
    (i) internal disposition
    (ii) modality of possibility and necessity
(f) 'have' existential versus 'be' possession

These are taken up in turn in the subsections that follow, with some general conclusions about impersonals in the Balkans at the end. Narrative imperatives are discussed in a separate section after that.

7.8.2.2.1 Atmospheric and Natural Phenomena

Expressions for atmospheric and natural phenomena in the Balkans show considerable diversity within genealogical Balkan language groups, as well as commonalities that cross genealogical lines. Whether or not the commonalities are contact-induced turns out to be difficult to judge, since it can be argued that any given verb of weather is inspired by nature rather than nurture. Moreover, a given language can have more than one expression, especially for degrees of intensity. Still, the developments seen in the Balkans, if one focuses on the most common expressions, are striking, and suggest a variety of interesting Balkan specificities.

For verbs of raining and snowing, Grk βρέχει and χιονίζει and Rmn *plouă* and *ninge* are noteworthy in that they preserve inherited specialized verbs translatable by the English verbs 'rain' and 'snow,' respectively. SDBR, i.e., Aromanian and Meglenoromanian, however, have each innovated independently: Aromanian uses *da* 'gives' and Meglenoromanian uses *meardzi* 'goes' plus respective nouns for 'rain' and 'snow.' In the Balkans, 'go' is not otherwise commonly used (although it can be encountered; it also occurs outside the Balkans, e.g., in Russian), and the only other language using 'give' is Romani, the verb being *del*, where, interestingly

enough, the subject is understood as *Devel* 'God,' as was the case in Sanskrit and Ancient Greek. The Balkan Slavic languages are striking in their innovation and diversity in this regard. Except in some dialects, none of them preserve the Common Slavic verb for 'rain' attested in Old Church Slavonic, *dъžditi*. Both Bulgarian and Macedonian have innovated weather-specific impersonal verbs that could be translated by English 'precipitate' since they can refer equally to rain and snow. In Bulgarian the verb is *vali* from an earlier meaning 'roll,' while Macedonian has a different verb, *vrne*. In BER I:211, it is speculated that *vrne*, which is attested in some so-called secret languages in Bulgaria, is a combination of *vali* and *răm*i 'drizzle,' but given the choice of 'roll' in Bulgarian, it seems reasonable to speculate that a perfectivization of *vrti* 'turn' with *-ne* could have produced **vrtne* whence *vrne*. In such a case, the Balkan Slavic speech area arguably shared a semantic shift, but with different lexicalizations. We can also note here that like Bulgarian and Macedonian, Turkish has a single specialized verb for raining and snowing, Standard Turkish *yağ-*, which, however, derives from 'pour, saturate' and also is the basis for the Turkish noun *yağmur* 'rain' (cf. *kar* 'snow').

Further to the north and west where Slavic is spoken, rain and snow 'fall' (e.g., BCMS *kiša pada* 'rain falls'), which is what they do in Albanian as well (e.g., *bie shi* 'falls rain'). While it is true that 'fall' is, quite literally, a natural verb to choose to describe what rain and snow do, it is nonetheless striking that in South Slavic, it developed precisely in that part where Albanian was spoken, keeping in mind that prior to 1878 mixed Slavic-Albanian populations extended east past the South Morava River and as far north as Aleksinac and into Bosnia. Then again, the Romanian borrowing of Slv *zăpadă* 'snow' (alongside inherited *nea*), for which the verb is native *cădea* 'falls,' points to a more widespread usage.

For expressions of cold weather of the type that in English are rendered *it is cold*, in the Balkans the two principal verbs used are 'be' and 'make, do.'[274] Both are well attested outside the Balkans as well, but their distribution within the Balkans is nonetheless suggestive. In Greek and Albanian, the verb of choice is 'make' (e.g., κάνει κρύο and *bën ftohtë* for 'it is cold'), in Bulgarian, Macedonian, Meglenoromanian, Romanian, and Romani, the verb 'be' is used (i.e., *studeno e*, *ladno e, iasti frig, este frig, šudro i[si]*, respectively, for 'it is cold').[275] Aromanian, however, straddles the two zones linguistically as well as geographically in this respect, having both types of expressions, i.e., both *fatsi arãstimi* and *easti arãstimi* or *arãstimi-i* for 'it is cold.' Given the use of 'make' elsewhere in Romance, it would appear that we have here a west-east divide, with Greek and Albanian representing the west, Slavic the east, and Balkan Romance influenced by Slavic, except Aromanian, which gives evidence of contact with both.

274 Other possibilities such as the Macedonian specialized verb *studi* 'it's [freezing] cold' are outside the scope of our consideration here, since the semantic parallels of auxiliary expressions are more readily comparable.

275 Turkish is also a 'be' language in this respect.

There is also an interesting difference between Macedonian and Bulgarian, as observed by Guentchéva 2010. Although Bulgarian expressions of the type *vali* 'it's raining' or *studeno e* 'it's cold' are generally viewed as impersonal and therefore subjectless, for expressive purposes the expletive neuter pronoun *to* can be used, as in (7.214) and (7.215):

(7.214) Abe to naistina valjalo!
wow it really rained.LPT.N
'Wow it really rained' (is raining)'

(7.215) To naistina e studeno!
it really is cold.N
'It really is cold'

In Macedonian, however, the equivalent sentences are not acceptable: *Abe toa navistina vrnelo / *Toa navistina e studeno. There is thus an interesting difference here in a structural detail, with Macedonian being stricter about the possible presence of expletive subjects than Bulgarian, and thus farther from a Slavic turn of phrase, since an expletive subject of this type is found sporadically in other Slavic languages.[276]

The comparative results are summarized in Tables 7.7 and 7.8. While Balkan expressions of weather do not show the across-the-board commonalities seen in other areas of idiomatic expression (see §4.3.11), nonetheless the differential distributions are instructive and some of the familiar sub-areal connections are to be found. For instance, on one feature or another, Bulgarian and Macedonian converge but with a key feature distinguishing them, and Greek and Albanian

Table 7.7 *Expressions meaning 'it is raining/snowing'*

	Greek	Romanian	Macedonian	Bulgarian	Albanian	Meglenoromanian	Aromanian	Romani
rain	βρέχει	plouă	vrne dožd	vali dăžd	bie shi	meardzi ploaiă	da ploai [e]	del biršim
snow	χιονίζει	ninge	vrne sneg	vali snjag	bie borë	meardzi neauă	da neauă	del iv

Table 7.8 *Expressions meaning 'it is cold'*

	Greek	Albanian	Aromanian	Romanian	Meglenoromanian	Macedonian	Bulgarian	Romani
is cold	κάνει κρύο	bën ftohtë	fatsi/easti arătsimi~arătsimi-i	este frig	iasti frig	ladno e/studi	studeno e	šudro i[si]

276 As Guentchéva herself notes, Skorniakova 2008 documents the occasional use of the ostensible neuter singular pronoun *ono* as an expletive subject with 'weather' verbs (and other impersonals) in both colloquial and literary Russian. Skorniakova also notes a corresponding use of the cognate *vono* in Ukrainian and *wono* in Lower Sorbian, this latter possibly under German influence.

also converge, but differ in these expressions from Slavic and most of Balkan Romance, while Aromanian straddles the dividing line between east and west, at least in idioms of temperature. The languages are clearly differentiated, especially in the occurrence of impersonal expletive subjects and the lexical items used for precipitation, but they are also linked in terms of the semantics of innovations in describing precipitation.

7.8.2.2.2 Experiencer Constructions

The expression of internal experience, including emotions and feelings such as regret, shame, being cold, and the like, involves an impersonal construction in many languages of the Balkans, and as with the weather verbs of §7.8.2.2.1, there is diversity in form but also some parallels suggestive of contact-induced convergence. The parameters of diversity are whether the expression is primarily noun-centered or verb-centered, what verb is used, how the experiencer is encoded, i.e., as a subject or an object, and, if an object, what case marking the experiencer receives.

By way of illustrating the situation with experiencers and the variation these constructions show across the languages, the Balkan forms for 'I am sorry,' 'I am ashamed,' and 'I am cold,' as representative examples of the class of constructions, are given in (7.216), (7.217), and (7.218), respectively, followed by some observations on the import of the structures evident here and the groupings that emerge from the data.

(7.216) 'sorry'

Grk	λυπάμαι 'I'm sorry' (lit., 'regret.1SG', nonactive verb form)
Alb	më vjen keq 'me.DAT comes.3SG.PRS bad'
Rmi	pharo si mange 'heavy is me.DAT'
Rmn	mi-este milă 'me.DAT-is pity'[277]
Aro	njilã nj-easti 'pity me.DAT-is'
Mac	žal mi e 'sorry me.DAT is'
Blg	žalno mi e 'sorry me.DAT is'

(7.217) 'shame'

Grk	ντρέπομαι 'I-am-ashamed' (1SG, nonactive verb form)
Rmi	ladžava 'I-am-ashamed' (Skt *lajje* 1SG, mediopassive verb form)[278]
Alb	më vjen turp 'me.DAT comes.3SG.PRS shame'
Rmn	mi-e ruşine 'me.DAT-is shame'
Mac	mi e sram 'me.DAT is shame'
Blg	sram mi e 'shame me.ACC is'

(7.218) 'cold'

Grk	κρυώνω 'I.cold' (1SG, active verb form)
Alb	kam ftohtë 'I.have (1SG) cold'

277 The Balkan Romance use of *milă/njilã* (with *m* > *nj* /__*i* via regular sound change) is an early borrowing from Slavic, and reflects the OCS meaning 'pity, compassion,' not the modern South Slavic meaning 'dear.'

278 Romani (Skopje, Arli Topaanli, etc.) also has *ladž i ma* (lit., 'shame is me.ACC.WK').

Rmi	šudro i mange / pahol man 'cold is me.DAT / colds me.ACC'
Rmn	mi-e frig 'me.DAT-is cold'
Aro	nj-easti-arcoari / nji-u-arcoari 'me.DAT-is-cold' (both)
Mac	ladno mi e / mi se studi 'cold me.DAT is / me.DAT INTR cold.3SG.PRS
Blg	studeno mi e 'cold me.DAT is'

In each of these data sets, Greek and Albanian stand out from the other languages, though in different ways. Greek shows verb-based expressions in each case, with the experiencer encoded as the subject of the verb, marked via the person/number endings (active or nonactive, depending on the verb itself); an overt subject nominal could of course occur, e.g., εγώ λυπάμαι, but since Greek is a pro-drop language (see §7.8.2.1), generally no such nominal is found. For 'shame,' Romani also has a fully verbal expression; in this case, though, as with all the Greek examples, it is a matter of a retention of a verbal construction from earlier stages of the languages: New Testament Greek had the verb σπλαγχνίζομαι for 'I feel sorry,' and Sanskrit has the verb *lajje* for 'I feel ashamed.' As noted in footnote 278, Romani can also have an accusative construction. Albanian typologically seems to be transitional in that it has a blend of a verbal and a nominal construction: the noun in (7.216) through (7.218) carries the primary semantic weight – *keq* '(something) bad,' *turp* 'shame,' and *ftohtë* 'cold,' respectively – but the verbal part is not a typical "light" verb; rather *vjen* 'come' and *kam* 'have' occur, each of which carries some semantic weight of its own.[279] The encoding of the experiencer in Albanian 'cold' is via the subject of the verb 'have,' but in the other two examples, the experiencer is a dative nominal (here the 1SG.DAT weak object pronoun *më*). This constructional variation between a verbal and a nominal construction, representing a different type of transitional state, occurs also in Romani for 'cold.' The other languages have a noun-based construction throughout, and encode the experiencer through an oblique case-marked nominal; Albanian, Balkan Romance, and Balkan Slavic all have a dative experiencer, while Bulgarian has a small number of predicates that take accusative-marked experiencers, a retention of a pattern in earlier Slavic (e.g., with OCS *sramъ*) that was extended somewhat, applying even with some loanwords, such as *enja* 'care,' a southeastern dialectalism from Greek.[280]

Thus, Greek is the outlier here in terms of these experiencer constructions, whereas Albanian and Romani show some parallelism with what is found elsewhere in the Balkans. The greatest convergence, however, is between Balkan Slavic and Balkan Romance, in regard to oblique marking of experiencers and the use of a noun-based construction. In this respect, Romanian and Aromanian pattern with Slavic, and in the case of Romanian, especially with Bulgarian (cf. Gołąb 1976); geographically and structurally, then, a contact-induced effect is

279 By contrast, a typical "light" verb would be 'be' or 'make,' as found in various paraphrases cross-linguistically.

280 Macedonian *sram te bilo!* 'shame on you' (lit., 'shame you.ACC be.N.LF') is a copy from BCMS, which has the accusative with the archaic optative use of the LF *bilo* (cf. §4.3.4.1.1). Romani also has an accusative construction with a DMS optative, e.g., *ladž te ovel olen* 'shame on them' (lit., 'shame DMS become ~ be.3SG.PRS them.ACC').

probably to be identified. Importantly, too, Balkan Romance diverges here from its Latin source. The corresponding Latin constructions involved verbs, whether expressed personally, as in (7.219a), or impersonally, as in (7.219bc):

(7.219) a. misereor 'I feel compassion for' (1SG, (medio)passive verb form)
b. me miseret 'I feel sorry for' (me.ACC be.sorry.3SG, active verb form)
c. me pudet 'it fills me with shame' (me.ACC shames.3SG, active verb form)

Thus, the experiencer constructions show some convergence that is likely to be contact-related, specifically involving Balkan Romance assimilating to Balkan Slavic nominal constructions, but overall the patterns of relationships that emerge within the Balkans for experiencer impersonals are different from those seen with the weather verbs; Albanian and Greek, for instance, do not match up here in the way they do with certain weather expressions.

7.8.2.2.3 Impersonal Constructions with Corresponding Personal Forms

The third type of impersonal subsumes those that, unlike the others discussed here, show a systematic correspondence, or alternation, with a personal construction with the same verb. This phenomenon is not unlike verb lability (see §6.2.6.4) except that instead of pairings of intransitive and transitive uses for the same verb form (e.g., Mac *šeta* 's/he walks, strolls,' *go šeta* 's/he walks it [the dog]'), the alternation here is between an impersonal expression with a source adjunct and personal counterpart with the source as subject. A paradigm case is the verb for 'drip,' e.g., impersonal '(it) drips from the faucet' and personal 'the faucet drips,' illustrated here with the verb italicized in each example:

(7.220)
	Alb	robineti *pikon* = *pikon* nga robineti 'the.faucet *drips* = *drips* from the.faucet'
	Mac	tapata *kape* = *kape* od tapata 'the.faucet *drips* = *drips* from the.faucet'
	Blg	trăbata *kape* = ot trăbata *kape* 'the.faucet *drips* = from the.faucet *drips*'
	Rmi	i češma *thavdela* = *thavdela* e češmastar / tar-i češma 'the faucet *drips* = *drips* the.OBL faucet.ABL / from-the faucet'
	Grk	η βρύση *στάζει* = *στάζει* από τη βρύση 'the faucet *drips* = *drips* from the.faucet'
	Rmn	ţeava *se scurge* = *se scurge* din ţeava 'the.faucet INTR *drips* = INTR *drips* from the.faucet'

In none of these is an overt expletive subject possible, so that, for instance, *αυτό στάζει από τη βρύση is impossible in Greek. Only Romanian marks the verb overtly as intransitive, through the use of the detransitiving/nonactivizing *se*. Despite the superficial structural similarities here, specifically the existence of the alternation itself and the fact that the same form of the verb is used in each function, there are no geographic distributional facts or striking or unusual convergent details that would warrant a contact explanation for the facts in (7.220). This conclusion is bolstered by the fact that the parallel verb in English, *drip*, appears both with a source subject (*The faucet was dripping*) and with an expletive subject

(*It was dripping from the faucet*), the English equivalent of the impersonals in (7.220), suggesting a naturalness to the Balkan alternation that would make a contact explanation unnecessary. Nonetheless, in an inventory of Balkan impersonals, the similarity in (7.220) is worth noting.

7.8.2.2.4 Impersonal Passives, Real and Potential

In nominative/accusative languages, passive constructions generally involve a change in valence of the verb and a change in the focus of the action expressed by the verb. For example, an active transitive verb, i.e., one which has an agentive subject and an affected object, passivizes with the object being treated as the subject of the verb and the agentive nominal being relegated to a different role, often as an instrumental or the like; cf. English *Robin found Batman* (active) / *Batman was found by Robin* (passive). It is possible also to have an unspecified agent in the active form (e.g., *Someone found Batman*), which in the passive construction can simply be omitted (e.g., *Batman was found*). In impersonal passive constructions, there is an unspecified agent, so that the focus is more on the action than on the participants, but unlike the *Batman was found*-type, with its clear logical object (*Batman*), there can be differences in the nature of the logical object, understood or overt or otherwise, as the discussion below indicates.

7.8.2.2.4.1 Impersonal Passives of Verbs of Speaking/Communicating

One type commonly referred to as an impersonal passive is found with verbs of speaking and communicating more generally, such as 'say' or 'write.'[281] These can occur in a passive form as a way of introducing the complement clause in a more focused manner, and thus with a 3SG form of the verb, reflecting the morphosyntactic properties of the complement clause; compare English *It is said* or *It is written* followed by a complement, where the *it* is an expletive required by English syntax, e.g., *It is said/written that the Balkans offer exciting opportunities to study language contact*. For the most part, such constructions are like ordinary passives built from a clause with an unspecified subject – compare *They say/write that the Balkans offer ...,* where the subject has no specific referent – and an object that is the complement clause to the verb of communicating (*that ...*). Still, constructions of this type possibly show some interesting effects as far as language contact is concerned, and the unspecified nature of the subject warrants considering them to be impersonals, as in Guentchéva 2010, and their passive form warrants placing them here.

As far as the Balkans are concerned, all of the languages of interest here show passives with verbs of communicating, formed in the usual way for each language, thus synthetically via passive voice morphology on the verb in Greek and Albanian

281 Hale & Buck 1966: §287a, for instance, say that "The name Impersonal is also conveniently applied to verbs that have an Infinitive or a Clause for subject," as would be the case with verbs of communicating.

and periphrastically via the use of a reflexive/detransitivizing element (labeled INTR in (7.221)), *se* in Romanian, Bulgarian, and Macedonian, and *pe* in Romani,[282] with the active verb form;[283] relevant examples are given in (7.221), along with a complementizer that would introduce the clausal complement:

(7.221)	Grk	λέγεται ότι …	('is-said.3SG.NACT that … ')
	Alb	thuhet se …	('is-said.3SG.NACT that … ')
	Rmn	se spune că …	('INTR say.3SG.ACT that … ')
	Blg	kazva se če …	('say.3SG.ACT INTR that … ')
	Mac	se kažuva deka	('INTR say.3SG.ACT that … ')
	Rmi	vakerela pe kaj …	('say.3 SG .ACT INTR that … ')

In Greek, the indefinite-subject third-person plural active form λένε 'they say' can be used, much as the English *they say* noted above; this use is found in Latin, e.g., *legunt* 'they say,' so that the Romanian use of its nonactive form here seems to represent a favoring, possibly contact-related given what is found in neighboring languages, of one alternate means of expression over another. Similarly, this indefinite active third-person plural usage is possible in Bulgarian, as Guentchéva 2010: 41 notes, and in Macedonian, though perhaps not as usual as the nonactive usage in (7.221), and actually occurs in Slavic more widely, for instance in Russian (*govorjat* 'they say'); whether there was such a pattern earlier in Slavic that modern Balkan Slavic has moved away from, however, is not clear.

Interestingly, however, in OCS, in such a construction, a third-person form could be used that was singular (not plural) and active (not passive); thus one finds *pišetъ* for 'it is written / they write' in the indefinite/impersonal sense under discussion here. This form occurs four times in Matthew (4:4, 4:6, 4:7, 4:10; cf. Diels 1963: II.5), and each time, importantly, it corresponds to a Greek passive verb, and a perfect passive at that (γέγραπται 'it has been written,' i.e., 'it is in a state of having been written,' thus 'it is written'). This possibility has been given up in modern Balkan Slavic, and the passive construction in (7.221), which was a possibility in OCS as well, has been selectively adopted and favored. The basis for the parallel favoring of a nonactive expression in Balkan Romance and Balkan Slavic is not clear, but it represents another way in which these languages show convergence in the domain of impersonal expressions. Similarly, the parallel Romani construction shows a movement away from earlier syntax, in that Sanskrit used here a quotative particle (*iti* 'thus'), with or without an overt verb of communicating; Romani thus has moved in the direction of Balkan Romance and/or Balkan Slavic, although the exact source for the calquing is difficult to pin down precisely.

282 As noted in §6.2.6.1, the INTR (= REFL) *se* / *pe(s)* morpheme of, respectively, Balkan Romance and most dialects of Romani uses a weak object pronoun agreeing with the subject in the first two persons, whereas the corresponding element in Balkan Slavic (and in some Romani dialects influenced by Slavic) is invariant, always in the *se* form in Slavic and *pe(s)* in Romani.

283 In the simple past tense (aorist) in Albanian, a periphrastic formation is used, with an element *u* that corresponds etymologically (in ways too complicated to allow for full presentation here, though see §6.2.6.1, footnote 379) to the *se* in Romance and Slavic languages.

7.8.2.2.4.2 *Impersonal Passives for Generalized Activity*

Impersonal passives with unspecified agents can also be formed in some of the languages from active intransitive verbs, i.e., verbs without an object at all. These passive forms thus focus entirely on the action expressed in the verb, inasmuch as there is no specified subject and no logical object, and represent an activity as going on in a general way. A typical translation for such a passive with a verb like 'dance' would be 'Dancing is taking place,' 'People dance here,' or the like. This impersonal passive use is found in most of the Balkan languages. A representative sampling is given in (7.222):

(7.222) Alb andej luftohej 'over.there fought.NACT.IMPF.3SG'
Mac tamu se boreše 'there INTR fought.ACT.IMPF.3SG'
Rmn atunci se lupta 'there INTR fought.ACT.IMPF.3SG'
Rmi odothe marela pe sine 'there fight.PRS.ACT .3SG INTR REM'
Blg tam se boreše 'there INTR fought.ACT.IMPF.3SG'
Grk πολεμούσαν 'fought.ACT.IMPF.3PL' (thus, 'they were fighting')
'There was fighting over there; fighting was taking place there;
they were fighting over there'

As is evident in (7.222), while Albanian uses an overt nonactive (also referred to as mediopassive) verb and Balkan Slavic, Balkan Romance, and Romani use their equivalents to a passive, consisting of an intransitive marker with an active verb, Greek is exceptional in that it uses a third-person plural active form with an understood indefinite subject, much like the alternant in §7.8.2.2.4.1 concerning verbs of speaking. The convergence that the other languages show is striking, and the Greek exceptionality parallels what is found with some of the other impersonals. The historical interpretation of all this is tricky, but some important observations can be made. In the case of Balkan Slavic, this formation may show a divergence from earlier states, to judge from the absence of such expressions in OCS. However, OCS is a limited corpus, and Večerka 1996: 241 cites a relevant example with the reflexive construction from a manuscript outside the canonical OCS corpus but of similar antiquity:

(7.223) pridetъ sę
comes.PRS.3SG INTR
'someone comes'

Further, this generalized activity impersonal is found elsewhere in Slavic, in Polish for instance (Kibort 2001; Gawelko 2005), as in (7.224):

(7.224) Tutaj było tańczone
here was.N.SG dance.PTCP.N.SG
'There was dancing here / Dancing took place here'

Thus that it is safe to assume that there were antecedents within early Slavic for the Bulgarian and Macedonian usage in (7.222), and the occurrence of such impersonal

passives in Baltic (e.g., in Lithuanian) and elsewhere in Indo-European would suggest as much.[284] Moreover, Latin also had this usage, as shown in (7.225):

(7.225) *itur* 'someone goes' ('go.PASS.3SG; lit., 'it-is-gone (by someone)')
pugnatum est acriter 'there was a fierce fight' (lit., 'fought.PST.PASS.PTCP.N.SG is.PRS.3SG fiercely,' i.e., 'was.fought fiercely (by someone)') (Caesar *Gallic War* 3.21.1)

This Latin usage is presumably the basis for the Balkan Romance usage, updated to the current morphosyntactic analogue to the Latin passive. Therefore, the striking convergence seen in (7.222) probably reflects a superficial parallel involving independent lines of descent into the modern languages rather than a contact-related convergence; such a conclusion is of course complicated by the lack of relevant historical evidence for Albanian. In the case of Macedonian, intransitive impersonal marking extends even into the quintessentially intransitive verb 'be,' e.g., *teško e da se bide glup, ima golema konkurencija* 'it is difficult to be stupid, there's a lot of competition' (lit., 'difficult is DMS INTR be.3SG.PRS.PFV it.has large competition').

7.8.2.2.4.3 Impersonal Passives Expressing Potential

The impersonal passive that expresses generalized activity can also be used with a particular functional nuance, to indicate potentiality for that activity to occur; an example is given in (7.226) for 'It is possible to sleep well here':

(7.226)	Alb	Këtu flihet mirë	'here sleep.NACT.PRS.3SG well'
	Mac	Ovde se spie dobro	'here INTR sleep.ACT.PRS.3SG well'
	Blg	Tuk se spi dobre	'here INTR sleep.ACT.PRS.3SG well'
	Rmi	Akathe sovela pe šukar	'here sleep.ACT.PRS.3SG INTR well'
	Rmn	Aici se doarme bine	'here INTR sleep.PRS.3SG well'
	Grk	Μπορεί να κοιμηθεί κανείς καλά εδώ	'can.3SG sleep.PRS.3SG someone well here'

By extension, such impersonal passives, when negated, are used to indicate impossibility, essentially an interdiction; an example is given in (7.227), where the expression is, more or less literally, 'there is no passing going on here,' thus 'it is impossible to go here,' i.e., 'Do not go here':

(7.227)	Alb	Këtej s'kalohet	'here NEG pass.NACT.3SG'
	Blg	Ottuk ne se minava	'from-here NEG INTR pass.ACT.3SG'
	Mac	Odovde ne se minuva	'from-here NEG INTR pass.ACT.3SG'
	Rmn	De aici nu se trece	'from here NEG INTR pass.ACT.3SG'

Once again (see §7.8.2.2.4.2), there is convergence involving most of the languages, with Greek as the outlier, using an overt modal verb, here μπορεί rather than a nonactive form.[285] The pattern of convergence is thus the same as with other impersonal passives, especially the generalized activity type. Here, however, it may be that the earlier stages

284 This construction is also found, e.g., in Sanskrit and Old Norse.

285 The nonactive verb form κοιμηθεί for 'sleep' in the Greek of (7.226) is not significant here, since this verb is deponent and has no active forms.

of at least Balkan Slavic and Balkan Romance did not have this nuance associated with the generalized activity impersonal construction, again judging from the evidence of OCS and Latin, so that the possibility that the convergence in (7.226) is contact-induced must be seriously entertained. Thus the convergence is clear, even if the exact causation and the directionality of external influence, if any, are not.

7.8.2.2.5 Impersonal Modals

As the potential impersonal passive of §7.8.2.2.4.3 indicates, impersonal constructions can be used in the expression of various types of modality. Going beyond the potentiality of the impersonal passives, and looking further at the semantics of other modal expressions, one can distinguish impersonal constructions which – somewhat akin semantically to the internal experience of feelings and emotion (see §7.8.2.2.2) – highlight nuances of speaker disposition and intent, as well as impersonals that express the modality of possibility and necessity. Both show interesting patterns of convergence in the Balkans.

7.8.2.2.5.1 Internal Disposition

Internal disposition, positive or negative, on the part of a speaker can be expressed in most, and to some extent all, Balkan languages through the use of an impersonal passive construction with a dative (or equivalent) personal pronoun indexing the experiencer. These forms translate as 'I feel like … ' or 'I don't feel like …, ' though literally they are 'to.me VERBs/is-VERBed … '; the examples in (7.228) all mean 'I feel like eating (X)' and are literally 'to-me eats/is-eaten (X)':

(7.228)			
	Mac	Mi se jade (burek)	'me.DAT INTR eats.PRS.3SG (burek)'
	Blg	Jade mi se (bjurek)	'me.DAT INTR eats.PRS.3SG (burek)'
	Alb	Më hahet (një pica)	'me.DAT eats.MDP.PRS.3SG (a pizza)'
	Aro	Nji-si mãcã	'me.DAT-3INTR eat.PRS.3SG'
	Megl	Ăń-ți máncă	'me.DAT-3INTR eat.PRS.3SG'
	Rmi	Hala pe mange (mas)	'eat.PRS.3SG 3INTR me.DAT (meat)'[286]

The parallelism among these languages is striking, both as to the form and as to the functional nuance associated with this form.

Not all of the languages show this construction; in Romanian, in particular, one would say simply *vreau să mănânc*, literally 'I.want DMS I.eat.' In both Romanian and Aromanian, there is also the straightforward expression of hunger via a literal 'to-me is hunger,' as in:

286 This particular example is based on a judgment given by an informant (VAF field notes); in a naturally occurring example, the 'feel like' construction was not chosen, even though when the speaker codeswitched into Macedonian he did use it:

i. Šukar i, na mangav te hav (Rmi) ne znam zošto, ne mi se jade (Mac)
good is not I.want DMS I.eat not I.know why, not me.DAT ITR eat.PRS.3SG
'It is good; I don't want to eat – I don't know why, I don't feel like eating.'

(7.229) Rmn Mi-e foame 'me.DAT-is hunger'
Aro Nji u-foami 'me.DAT is-hunger'

which appears to be more a Romance turn of phrase than a Balkan one, as suggested by French *j'ai faim* and Spanish *tengo hambre*, both literally 'I.have hunger'; the use of the verb 'be' with a dative pronoun in these Balkan forms rather than 'have' would seem to continue the Latin dative of possession construction and so it both parallels the French and Spanish use of 'have' and represents an archaism in Balkan Romance rather than an innovation.

Greek in general is an outlier here, as with the impersonal passives more broadly, expressing the "inclination" sense of (7.228) via a phrase that literally uses 'have' plus a noun for 'mood, disposition':

(7.230) Έχω διάθεση να φάω (πίτσα) / Έχω διάθεση για πίτσα
have.1SG mood DMS eat.1SG (pizza) for
'I feel like eating (pizza)' / 'I'm in the mood for pizza'

There is, however, an interesting dialectal point of contact involving Greek. In particular, Papadamou & Papanastassiou 2013 (see also Papadamou 2018, 2019a) report that in the Greek of the northern prefecture of Kastoria, there are villages that have the following constructions:

(7.231) a. Mi trójiti
me.ACC eat.NACT.3SG
'I feel like eating' (lit., '(to.)me it.is.eaten')
b. Mi píniti
me.ACC drink.NACT.3SG
'I feel like drinking' (lit., '(to.)me it.is.drunk')

which mirror the Balkan Slavic and Albanian type in (7.228) exactly, with nonactive voice forms and with an indirect object weak pronoun.[287] Given the geographical limitation of this type within Greek to just those northern varieties in contact with Macedonian and/or Albanian, it can be safely assumed that this represents a contact-induced innovation in Greek.[288] Moreover, since the construction is found throughout Slavic, and is absent from Balkan Romance north of the Danube as well as from Romani dialects not in contact with Slavic, it is safe to conclude that the usage is of Slavic origin in the other Balkan languages, although, as usual, the lack of ancient evidence for Albanian leaves a degree of uncertainty.

287 The verbs in (7.231) show the vowel raising characteristic of a northern dialect, so that they are from an earlier [trójete] / [pínete], the forms found in southern dialects (τρώγεται / πίνεται). Similarly, the experiencer object is accusative here, reflecting the northern use of the accusative for indirect objects, parallel to the genitive in southern dialects of Greek and datives in other languages (see §6.1.1.1.3.1).

288 Nikolaos Lavidas (National and Kapodistrian University of Athens) and Ianthi Maria Tsimpli (University of Cambridge) have informed us (p.c.) that in their corpus of West Thracian Greek, this construction does not occur.

7.8.2.2.5.2 Modality of Possibility and Necessity

The modality of possibility and necessity reflects speakers' interest in worlds as they could be and worlds as they should be, that is to say, possible world modality and deontic modality. To a certain extent, both of these reflect reference to unreal conditions, to irrealis modality. As far as Balkan impersonals are concerned, an impersonal form of the verb plays a prominent role in the expression of the Balkan periphrastic future tense and from a semantic standpoint, reference to futurity, as a kind of possible and unreal, or as yet unrealized, state, can be subsumed under the rubric of irrealis modality. However, the future tense also involves interaction with personal verb forms, and in any case, the Balkan future is intimately tied to morphosyntax and to the morphosyntactic categories of tense and modality. It is therefore discussed in Chapter 6 (§6.2.4.1), although a few relevant observations are made here. This section then concentrates on nonfuturate irrealis modality and its realization in impersonal constructions.

Possibility and necessity are expressed similarly in the various languages, with impersonal verb in a third-person singular form followed by a complement with a personal verb headed by the DMS. In that way these impersonals are parallel to auxiliaries in other languages in terms of their function but behave like regular complement-taking main verbs. Examples are given for possibility in (7.232) and for necessity in (7.233); in (7.232) the sentences mean 'Perhaps/It-might-be-that I will eat (burek)' and literally are '(it.)can that I.eat (burek)'), while in (7.233) the meaning is 'I must eat' and the forms literally are '(it.)must that I-eat':

(7.232)
Mac	Može da jadam burek
Blg	Može da jam bjurek
Alb	Mund të ha një byrek
Grk	Μπορει να φάω ένα μπουρέκι
Rmi	Šaj te hav bureko
Aro	Poati s-mãc
Megl	Pọti să mănânc
Rmn	Poate să mănânc

(7.233)
Mac	Treba da jadam burek
Blg	Trjabva da jam bjurek
Alb	Duhet të ha një byrek
Grk	Πρέπει να φάω ένα μπουρέκι
Rmi	Valjani/Trebul te hav bureko
Rmn	Trebuie să mănânc
Aro	Lipseashti s-mãc

Despite the similarities evident in (7.232) and (7.233), there are some differences to be noted. For instance, in some of the languages for some of these verbs, conjugated personal forms are possible while others have only the 3SG forms. Albanian, for instance, has a set of personal verb forms for 'can,' as do Greek and Romanian: the Albanian forms are all nonactive (e.g., *mundem* (1SG), *mundesh* (2SG), etc.) while the Greek and the Romanian forms are active (Grk μπορώ (1SG), μπορείς (2SG), etc.; Rmn *pot* (1SG), *poti* (2SG), etc.). Albanian also has personal forms for

'must,' e.g., *duhen* 'they-must,' with a different construction, namely with the participle as the complement, e.g., *duhen konsideruar* 'they must be considered.'-[289] Regarding complementation, it can be noted too, as discussed in §7.7.2, that 'can' is one of the last verbs in Romanian that still allows an infinitival complement, e.g., *eu pot merge* 'I can go' (alongside *eu pot să merg*). In Balkan Slavic, conjugated verb forms have different pragmatics from the impersonal ones.

An important diachronic development with these particular impersonals is that in some instances they have developed, or can be assumed to have developed, out of personal constructions. For instance, inflected 'can' (root *mog*-) is found in all Slavic languages, and impersonals with this root are, outside of South Slavic, more likely to involve a predicative adverb *možno* (< **mogĭno*-), as in Russian, than a verb, so that the impersonal *može* of South Slavic is likely to be the innovation. Cf. also the use of *puede* in Macedonian Judezmo in (7.164) . In Albanian, inasmuch as *duhe*- is formally the mediopassive of *dua* 'want, love,' it can be assumed that the 3SG *duhet* 'it is necessary, it is needed' is a specialization of a personal *duhet* 'she/he/it is wanted.'[290] Similarly, in Ancient Greek, the root πρεπ- could occur in a personal construction meaning 'be visible; be like' as well as, more rarely, 'be fitting,' while the 3SG, impersonal, form meant 'it is fitting' (with dative of the affected nominal plus and infinitive complement), so that most scholars assume that the impersonal usage developed from the personal. The same development occurred also with the modal verb μέλλω 'be about to,' which in its impersonal form, 3SG μέλλει, came to serve as an alternative future expression in early Postclassical Greek, a usage that continued into the medieval period.[291] This coincides with the fact that the volitionally based futures in the Balkans, as detailed in §6.2.4.1, all show a depersonalization in their development, with a conjugated finite verb form of 'want' as a predecessor to the emergence of invariant marker;

289 Geg has personal *duhem* + INF, e.g., *duhem me shkue* 'I must go.' The Albanian participial construction has a parallel in Romanian, e.g.:

i. Trebuie spus numai adevarul (Pănă Dindelegan 2013: 104)
must.3SG say.PPP only truth.NOM.DEF
'Only the truth must be said' (lit., ' … must said')

The relationship between the Albanian and the Romanian constructions is not entirely clear, but the absence of such a pattern in French and Italian is suggestive of a substrate origin for them (as with the old lexical parallels noted in §4.2.1.1). Alternatively, given that a similar construction occurs dialectally in at least American English, e.g., *The car needs washed*, possibly based on a Celtic model, it could reflect an Indo-European inheritance or even parallel but independent developments in each language. To this can be added the Macedonian use of *saka* 'want' or *bara* 'seek, want' plus a verbal noun in the meaning 'needs to be,' e.g., *sirenjeto saka jadenje* 'the.cheese needs eating/to.be.eaten,' *Na kolata i treba mienje* 'the car needs washing/to.be.washed' (lit., 'to the.car needs washing'), *zgradata bara renoviranje* 'the building needs renovating/to.be.renovated.' Cf. also Aro *Va multu imnari pana Bituli* (= Mac *Saka mnogu odenje do Bitola*) 'It takes a long time (lit., 'it.wants much') to get (lit., 'going') to Bitola' (Markovikj 2007: 168), with the verbal noun (*imnari*).

290 There is a general connection between 'want' and 'need' (where 'want' is primary and 'need' is secondary), in many languages, including all the Balkan ones, e.g., Grk το στιφάδο θέλει λίγο αλάτι, Mac *čomlekot saka malce sol* 'the meat.stew wants/needs a.little salt.' See Roussou 2005 on the Greek construction and note the discussion of Eng *wanting* in §4.3.10.

291 See Markopoulos 2009 and Lucas 2013 for a full account of future μέλλω, and cf. §6.2.4.1.1 and §6.2.4.1.4.

thus earlier Greek θέλω να γράφω 'I-will DMS I-write' is the starting point for the later future marker θα, and a similar starting point can be assumed in all of the Balkan 'want'-based futures, with a depersonalization of the 'want' verb to a 3SG form (e.g., Grk θέλει να γράφω) as the first step on the way to an invariant (reduced) marker, e.g., Lat *vol-* in the case of Rmn *o*, Aro *vã*, Megl [TsR] *ã[s]*, early Slv **хъtje-* for Mac *kje*/Blg *šte*, Alb *dua* for later 3SG *do*, and Rmi *kam-* for *ka[m]*, with *mang-* 'want' > *me* in some dialects (e.g., Drindari, a Balkan II dialect in southern Bulgaria; Boretzky & Igla 2004: 2, Map 138). In this last case, *kam-* is etymologically 'love' but is lost in many dialects except in its use as a future; *mang-* originally meant 'beg' and underwent a semantic shift to 'want.' Torlak BCMS gives evidence that the 1SG is the last to be replaced, as some dialects preserve person marking precisely and only in the 1SG (1SG *ču* vs. elsewhere *če*).

The source of this depersonalization is not clear, and it has been speculated, at least regarding the future (Joseph 1978/1990; Joseph & Pappas 2002), that it is due to a drive to reduce redundancy. While there may be some validity to such a notion for at least certain types of language change, other morphosyntactic developments in the Balkans, in particular the replacement of the infinitive by finite forms, add redundancy via personal marking on subordinate clause verbs. To the extent that increased redundancy with the loss of the infinitive might be due to the communicative advantage it offers in a multilingual contact situation, as suggested in §7.7.2 (see also Joseph 1983a: chapter 7), the reduction of redundancy with the depersonalization of modals might indicate that contact does not play a major role, if any, in this development. Indeed, the chronology of the depersonalization of μέλλω in Greek would point in the same direction.[292] Nonetheless, whatever the basis for their emergence from personal forms, depersonalized modals contribute to the superficial parallelism in Balkan syntax, even if contact is not responsible, and the depersonalizing specifically in the future shows that futurity in the Balkans behaves like a mood.

7.8.2.2.6 'Have' Existential versus 'Be' Possession

The final type of impersonal to be discussed here has to do with the expression of existence and possession. It is well known that these two notions show relationships in terms of their formal expression cross-linguistically, with what might seem like the

292 Moreover, depersonalization of an auxiliary seems to have occurred independently in two outlying Greek dialects. Spiro 2008 presents examples of impersonalized 'have' in the perfect system of the Greek of southern Albania:

i.
ixe	ipoférune	ta	xoría
had.3SG	suffer.3PL	the-	villages

'The villages had suffered' (lit., 'it.had they.suffer the villages') and Sitaridou 2014a gives instances from Pontic Romeyka of inflected infinitives in the pluperfect with an invariant 3SG form of 'have' as the perfect auxiliary:

ii.
ixe	ipina
had.3SG	say.INF+1SG

'If I had said … '

quintessential verb of possession, 'have,' often figuring in existential constructions, and the quintessential verb of existence, 'be,' in possessive constructions. Both constructions are found in the Balkans, but in some instances, the occurring patterns reflect convergent inheritances rather than sprachbund-related contact.

The basic facts concerning existential constructions with 'have' in the Balkans are that Greek, Albanian, Balkan Slavic, and Balkan Romance all signal existence, and by the same token nonexistence, via an impersonal 3SG form of the verb 'have,' negated if appropriate, with the nominal for which existence is predicated expressed in the accusative case, for those languages that distinguish nominal case; some examples are given in (7.234):

(7.234)
	Mac	Ima ženi tuka 'There are women here' (lit., 'has women.PL here')
	Blg	Njama ženi tuk 'There aren't any women here' (lit., 'not.has women.PL here')
	Alb	Ka dikë këtu 'There is someone here' (lit., 'has someone.ACC here')
	Grk	Έχει αχινούς στη θάλασσα 'There are sea-urchins in the sea' (lit., 'has urchins.ACC in.the sea')
	Aro	Aoatsi-ari multsã mileti 'There are a lot of people here' (lit., 'here-has much people')
	Megl	Ari lucru 'there is work' (lit., 'has work')

The Balkan Romance situation is somewhat more complicated. Caragiu-Marioțeanu 1958 states that 'have'-existentials occur in widespread use in both Aromanian and Meglenoromanian and notes that they occur in Romanian, but only in two contexts, with a negative plus indefinite object as in (7.235a), with either an infinitival or a DMS complement, and in the fixed phrase in (7.235b):

(7.235) a. N-are cine aici să ajute
NEG-has.3SG whom.ACC here DMS helps.3SG
'There is no one here to help'

b. N-are nimic 'Never mind!' (lit., 'not.has nothing')

The fixed expression in (7.235b) matches Albanian *s'ka gjë* 'it's nothing; it's all right' (lit., 'not.has thing'), though the provenance of this parallel is unclear as to whether it involves calquing in some way or is simply an independent development in each language.

The rest of the history here, however, is fairly clear. Greek has had an existential construction with 'have' like that in (7.234) since Roman times in the Postclassical period, but not in Classical Greek; given the chronology of the emergence of this construction in Greek, it is most likely a calque on a Late Latin model. The Latin construction, with the form *habet* '(it-)has' plus accusative and at least for Western Romance the adverb *ibi* 'there' as well, is quite likely the source, via calquing, of the Albanian construction, and is definitely the ancestor of the Balkan Romance construction.[293] However, the restricted nature of the 'have' existential in Romanian needs to be explained. Influence from early Slavic is a possibility, since the Balkan

293 The presence of *ibi* is posited to account for the *y* in the Spanish existential *hay* and the French existential *il y a*.

Slavic formation of (7.234) with 'have' represents an innovation away from the early Slavic type with 'be.' The 'be'-type is seen in OCS (present tense *jes-*, future *bǫd-*, past *bě-*; cf. Lunt 2001: 164) and is still found elsewhere in Slavic, e.g., Russ *jest'* 'there is/are,' even though the 'have'-type occurs to some extent outside of South Slavic.[294] A reasonable scenario, therefore, is that the Aromanian and Meglenoromanian unrestricted use of 'have' in existentials reflects an inheritance from Late Latin *habet*, as elsewhere in Romance, while the Romanian restriction and spread of 'be' in existentials reflects Slavic influence, either learnèd influence from Church Slavonic or influence at a more colloquial level before Balkan Slavic innovated the use of 'have.' The South Slavic innovative generalization of a 'have'-existential (see footnote 294) is perhaps best seen as the result of the SDBR substratum (cf. Gołąb 1976). The spread into BCMS (but not Slovene), where 'have' in existentials occurs in an invariant 3SG form, *ima*, can be compared to the spread of the 'will' future, which affects all of South Slavic except Slovene and the Kajkavian dialects of Croatian. Both contact with old populations of Balkan Romance speakers and the influence of Greek in the Orthodox Church may also have been factors.

A further point of convergence with the 'have'-existential, but also an area for some divergence among the languages, is in the possibility of the use of an (ostensibly definite) object pronoun as the entity for which existence is being predicated. Thus Macedonian (and, *mutatis mutandis*, Bulgarian, and the BCMS complex), Albanian, and Aromanian allow this structure, while Greek does not, as the examples in (7.236), for 'they exist, there are … ' (lit., 'them (it-)has'), show:

(7.236) Mac *gi ima*
Alb *i ka*
Aro *l-are*
Grk *τα έχει (OK: Ø έχει)

This is potentially of typological interest since the predicated entity in existentials is typically indefinite when full noun phrases are involved, as in (7.234), and excludes definite pronouns (cf. English *There is a man* / *There is him). However, there are non-Balkan languages that allow such pronouns; in Spanish, for instance, *Lo hay*, lit., 'him it.has,' is possible for 'There is one (e.g., a book).' Therefore, the facts of (7.236) may reflect independent developments.

Thus there is certainly convergence in form in the Balkan languages regarding the use of an impersonal 'have' in existential expressions and contact is relevant for some aspects of the convergence. However, it is not fully a sprachbund-type of contact-induced convergence – Greek borrowed the construction early and Albanian may have too; Balkan Romance inherited its 'have' construction but Romanian innovated away from that due to contact;

294 Both Polish and Ukrainian show a limited use of impersonal 'have' in existentials, specifically in negative constructions in the present tense (*nie ma* / *nemaje*, respectively), as discussed in Twardzisz 2012. Such facts suggest that the 'have' existential may have begun in late Common Slavic, since it is found in all branches of Slavic. Still, the complete generalization of the 'have'-type in South Slavic can be taken as significant, differentiating it from East and West Slavic.

Balkan Slavic innovated the construction, possibly due to contact with Greek. So contact is involved, and a subset of the convergence may qualify as Balkanologically significant, but the overall convergence is probably not a Balkanism in the sense developed here.[295]

As for possessives, the verb 'be' is found with a dative of possessor in earlier stages of all Balkan languages for which ancient or medieval attestations exist, i.e., Classical Greek, Latin, Old Church Slavonic, and Sanskrit.[296] This most likely is a Proto-Indo-European construction that was inherited in each of those languages. In all except Sanskrit, a verb 'have' developed in ancient times and occurs alongside the 'be' possessive.[297] The 'be' construction is not technically an impersonal since the possessum is the subject of 'be,' but it is relevant here since it gives way in all but Romani (among the Indo-European Balkan languages) to the use of 'have,' so that the later instantiations of these languages in the Balkans use a verb 'have' for possession; this construction therefore provides an interesting counterpart to the existential 'have.' The Romani construction, as in Sanskrit, is not a true impersonal with the verb 'be,' the possessor is underlyingly or overtly in the accusative (versus Sanskrit genitive), and the possessum is nominative as seen in (7.237), where (7.237b) has a nominative as a dislocated topic along with the accusative pronoun and nominative possessum:[298]

(7.237) a. (Man) si man duj čhave
(me.ACC) is me.ACC two children/sons
'I have two children' (lit., '(to-)me is two children')

b. Jekh daj si la duj čhave
one mother is her.ACC two children/sons
'One mother, she has two children' (lit., 'one-mother is (to-)her two children')

Relevant here too is the Romani use of 'be' with accusative in the future construction (see also §6.2.4.1.5), as it shows impersonal syntax, with the Romani counterpart, via calquing, to the negative future with 'have' in Balkan Slavic:[299]

(7.238) Nae man te hav
NEG.is me.ACC DMS eat.1SG
'I will not eat'

7.8.2.2.7 Conclusions Regarding Balkan Impersonals

Several observations can be made that sum up what the foregoing sections show about impersonals in the Balkans. First, the languages on the territory of the Ottoman vilayets with a majority of Macedonian or Albanian speakers pattern

295 Note that both Banfi 1985 and Sh. Demiraj 2004 consider the 'have' existential to be a Balkanism.
296 Actually, in Sanskrit it is a genitive possessor with 'be,' but the genitive subsumes many dative functions in Classical (and later) Sanskrit, so this is equivalent, in a sense, to a dative possessor. Turkic also has existential possession using special particles.
297 See McAnallen 2011 on possession in OCS, with relevant literature, where she notes the marginal occurrence too of a construction with the locative preposition *u* with the genitive for the possessor.
298 Some Romani dialects in the Balkans have a lexical verb for 'have' based on the root *ther-* 'hold.'
299 Some Romani dialects in the Balkans can also form a positive future with the possessive construction, e.g., *si man te džav* 'I have to/will go' (cf. Boretzky & Igla 2004: 2, Map 138).

together most closely, suggesting a relevance to contact as an explanation for convergences there. Second, there is a general tendency to move from the presence to the absence of person marking in constructions with modal senses, which may reflect universal tendencies possibly encouraged by contact. Third, for those languages with ancient attestations and with both Balkan and non-Balkan modern forms, noncanonical/impersonal/nonperson-marked subjects are generally more common in the Balkan varieties than in the non-Balkan varieties. This is especially evident in Romance and Slavic. Thus there are pockets of contact-induced change in impersonals, as well as independent inheritances that give parallelism in some of the languages, so that overall the situation with the wide range of Balkan impersonals shows that both inheritance and contact must be taken into account in understanding convergent structures in the Balkans. The spread of the dative to older accusative impersonal constructions in Macedonian (§7.8.2.2.2) is probably one of contact-induced convergence, and the dative impersonal of internal disposition in Kastoria Greek (§7.8.2.2.5.1) is clearly another.

7.8.2.2.8 Narrative Imperatives

As noted in §7.8.2.2, there is a use of the imperative in the Balkans that falls under the rubric of impersonals, in an extended sense, the so-called narrative imperative (Friedman 2012d, on which much of this section is based). In this construction, an imperative – an ostensibly second-person form but here without a specified second-person subject – occurs with other persons as subject and is interpreted as such. The imperative here renders the narration of past actions particularly vivid in a shorter passage. Unlike the historical present, which can be maintained over a longer stretch of narrative, the narrative imperative is rarely more than a sentence in a longer narrative.

The following example from Macedonian (Hacking 1997b: 215), is illustrative, with the imperatives and the corresponding tensed and personal interpretations marked in bold and the narrative imperative subjects in italics:

(7.239) Nie sedevme vo kolata vo blizina na avtobuskata stanica. Gi čekavme David i Aneta da dojdat so avtobus. Čekavme, čekavme – avtobusot go nemaše i što pravevme? *Jas* **izlezi**, **vlezi**, a *Elena* **gledaj, gledaj**, no niv gi nemaše.
'We were sitting in the car near the bus stop, waiting for David and Aneta to arrive on the bus. We waited and waited but there was no bus. What were we doing? *I* **got in** (and) **got out**, *Elena* **looked** (and) **looked**, but there was no sign of them.'

In her discussion of the phenomenon in Macedonian, Hacking 1997b: 218, following Jakobson 1957 on Russian, argues for a single invariant meaning for both exhortative and narrative imperatives, namely "imposition," the difference between the two being the locus, i.e., the speaker's desire in the former and some external imposition in the latter. She is explicitly interested not in the origin of the narrative imperative (Hacking 1997b: 218) but rather in its synchronic position in modern Macedonian, where it is structurally identical to the exhortative imperative.

The narrative imperative occurs not only in Slavic, but also in Albanian, Balkan Romance, Romani, and Turkish, but is not found generally in Greek. From the point of view of language contact, five facts to be illustrated in §7.8.2.2.8.1–9 below stand out as particularly significant:

(i) The phenomenon is found throughout Slavic, but the restrictions on its occurrence and usage point to specifically Balkan developments for Balkan Slavic.
(ii) The usage is rare in Greek but also in the Aromanian dialects of Greece (cf. Friedman 2008b on similar distribution with respect to object reduplication).
(iii) The Romanian usage is characteristic specifically of those dialects that were longest in the Ottoman Empire – Wallachia and Moldavia – and at the same time had more influence from East Slavic as well as Balkan Slavic.
(iv) Today the usage is best preserved in Macedonian and Albanian and the languages in contact with them.
(v) The distribution and vitality of the narrative imperative in the Balkans suggest that it is an areal rather than a typological phenomenon.

The occurrence of the narrative imperative in the various languages is outlined in the sections that follow; a brief discussion of a nonfinite construction that appears to be functionally related closes this section.

Before embarking on this survey, however, a terminological clarification is needed. Asenova uses the term *narrative imperative* to refer to a construction that involves the DMS plus present tense verb form to express an unexpected (and undesired) past action, as in the following examples: (7.240) and (7.241) are Albanian, given with their Bulgarian translations (all from Asenova 2002: 193), which show the same phenomenon. The DMS constructions and their translations are boldfaced. Example (7.242) is from Macedonian (from Palmer 1997) with the same effect. The English translations here attempt to convey the stylistic effects of the passages while staying as close to the original as possible:

(7.240) **Të kthehet** në mes të natës i pirë,
DMS return.3SG.PRS.MDP in middle PC.GEN night.GEN PC.M.NOM drunk
të ngrejë më këmbë shtëpinë dhe **të na**
DMS raise.3SG.SBJV to foot house.DEF and DMS us.DAT
bëjë turp për botën! (Alb)
make.3SG.SBJV shame for world.DEF.ACC
Da se vărne posred nošt pijan, **da vdigne**
DMS ITR return.3SG.PRS in.the.middle night drunk.M DMS raise.3.SG.PRS
kăštata na krak i **da ni napravi na smjah** na horata! (Blg)
house.DEF to foot and DMS us.DAT make.3SG.PRS to laugh to people
'[Just imagine!] **He returns** in the middle of the night drunk, **he wakes up** the whole house, and **he makes a laughingstock of us** in front of everyone!'

(7.241) **Të rrish** një orë në rradhe dhe **të mos marrësh** bukë! (Alb)
DMS sit.3SG.SBJV one hour on line and DMS MNEG take.2SG.SBJV bread
Da sediš edin čas na opaska i **da ne vzemeš** hljab! (Blg)
DMS sit.3SG.PRS one hour on tail and DMS NEG take.2SG.PRS bread
'[Just imagine!] **You stand** in line for a whole hour and **you don't get** any bread!'

(7.242) **Da mi ti vlezi** nekoe kuče vo gradinata,
DMS me.DAT you.DAT enter.3SG some dog into garden.DEF
sè **da mi ti uništi**. (Mac)
all DMS me.DAT you.DAT ruin.3SG
'[Can you believe it?!] Some dog **got into** my garden and **wrecked** everything'

Asenova 2002: 192–193 adduces similar examples from Romanian and Greek, and writes that they are typically colloquial, dialectal, or folkloric and that in examples taken from (belletristic) literature they indicate the speaker's involvement with the action, rendering the narration vivid and dramatic. On the basis of the occurrence of the construction in Albanian, Bulgarian, Greek, and Romanian (and Macedonian as well, albeit not noted by Asenova), she describes the phenomenon as *obštobalkansko*, i.e., a Balkanism, an eminently reasonable conclusion to draw here.

Still, the designation *narrative imperative* for such examples seems inappropriate because an imperative is not involved, even if the DMS with a second person verb form, as in (7.241), can in principle substitute for a true imperative (see §6.2.4.3.2).

Asenova also cites Nicolova 1974, who connects the DMS usage in Bulgarian with the narrative use of the (synthetic) imperative, the construction that we are calling here the narrative imperative. Further, Asenova claims that such a use of the imperative proper does not occur in the other Balkan languages and is only dialectal in Bulgarian. However, the survey in the following sections makes it clear that Asenova is mistaken.

7.8.2.2.8.1 Macedonian

In Macedonian, a perfective narrative imperative can be used if the action is completed and iterative as in example (7.243), from Hacking 1997b: 215, and (7.244), slightly modified from Koneski 1967: 418; as in earlier examples, the imperative and its interpretation are in bold and a subject, if overt, is in italics – note that, owing to pro-drop in Macedonian in general (see §7.8.2.1), these imperatives do not require an overt subject:[300]

(7.243) Tetin Nomče beše orač. Eden vol **kupi**,
Uncle N. was.IPFV.IMPF ploughman one ox buy.PFV.IMPV
drug **pcovisaj**.
other die.IPFV.IMPV

300 Thus, *kupi* in (7.243) has an understood subject filled in by reference to *tetin Nomče*; note too the absence of overt subjects in (7.245) and elsewhere.

Cel život toa raboteše.
whole life that.N work.IPFV.IMPF
'Uncle Nomče was a ploughman. He'd **buy** one ox and *another* would **die**. His whole life went like that.'

(7.244) Se vrakjavme pijani: *toj* **padni**, *jas* **stani**, *jas* **padni**, *toj* **stani**
INTR return.IMPF.1PL. drunk.PL he fall.IMPV I stand.IMPV I fall.IMPV he stand.IMPV
'We were returning drunk: *he* **fell**, *I* **stood** up, *I* **fell**, *he* **stood** up'

Imperfective imperatives can also be used in Macedonian for repeated or habitual actions that are atelic or gnomic and hence imperfective as in examples (7.245) and (7.246), from Koneski 1967: 418:

(7.245) **Teraj**, **teraj**, **napinaj**, **napinaj** – ne
pull.IPFV.IMPV pull.IPFV.IMPV struggle.IPFV.IMPV struggle.IPFV.IMPV not
biduva, ne biduva – ostana kolata v kal
be.possible.PRS.3SG not be.possible.PRS.3SG remain.AOR.3SG wagon.DEF in mud
'He **pulled** and **pulled**, **struggled** and **struggled**, it was no use, the wagon remained in the mud'

(7.246) Starite **naveduj** **se**, mladite **rasti**,
old.PL.DEF bend.over.IPFV.IMPV INTR young.PL.DEF grow.IPFV.IMPV
taka vrvi vekov
thus go world.DEF.PX
'The old **get bent over**, the young **grow up**, that's the way of this world'

Such examples make it clear that the usage, while colloquial, is quite normal in modern urban Macedonian speech. This in turn is consistent with choices made by Macedonian intellectuals in their valorization of many colloquial features. The Macedonian usage, however, may also have been strengthened by language contact.

7.8.2.2.8.2 Bulgarian

In Bulgarian, Teodorov-Balan 1940 gives examples of the narrative imperative, but none of the later (post-World War Two) normative grammars include such usage, and Nicolova's 1974 Bulgarian examples are all from dialect studies. Educated speakers of modern Bulgarian find the usage stylized, dialectal, and archaic, and some have observed that even in nineteenth- and early twentieth-century Bulgarian literature, the narrative imperative was used to evoke peasant speech.[301] It would thus appear that this usage was disfavored by the eastern Bulgarian intellectuals whose dialect became the basis of the Bulgarian standard, despite the fact that the construction occurs in eastern Bulgarian dialects.

7.8.2.2.8.3 Bosnian-Croatian-Montenegrin-Serbian

Stevanović 1986: 708–709 cites numerous examples from nineteenth-century authors of the BCMS canon from all over the former Serbo-Croatian-speaking regions of former Yugoslavia, e.g., Njegoš (southern Montenegro) and Kovačić (from Lukovdol, near the Slovenian border), but he specifies the usage as being especially

301 These assessments are based on interviews by VAF with Bulgarian speakers (2010).

characteristic of Montenegro. Maretić 1963: 625, writing on what was then the western variant of Serbo-Croatian, on the other hand, marginalizes the usage, noting that neither Vuk's nor Daničić's (nineteenth-century Serbian) Bible translations have a single narrative imperative, and that they are rare in Vuk's writings. He goes on to observe that the usage is "very common in the speech of the southern regions, but it is a dialectism" (Maretić 1963: 626).[302] As in Macedonian, examples involve both perfective and imperfective verbs. From this it is clear that already before the break-up of the former Serbo-Croatian, narrative imperatives were identified as being more "Serbian" than "Croatian," and the usage was viewed as typical of those dialects closest to, or part of, Balkan Slavic.

7.8.2.2.8.4 East and West Slavic and Questions of Origin

In East Slavic the perfective imperative is used to denote sudden past actions (Vinogradov 1972: 434–437). The East Slavic usage thus differs from the South Slavic in two important respects: it is semelfactive rather than iterative, and it is limited to perfectives. In West Slavic, the construction occurs in Czech only in fixed expressions, and in Slovak it is considered an East Slavonicism except in such expressions. It is almost nonexistent in Polish (Nicolova 1974). Vinogradov notes the suggestion that the Russian usage derives from a reinterpretation of the Common Slavic 2/3SG aorist, which in some conjugations looks like a singular imperative, but the South Slavic usage cannot be so interpreted since the aorist is still alive there, and, what is more, it is precisely in the southern part of South Slavic territory, where the aorist is best preserved as a functional (and morphological) category, that the narrative imperative – which overlaps with only a limited, unproductive class of aorists – is also most alive. Koneski 1967: 420 suggests that the Macedonian usage actually derives from the 2/3SG present perfective, which is sometimes indistinguishable from the aorist and which no longer stands alone (but this last condition applies only to the western Macedonian dialects, which in this respect reflect the standard). He notes that present perfectives can be so used in BCMS. However, given that BCMS also has the narrative imperative, and that the usage also occurs with imperfectives, this argument appears unlikely. Unfortunately, probably because the construction is highly colloquial and does not occur in Greek, it is not attested in OCS.[303] While the difference between the aspectual force of the East Slavic and South Slavic usages could be invoked to support a claim of separate developments, it could also be argued that this was a Common Slavic colloquial innovation that developed differently in East and South Slavic. The fact that it is most robust precisely in Macedonia and Montenegro is revisited in §7.8.2.2.8.10 below.

302 In the original: " … *je veoma običan o govoru južnih krajeva, ali to je dialektizam*."

303 Večerka 1989, 1993, 1996, 2002, 2003 show no examples. Pandev 2000: 62–63, however, does have examples from the *Tikveš zbornik*, a late medieval Macedonian Church Slavonic manuscript that shows colloquial influences.

7.8.2.2.8.5 Albanian

Examples of the narrative imperative in Albanian are cited as emotive usage in Buchholz & Fiedler 1987: 150:

(7.247) Po vajta, se s' kisha ç' të bëja, vajta
so I.went.AOR that not have.IMPF.1SG what DMS do.IMPF.1SG I.went.AOR
këmbadoras, aty **ngreu**,[304] aty **rrëxohu**
foot.hand.ADV there get.up.IMPV.INTR there fall.down.IMPV.INTR
'So I went on, for there was nothing else for me to do, I went on all fours, I **stood up**, I **was knocked down.**'

The imperatives in (7.247) literally represent 'stand up!,' 'be knocked down!'

7.8.2.2.8.6 Turkish

Turkish has the same type of usage of the imperative, as can be seen in examples (7.248) and (7.249), which are translations of (7.243) and (7.244) above. Here the (a) examples are in standard Turkish and the (b) examples are the West Rumelian Turkish (Gostivar) dialect of North Macedonia:

(7.248) a. Nomçe enişte çifçi-ydi: bir öküz **al**, diğeri **geber**.
N. uncle ploughman-was.PCOP one ox buy.IMPV other die.IMPV
b. Nomçe inişte idi çifçi: **al** bir ukuz, übürisi **geber**
N. uncle was.PCOP ploughman buy.IMPV one ox other die.IMPV
'Uncle Nomče was a ploughman. He'd **buy** one ox and *another* would **die**'

(7.249) a. Sarhoş dön-üyor-du-k: *o* **düş**, *ben* **kalk**, *ben* **düş**, *o* **kalk**.
drunk return-GPRS-PST-1PL he fall.IMPV I stand.IMPV I fall.IMPV he stand.IMPV
b. Sarvoş din-er-dı-k: *o* **düş**, *ben* **kolk**, *ben* **düş**,
drunk return- GPRS-PST-1PL he fall.IMPV I stand.IMPV I fall.IMPV
o **kolk**
he stand.IMPV
'We were returning drunk: *he* **fell**, *I* **stood up**, *I* **fell**, *he* **stood up**'

In the case of example (7.249), Standard Turkish would prefer a gerundive construction of the type *düşe kalka* 'falling, arising,' but the construction with the imperative is also permissible.

7.8.2.2.8.7 Romani

The narrative imperative occurs in all of the three main dialect groups spoken in North Macedonia. Example (7.250), which translates (7.244), is in the Skopje Arli dialect, but speakers of the other dialects judge the construction as normal as well.[305]

(7.250) Irinaja sine amen mate: *ov* **per**, *me* **ušti**, *me* **per**, *ov* **ušti**
we.return REM we drunk: he fall.IMPV I stand.IMPV I fall.IMPV he stand.IMPV
'We were returning drunk: *he* **fell**, *I* **stood up**, *I* **fell**, *he* **stood up**'

304 The form *ngreu* would be *ngrihu* in the post-1972 standard. 305 VAF field notes from 2009.

Matras 1994b does not note such usage for the Northern Vlax dialects he investigated, and while such an argument *ex silentio* cannot be considered definitive, nonetheless, given his discourse-oriented approach, its absence is striking. Note also that Cech & Heinschink 1999: 125 cite examples from Turkey, Kosovo, and Serbia, but not Greece. The usage does occur in the Romani of Ágios Athanásios (formerly Ali Bey Köy), a village that has become a suburb of Sérres in Greek Macedonia. As noted in Friedman 2012f, this dialect shows very old contact with Slavic, and while the town of Sérres itself apparently had a large Greek population even before the Balkan Wars, the village of Ali Bey Köy was a little more than half "Bulgarian," i.e., Macedonian (Kănčov 1900: 177), the remainder being entirely Romani.

7.8.2.2.8.8 Romanian

According to modern speakers, the narrative imperative is characteristic of Wallachia and Moldavia and has a somewhat archaic or dialectal feel to it now. Example (7.251), from Graur et al. 1966: 223, is from the nineteenth-century Wallachian writer B. Delavrancea.

(7.251) Cartea e deschise la foia 80; şi *eu* **trage**-i tare şi deluşit …
book.DEF is opened to page 80 and I read.IMPV-it aloud and clearly
'The book opened to page 80; and *I* **read** it aloud and clearly … '

The examples in Irimia 1976: 126 are likewise nineteenth century and from Wallachia and Moldavia. The example in Romalo 2008: 383, which as a work makes an effort to use recent data, is from the Wallachian writer Ion Caragiale's story *Kir Ianulea* published in 1909. Pană Dindelegan 2013 does not mention the usage at all, but Vasilescu 2013: 398 does provide an example in a completely different context, discussing the colloquial use of expletive pronouns:

(7.252) **Dă**-i cu bere, **dă**-i cu vin!
give.IMPV-3SG.DAT with beer give.IMPV-3SG.DAT with wine
'They keep drinking glasses of beer and wine one after the other!'

7.8.2.2.8.9 Aromanian, Meglenoromanian, and Greek

Like Macedonian, the Aromanian dialects of North Macedonia have narrative imperatives, as in example (7.253) from Frasheriote Ohrid Aromanian:

(7.253) Shi *mini* **fudz** shi *năs* **fudz**
and *I* **run**.IMPV and *he* **run**.IMPV
'I ran and he ran'

In the Aromanian dialects of Greece, such usages are unusual but they do occur, as seen in (7.253') and (7.253") from Bara et al. 2005: 207, where (7.253') was used to elicit (7.253"):

(7.253') Κι αυτοί τι κάνουν, μια **έλα** εδώ, μια **φύγε** εκεί
and they.NOM what do.3PL one **come**.IMPV.SG here one **leave**.IMPV.SG there
'And they, what do they do? Sometimes they came [lit., 'come!'] here, sometimes they went off [lit., 'leave!'] there!'

(7.253") … unã ela aua, unã du-ti aklo
one **come**.IMPV.SG here one **go**.IMPV.SG-you.ACC.INTR there
‘ … one **came** here, one **went** there’

Meglenoromanian also has a narrative imperative as in examples (7.254), based on (7.244), and also (7.255). These examples describe a general state of affairs in the past:

(7.254) Nă turnăm beac, **kaz-** **skoala-** ti
we return.IMPF.1PL drunk, fall.IMPV rise.IMPV youACC.INTR
‘We were coming home drunk, we kept **falling down** and **getting up**’

(7.255) Toata ziua **mănkă**, **mănkă**, nu si dumănkă
all day eat.IMPV eat.IMPV not INTR get.full.3SG.PRS
‘All day he **ate** and **ate** and didn’t get full’

7.8.2.2.8.10 Summary Regarding Narrative Imperatives

We can thus conclude that, *pace* Asenova 2002: 193 as noted above, the true narrative imperative occurs not only in Slavic, but also in Albanian, Romanian, Aromanian, and Meglenoromanian as well as Romani and Turkish, and, to a limited extent, also Greek. Moreover, two basic facts about the distribution of the Balkan narrative imperative are striking: (a) its use is more frequent as one moves south for both Balkan Slavic and Balkan Romance, but it is unusual in Greek and dialects of other languages in contact with Greek (cf. Friedman 2008c on similar phenomena with respect to object reduplication), and, with the exception of Slavic, it does not occur in related languages/dialects outside the Balkans;[306] and (b) it is most vital in languages/regions in contact with Slavic (Albanian, Balkan Romani, Wallachia, Moldavia), with the west being favored over the east. Moreover, the East Slavic narrative imperative is quite distinct from the South Slavic, but it is not unreasonable to speculate that both have their origins in a colloquial Common Slavic usage that developed differently in the two regions, and in its Balkan manifestation developed most robustly precisely in the regions of the most complex contact. The use of a 2SG imperative – which cross-linguistically is often a bare (or nearly bare) root – to render a narrative more vivid makes sense typologically. However, given the distribution in the Balkans and absence beyond the Balkans, the narrative imperative, which is better preserved in Macedonian than in Bulgarian, appears to be a Balkanism, i.e., an areal rather than a typological phenomenon.

306 Data for Arvanitika or Arbëresh would help in dating the phenomenon. It would be worthwhile to look at Pontic, Cappadocian, and Greek Cypriot as well. This highly localized distribution within Greek for the narrative imperative, apparently found just in a region in which Aromanian is also spoken, is reminiscent of the geographically restricted occurrence within Greek of the impersonal of internal disposition described in §7.8.2.2.5.1, and is further testimony to the importance of geography in Balkan language contact.

7.8.2.2.8.11 A Functionally Related Usage: Narrative Verbal Nouns

As discussed in the preceding subsections, the narrative imperative is associated with a certain vividness of narration. Moreover, the imperative in this construction is not overtly marked for the person of its associated subject. There is another construction found in the Balkan languages involving a form with no marking for a notional subject that adds vividness to a narration. The relevant form in this functionally related construction is a deverbal nominal that, as a nominal, has no person marking for subject. Illustrative examples from various of the Balkan languages can be found in parallel passages in translations of Aleko Konstantinov's 1895 novel *Bai Ganyo*, originally written in Bulgarian. Bai Ganyo is chasing a train that he thinks is leaving the station but is in fact simply changing tracks, and he explains to the narrator that he set off after the train, " … and I was running and running …":

(7.256) a. pa bjag pa tičane … (Blg; Konstantinov 1895)
and.then running.VBLN[Ø] and.then running.VBLN

b. pa trčanje, pa trčanje … (Mac; Konstantinov 1967)
and.then running.VBLN and.then running.VBLN

c. koşmak ama, ne koşmak … (Trk; Konstantinov 1972)
running.INF but what running.INF

d. και φευγιό και τρεχάλα (Grk; Konstantinov 1922)
and fleeing.VBLN and running.VBLN
' … and I ran and I ran … ' *or* 'and I was running and running'

While the occurrence across several languages of verbal nouns here may partly be a reflection of the translation process, since the Bulgarian original has verbal nouns, the fact of their availability for use in such vivid narration is what is of particular interest here. It is clear, for instance, that these deverbal nouns have a functional similarity to narrative imperatives, given the expressiveness they show. Moreover, this similarity is demonstrated by the fact that in other languages, imperatives occur in the parallel translations; Romanian has only (narrative) imperatives (*fugi*, repeated) whereas Albanian and Serbian have an interesting combination of a narrative imperative (*nxito* and *trči*, respectively) and verbal noun (*vrap* and *trk*, respectively, both bare root verbal nouns):

(7.257) a. şi fugi, şi fugi! (Rmn; Konstantinov 1964)
and run.IMPV and run.IMPV

b. e nxito e vrap (Alb; Konstantinov 1975)
and run.IMPV and running.VBLN

c. pa trk, pa trči (Srb; Konstantinov 1955)
and running.VBLN and run.IMPV
'I ran and I ran'.

7.8.3 Interrogation-Related Sentences

In various places in Chapter 4, where lexical Balkanisms are the focus, instances of the borrowing of material related to questions, and, in some cases, answers as well,

are adduced, specifically in §4.2.3.5, §4.2.3.1.1, and §4.2.4.2.2. Similarly, a convergence in yes-no question intonation, as revealed by Lehiste & Ivić 1980, is discussed in §5.5.4. Since there are apparent contact-related lexical and prosodic effects in the realm of interrogation, it is not surprising that there should be some syntactic effects as well where contact might be or definitely is involved, although determining conclusively that contact is responsible for some convergences in this domain can be difficult. Nonetheless, as with some of the other parallels or convergences discussed in this chapter, the superficial similarity is striking and thus worthy of presentation here, even if some of it is inconclusive with regard to contact.

7.8.3.1 Yes-No Questions

Several of the Balkan languages have a special word that is used in the formation of a yes-no question. In particular, Albanian (*a*), Balkan Slavic (*li)*, and Turkish (StTrk *mI,* WRT *mi*; Friedman 1982c) have native formations, Aromanian in North Macedonia has borrowed the Slavic one, Romani borrows freely from all three languages, and some WRT dialects borrow Albanian *a* (Matras & Tufan 2007: 219) depending on the particular dialect. The examples in (7.258) are illustrative:[307]

(7.258) a. A keni cigare? (Alb)
Q have.2PL cigarette

b. Dali imate cigara? (Mac)
Q have.2PL cigarette

c. Imate li cigara?[308] (Blg)
have.2PL Q cigarette

d. Sigaran var mı? / Sende sigara var mı? (Trk)
cigarette.2PL.POSS exist Q you.LOC cigarette exist Q
'Do you have a cigarette?'

e. A gittin Stambola? (WRT)
Q go.2SG.PFV.PST Istanbul.DAT
'Did you go to Istanbul?'

f. Gideys mi sinemaya (WRT)
go.1PL.OPT Q movies.DAT
'Shall we go to the movies?'

g. Aoa li eshtsã (Aro)
here Q you.are.2SG.PRS
'Are you here?'

307 Romani, Balkan Slavic, and Aromanian can omit Q particles in questions, but in Albanian and Turkish, where the Q particles are quite different, they are the norm.

308 This sentence could also be used in Macedonian, but as Englund 1977 demonstrated, the use of *li* in these contexts is vastly more typical of Bulgarian than of Macedonian. By contrast, Blg *dalí* (vs. Mac *dáli*) is not used to introduce ordinary questions but has different functions of the type 'I wonder if,' which is a type of yes-no question, but not necessarily a request. Thus (7.258bc) offer yet another salient difference between Macedonian and Bulgarian.

The specific forms (leaving aside the obvious copying in WRT, Romani, and Aromanian) are different, but if the function and syntactic slot are similar enough, some contact effects could be posited. Albanian is clearly an outlier here, and Greek appears to be irrelevant except for its influence on West Thracian Turkish, i.e., the ERT dialect of Greece as reported by Petrou 2018: 301–302, 2021: 276–277, where *mI* can be omitted in questions and interrogation is indicated simply by intonation, as in Greek. As Petrou notes, the same use of intonation and absence of *mI* occurs in Cypriot Turkish (probably also under Greek influence) and in the Turkish of eastern Anatolia (in contact with Iranian), but we can say that those parallels actually strengthen the case for a contact-induced explanation, given the larger Turkic picture described in the next paragraph.

Convergences in the usage of Slavic *li* and Turkish *mI* (especially the Romani instantiations of both) are covered in detail in §6.2.5.3 (cf. also §6.1.2.5). However, the Turkish and Slavic particles show important differences, and the Albanian is even more different. While Turkish *mI* is obligatory in yes/no questions, Albanian *a* and Slavic *li* are not. Second, *mI* and *li* are enclitic, occurring after the element they serve to focalize in the question. Moreover, Albanian *a* is sentence-initial, while the Turkish and Balkan Slavic Q words have different syntactic requirements. Most importantly, *mI* and *li* are found outside of the Balkans, respectively, in most other Turkic and all other Slavic languages, respectively, e.g., OT *mU,* Uzbek *-mi,* Kazakh / Karakalpak *MA*, Bashkir / Tatar *-mĔ,* etc, and all of Slavic for *li*. Moreover, within Balkan Slavic, there are differences in the details of placement and use of *li* across all of Slavic and even between Macedonian and Bulgarian within Balkan Slavic; see Rudin et al. 1999 for a careful and thorough study. Given all the differences in these various elements used in forming yes-no questions, there is no reason to think of the occurrence of such a marker here as anything other than a mere parallelism.

One other interrogative construction discussed earlier (in §§7.4.1.2.2.2.2 and 7.6.3.2) deserves further treatment here. In Albanian, Greek, and Romani, the respective modal negators *mos,* μη(ν), and *ma* go beyond the various functions discussed above in §§7.6, 7.6.2–7.6.4, and 7.8.1.2, and are found as markers of highly tentative yes-no questions. The modal negator in these questions, exemplified in (7.259), signals greater doubt on the part of the questioner so that with the use of *mos/*μη(ν)/*ma*, it is not just a simple polar yes-or-no answer that is being sought but rather a genuine absence of any surety as to the situation is being expressed (the Greek now sounds old-fashioned; see Thumb 1912: §255 for other examples):

(7.259) a. (A) mos e njihni atë? (Alb)
Q MNEG him.ACC.WK know.2PL him.ACC.STR
'Do you perhaps know him?'

b. Μην είδες το παιδί; (Grk)
MNEG saw.2SG the child
'Did you perhaps see the child?'

c. Ma sijan nasvalo? (Rmi, Topaanli Arli)
MNEG are.2SG sick.M
'Are you perhaps sick?'

What makes the use of this particular element in just this way of some interest in a Balkanological context is that precise parallels seem not to be found elsewhere in the Indo-European languages that have a reflex of the modal negator *meH_1; rather, although it is admittedly difficult to measure precisely, the particularly high degree of tentativity seen with the m-negator in (7.259) seems to be restricted to Albanian, Greek, and Romani, and to be lacking elsewhere. It is true that Classical Armenian *mi* is used in questions expecting a negative reply (Klein 2011), and that Sanskrit *mā* is used in questions that seem to express some hesitation, as in (7.260), from Monier-Williams 1899: s.v.:[309]

(7.260) mā bhūd āgataḥ (Amaru, *Amaruśataka*)
MNEGP may.be.3SG arrived.PTCP
'Can he not (i.e., surely he must) have arrived?'

Nonetheless the questions in (7.259) seem not to signal a specific answer – positive or negative – in the speaker's mind. Interestingly, in Ancient Greek, the degree of tentativity with μή in questions seems quite high; it is described in especially strong terms in Emde Boas et al. 2019: §56.8, viz. "[μή] indicate[s] that the speaker is reluctant to accept a positive answer as true: 'is it really the case that …, ' 'it isn't the case that …, is it?', 'surely not … ?'"; an example is given in (7.261), from Emde Boas et al. 2019: §38.8, and see also (7.86) in §7.6.3.2:

(7.261) Ἀλλὰ μὴ ἀρχιτέκτων βούλει γενέσθαι; (Xen. *Mem*. 4.2.10)
but MNEG architect want.2SG become.INF
'But you don't want to become an architect, do you? / But surely you don't want to become an architect?'

These facts suggest – refining somewhat the discussion in §7.6.3.2 – that expressing some tentativity could well have been an inherited use for reflexes of *meH_1, but a heightened degree of tentativity, the particular point of convergence among the modern languages seen in (7.259), may have deeper roots in the Balkans. As noted in §7.6.3.2, Greek and Albanian are not generally believed to form a subgroup within Indo-European (though note recent arguments to the contrary in Hyllested & Joseph 2022), so it would seem that this situation might not easily reflect a shared innovation that the two branches undertook together, at some prehistoric point where they might have been a single speech community. Instead, in such a view, contact is more likely to be involved. Still, the directionality and age of the contact are not readily determinable. There are some ancient loans from Doric Greek into Albanian, e.g., *mokër* 'millstone' (from μᾱχανᾱ́ 'device'), so that an assumption that Albanian took this usage from an innovative use in Greek at an early period would be reasonable, but in principle Greek could have gotten it from an innovating prehistoric Albanian; moreover, it could even be a more recent entry into Albanian from contact with Postclassical, even Medieval, Greek. Whatever the exact nature of the contact scenario, however, the convergence

309 It should be noted that the verb form used here, the so-called injunctive mood, imparts some of the modality, and thus the tentativity seen in this example.

seen here is of interest, whether for Balkanology of ancient times or of the more recent period. The similar use of *ma* in some dialects of Romani could have arisen at the time of contact with Byzantine Greek.

There are, moreover, other Balkanologically significant angles involving these tentative questions surrounding possible contact effects. In the Romani of Agía Varvára (AV), outside of Athens, as described by Igla 1996, the expression of a tentative question uses *mipos*, an Ancient Greek composite of μή with πως 'in any way' that in Modern Greek is a variant of μη(ν) in some of its uses, and can replace it acceptably in (7.259b); Igla 1996: 162 gives the following AV Romani example:

(7.262) Mípos páli phiradán tut lása
Q again went.2SG you.ACC her.INST
'You haven't been going around with her again, have you?'

Igla makes the observation also that in AV Romani, *mipos* is always sentence-initial; since this is the usual position for this form as a question word in Greek, this means that the lexeme has been borrowed along with its syntactic characteristics, a phenomenon seen with other lexical borrowings, e.g., with Turkish postpositions borrowed into various Balkan languages as postpositions, as discussed in §4.3.3.2.

In this regard, it is worth examining such dubitative questions in a broader Balkan context. The equivalent of the interrogative use of the modal negator in other Balkan languages is BSl *da ne*, BRo *să/si/s' nu*, as well as Aro *naca/nacã,* Rmi *te na* and *ma te* (see §§7.3.6.2, 7.6.2, 7.7.2.1.3.2.2) and Rmn *nu cumva* (lit., 'not perhaps').[310] For illustration, a translation of NTGrk μήτι, a variant of μή (a composite of μή + τι 'something'), into various Balkan languages, including Modern Greek, is given in (7.263), a passage from Matthew 12:23:

(7.263) a. Μήτι οὗτός ἐστιν ὁ υἱὸς Δαβίδ; (Grk, Rcvd text)
MNEG this is the son David?
b. Μήπως εἶναι οὗτος ὁ υἱὸς τοῦ Δαβίδ (ModGrk, Bible 1962a)
MNEG is this the son him.GEN David
c. A mos është ky biri i Davidit? (Alb, Bible 1980)
Q MNEG is this son.DEF PC David.DEF.GEN
d. Te na ovela bi kakava e Davidovesko Čavo (Blg/Rmi, Bible 1932)
DMS NEG is COND this DEF.OBL David.GEN son
e. Da ne bi tozi da e Davidovijat sin? (Blg, Bible 1995b)
DMS NEG be.3SG.OPT[311] this DMS is David.GEN son
f. S-nu hibã taha aestu Hilju-al David (Aro, Bible 2004)
DMS-NEG is.SBJV DUB this son-DAT David
g. Ne e li ovoj Hristos sinot Davidov (Mac, Bible 1990a)
NEG is Q this Christ son.DEF David.ADJ

310 The Turkism *zar (ne)* can also introduce dubitative yes-no questions in Balkan Slavic, and can also be used question-finally as a tag (cf. §4.3.4.2.2).

311 In synchronic terms, *bi* could only be a 3SG aorist or a 2/3SG conditional marker. As a conditional marker, however, it would require an *l*-participle main verb, and as an aorist it would not cooccur with the DMS. The form is thus best treated as a frozen optative, which is its historical origin.

h. Dali nanaj adavkha okova eja Davidesko čhavo (Blg/Rmi, Bible 1995a)
Q NEG this that DEF.OBL David.GEN son

i. Nu cumva este aceast Fiul lui David (Rmn, Bible 1962b)
NEG somehow is this son.DEF him.DAT David
'Is not this [perhaps] the Son of David?'[312]

As can be seen from (7.263c), the translation of the New Testament Greek into Albanian using MNEG is consistent, while in Romani, although the morpheme *ma* is available, it is not used.[313] It is also worth noting that although DMS + NEG occurs in all of Balkan Slavic, Balkan Romance, and Romani as an option – and it is the choice in Aromanian, Bulgarian, and one of the Romani translations – in Macedonian, Romanian, and a different one of the Romani (7.263h) translations, other means of expressing interrogation are used and the dubitative quality is erased.

Aromanian *naca/nacã* corresponds exactly to the above-mentioned interrogative use of the Albanian, Greek, Romani modal negator, Romanian *nu cumva* and Balkan Slavic *da ne ... [nešto]* 'is [not] (by any chance/perhaps [lit., 'something'] …)', i.e., the same usage as modal negator in the languages that have it. According to Papahagi 1974: s.v., Aro *naca/nacã* combines Grk να with Aro *cã*. This etymology, however, does not properly explain the second vowel in the form *naca*, which is the more frequent realization. Moreover, in some Aromanian dialects *naca* takes the subjunctive.

The examples in (7.264) provide further illustrations of this particular usage across a variety of Balkan languages, and see also (7.259), and in §7.6.3.2, (7.87):

(7.264) a. Μην είσαι τίποτα άρρωστος; (Grk)
MNEG are.2SG anything sick
'You're not feeling sick, are you?'

b. Μήπως είσαι άρρωστος;
NEG.Q are.2SG sick (Grk)
'Perhaps you are sick?' / 'You're not feeling sick, are you?'

c. Naca esht (tsiva) landzit? (SAro)
NEG.Q are.IND.2SG anything sick
'You're not feeling sick, are you?'

d. Naca hii (tsiva) niptut? (NAro)
NEG.Q are.SBJV.2SG anything sick
'You're not feeling sick, are you?'

e. Nu cumva esti bolnav (Rmn)
NEG perhaps are.2SG sick
'Perhaps you are sick?'/'You're not feeling sick, are you?'

f. Te na sijan nasvalo? (Rmi, Arli, Skopje)
DMS NEG are.2SG sick.M
'Are you perhaps sick?'

312 Note that the King James version, like the Macedonian translation (7.263g) and one of the Romani ones (7.263h), omits the dubitative quality, which in the other Romani one (7.263d) is marked by a conditional and in the Aromanian one (7.263f) is marked by the dubitative particle *taha,* borrowed from Greek.

313 These facts are true of all available translations, which, for the sake of space, are not cited here.

g. Ma te san nasvalo? (Rmi, Cergar [S.Vlax], Tirana)
MNEG DMS are.2SG sick.M
'Are you perhaps sick?'

h. Ma te isinan namborime? (Rmi Arli, Korça)
MNEG DMS are.2SG sick.M
'Are you perhaps sick?'

7.8.3.2 Multiple WH-Questions

Question words, specifically content-based question words, so-called "WH-question words," the equivalents of *who*, *what*, *why*, *where*, etc. in English, figure in a convergence that occurs in all the Balkan languages except Greek. Except for Greek, multiple instances of a WH-question word occur fronted at the left edge of a sentence or clause, as in (7.265).[314]

(7.265) a. Koj kogo vižda? (Blg; cf. Rudin 1988b)
who whom see.3SG
'Who sees whom?' (lit., 'who whom sees')

b. Koj kogo vide? (Mac; cf. Kochovska 2006)
who whom saw.3SG
'Who saw whom?' (lit., 'who whom saw')

c. Cine cu ce merge? (Rmn; cf. Rudin 1988b)
who with what goes.3SG
'Who goes by what (means of transportation)?'

d. Cari (pi) cari vidzu?[315] (Aro)
who DOM whom saw.3SG.AOR
'Who saw whom?'

e. Dihet kush çka ka bërë në vitet '90? (Alb)[316]
know.MP.3SG who what has done in years '90
'It is known who did what in the '90s'

f. Ko kas dikhlja? (Rmi; McDaniel 1989: 590)
who whom saw.AOR.3SG
'Who saw whom?'

g. Kim kiminle nerede ne yapıyor? (Trk)
who with.whom where what is.doing
'Who is doing what with whom where?'

314 Since Turkish is an SOV language, the left edge is not the "front" in the same sense as in the Indo-European Balkan languages. Although SOV order is less strict in WRT than in Standard Turkish, those differences are irrelevant here.

315 See §6.1.1.1.2 for discussion of *pi* in Aromanian. Thede Kahl (Austrian Academy of Sciences, University of Vienna) has the following to say about multiple WH-words in Aromanian (p.c.): "For the dialects without *pi*, both *Cari vidzu cari* and *Cari cari vidzu* are possible. *Cari tsi vidzu* 'who saw what' gives the feeling that something is missing at the beginning of the sentence, e.g., *Nu mata shtiu* 'They do not remember' (*cari tsi vidzu*). At the beginning of the sentence *Cari vidzu tsi* would be preferred, in which case *tsi* is emphatic; however, *Cari tsi vidzu* is not wrong and is understandable."

316 <https://telegrafi.com/jahjaga-dihet-kush-cka-ka-bere-ne-vitet-90/>

Moreover, as Rudin 1988a demonstrates for Bulgarian and Romanian, and as is also the case for Romani (McDaniel 1989) and the other languages, the fronted question words form a constituent, as they cannot be separated by clitic pronouns or by discourse parentheticals. Further, Rudin argues convincingly for Bulgarian and Romanian, and so also for the other languages, that WH-words are arranged with respect to one another in a hierarchical structure rather than exhibiting a flat structure.[317]

Greek differs from the other Balkan languages in that even though it allows multiple instances of a WH-word in a sentence, it is not a multiple-WH-fronting language.[318] Thus, the equivalents of (7.266) are ungrammatical with multiple fronted WH-words; rather, only one of the WH-words can appear on the left edge, e.g.:

(7.266) Ποιος χτύπησε ποιον; (Grk; cf. *Ποιος ποιον χτύπησε;)
who.NOM hit-3SG whom.ACC
'Who hit whom?'

The ability of multiple WH-questions to occur on the left edge is a property found all over Slavic; Rudin 1988a reports that Russian, Polish, and Czech are multiple WH-fronting languages, and others could be added, including, within South Slavic, BCMS (Browne 1976).[319] Moreover, it is found in various languages outside of Slavic and Romance, e.g., Hungarian (Tomaszewicz 2011) and Armenian (Tamrazian 1994), as well as Albanian and Turkish, as noted above. It is thus fair to raise the question of how striking the Balkan convergence actually is. The multiple fronting in Balkan Slavic could be a matter of inheritance and the Balkan Romance fronting could conceivably be an independent innovation, since geographically and typologically distinct languages like Armenian can show this pattern.[320]

However, the particularity of hierarchical structure for the configuration of fronted elements in Balkan Slavic sets Bulgarian and Macedonian off from other Slavic languages, including BCMS, suggesting that there is a special aspect to the Balkan Slavic realization of this characteristic. And, within Romance, the Balkan Romance facts are unique, since other Romance languages, e.g., French or Italian, do not even allow fronting of multiple WH-question words. It is thus reasonable to draw the inference that the Balkan Slavic pattern represents an innovation, and, that further, given the distribution of the feature within Romance, being restricted just to the Balkans, Romanian and Aromanian acquired this pattern through contact with Balkan Slavic speakers. As already noted, in Turkish such multiple fronting is to be expected on typological grounds.

317 McDaniel 1989: 590 indicates that the hierarchy is preferred but not strict in the Prishtina Arli that is the source of her data. See also footnote 315 on Aromanian.

318 There are languages that do not allow more than one WH-word in a sentence, e.g., Irish and Somali (Stoyanova 2008).

319 Slovak too, for instance, based on information from Dr. Marcela Michalkova (p.c. 2004).

320 In principle, Hungarian could show this multiple fronting pattern due to contact with Slavic languages, including the possibility of a Pannonian substrate. Eastern Uralic, like Turkic, is SOV, but Western Uralic is SVO so there are various possible contact scenarios.

The occurrence of multiple question words in a prominent position in the sentence is presumably a salient feature of surface structure that would be noticeable in conversational interactions between speakers of a multiple fronting language and speakers of a language allowing just a single WH-word to be fronted. Thus the spread of the pattern from Slavic into Balkan Romance can be taken to be a matter of calquing by Romance speakers of the surface positioning shown by these question words in Slavic. Alternatively, assuming that it was not the case that Slavic speakers shifted to Romance, the appearance of the multiple fronting in the usage of Balkan Romance speakers could reflect a reverse interference effect into Romanian and Aromanian based on their use of Slavic. The parallelism between Balkan Slavic and Balkan Romance, however, goes deeper than just the surface position the question words take, for the languages converge as well on the matter of a hierarchical configuration of the fronted elements relative to one another. If borrowing and transfer across languages is a surface-oriented phenomenon, as argued in §3.2.1.7 and §3.2.2.8, it can be assumed that Balkan Romance speakers in essence reconstructed for themselves the hierarchical structure based on cues in the observed surface patterns, cues such as the absence of any intervening material between the fronted question words. Here the fact that Aromanian is less strict than Romanian could have to do with social or historical differences in the calquing.

It can be concluded, then, that there is a characteristic Balkan variety of multiple-WH-fronting, and that it forms a point of syntactic convergence between all the Balkan languages except Greek, with Aromanian being between Greek and the rest of the Balkan languages.

7.8.4 Double Accusative Clauses

There are several constructions in the Balkans that show two accusative case-marked nouns governed by the verb where there are matchings across various languages.[321] One accusative is the affected patient (the direct object) while the other covers a wide and not necessarily unified range of grammatical/semantic relations, including recipient (indirect object), means of filling, and object of privation. Sandfeld 1930: 201–203 draws attention to the double accusative construction with 'teach someone something,' and 'ask someone something' in Romanian, Albanian, Greek, and Macedonian (but not in Bulgarian, where, as Sandfeld points out, the indirect object marker *na* precedes 'something,' corresponding to the dative elsewhere in Slavic). For some of these languages, the occurrence of double accusative construction(s)

321 Some dialects of Romani permit a double accusative construction with both possessor and possessum in the accusative, but this appears to be a dialectal rarity (see §7.8.2.2.6 for the usual Romani construction).

most likely represents an inheritance. In Greek, for instance, the construction is clearly old, found in Ancient Greek and New Testament Greek not only with these verbs but with other verbs as well e.g., 'hide' and 'deprive' (cf. Jacquinod 1989; Blass & Debrunner 1961: §155, 159).[322] For Romanian, this construction may continue a Latin syntagm, but it is interesting, as Sandfeld notes, that in no other Romance language was the Latin type preserved; this suggests that its occurrence in Romanian may be a matter of a retention in a contact zone (cf. §3.2.2.8).[323] For Albanian, the prehistoric situation is not clear, so the provenance of this construction cannot be assessed. The double accusative with *učі* 'teach' in Macedonian is clearly an innovation, but double accusatives do occur in OCS, as seen in Matthew 4:19, given here in the version of Codex Marianus (Jagić 1960), the received Greek text, and some translations into other Balkan languages:[324]

(7.267) a. stvor[j]ǫ va lov[ь]ca čl[ově]komъ (OCS; Mar.)
make.PRS.PFV.1SG you.DU.ACC catchers.DU.ACC people.PL.DATP

b. ποιήσω ὑμας ἁλιεῖς ἀνθρώπων (NTGrk; Rcvd text)
make.1SG.PRS.PFV you.PL.ACC fishermen.ACC people.PL.GEN

c. do t'ju bëj peshkatarë njerëzish (Alb; Bible 1980)
FUT DMS'you.PL.ACC make.1SG.PRS fishermen.ACC people.PL.ABL

d. va vã fac piscari di oaminji (Aro; Bible 2004)
FUT you.PL.ACC make.1SG.PRS fishers.ACC of people.PL.GEN
'I will make you fishers of men'

Here the influence of the Greek on the Balkan Slavic seems clear. All the modern non-Balkan Slavic languages, including the former Serbo-Croatian, have 'fishers' in the instrumental rather than the accusative, but all subsequent Balkan Slavic versions keep the accusative. It is worth noting here that Ostromirovo Evangelie, the oldest East Slavic text, follows the OCS translation and uses a double accusative, but this was arguably the result of Balkan Slavic influence in the East Slavic Church Slavonic tradition.

Sandfeld also notes that with verbs of supplying, furnishing, and filling, there is a double accusative in Greek that is also found in Macedonian and Albanian. The Greek construction is quite old, cf. Sandfeld 1930: 202; New Testament Greek has ἐγέμισεν αὐτὸ ὕδωρ 'he filled it [with] water' (*Evang. Thomae* 11.2 [Tischendorf 1876]), and Modern Greek has γεμίζει το ποτήρι νερό 'he.fills the glass [with] water.' In Albanian and Macedonian, most of Sandfeld's 1930: 202–203 examples

322 Feldman 1978 denies that Ancient Greek had a construction in which an indirect object is realized as if it were a direct object, based on the syntax of ἐπιτάσσω 'lay upon.' Lavidas 2013 and Benedetti 2020, however, affirm the traditional view that there is such a construction. Even if Feldman is right about ἐπιτάσσω, the fact that there are some verbs that participate in this construction is all that matters from the Balkanological perspective.

323 This construction may well be an inheritance from PIE in Greek and Latin, as it also occurs in Sanskrit with these verbs. For the emergence of this construction in later Balkan Slavic and its retention in Balkan Romance, its prehistory in Greek and Latin is irrelevant.

324 It is worth noting here that Balkan Romani NT translations do not use double accusatives in this context but rather favor an ablative for the 2SG pronoun.

come from folklore texts. This in turn indicates that such constructions in Macedonian and Albanian are either more limited in dialectal distribution or have retreated, perhaps under pressure from other languages. Consider, in this regard, examples (7.268ab) from Macedonian and (7.268cde) from Albanian:

(7.268) a. Go napolni drvenoto lule so tutun (Janevski 1956)
it.ACC filled.AOR.3SG wooden.DEF pipe with tobacco
'He filled the wooden pipe with tobacco'

b. Napolni go peralnikot vino (Popov 1953)
fill.IMPV.2SG it.ACC kettle.DEF wine
'Fill the kettle with [with] wine'

c. I dha çiftit të ri një sënduk plot me flori[325]
it.DAT gave.3SG pair.DEF.DAT PC young a chest full with gold.coin
'He gave the young couple a chest full of gold coins'

d. Daroviti atje plakën darova të mëdha (Sandfeld 1930: 202)
presented.3SG there old.woman.ACC presents PC big
'There he presented the old woman [with] large wedding presents'

e. Vetijën' e ka të mirë, zëmrënë plot mëshirë
feature.ACC it.ACC has PC good heart.ACC full compassion
shpirtinë me ëmbëlsirë (Sandfeld 1930: 203)
spirit with sweet
'He has the good feature, a heart full [with] compassion (and) spirit (and) with sweetness'

Example (7.268a) is what is now considered normal in Macedonian, while (7.268b) is intended to sound like rural dialect. Modern urban speakers would not use it. Similarly, (7.268c) is current standard Albanian, (7.268d) comes from a southern Tosk folktale originally published in the late nineteenth century with archaic vocabulary that needed to be glossed when it was republished in 1954 (Sakos 1954: 7–13), and (7.268e) is from a pamphlet by Naim Frashëri first published in 1896 (Jokl 1926), where it can be seen that the second accusative could also be expressed as a prepositional phrase in the same sentence. This double accusative construction dates in Greek to at least Hellenistic Greek as it occurs in the Septuagint (e.g., Exodus 31:3 ἐνέπλησα αὐτὸν πνεῦμα θεῖον 'I.filled him [with] spirit divine'; Blass & Debrunner §159[326]) though limited in the New Testament to being a *varia lectio* in Acts 2:28 (πληρώσεις με εὐφροσύνην 'you.will.fill me [with] gladness') for the more common genitive (εὐφροσύνης). Sandfeld thus sees Greek as the basis for this construction in the other languages. Koneski 1981: 107–108 notes that for Macedonian, such constructions are old (eleventh to thirteenth centuries) and even appear in some dialects of the former Serbo-Croatian. He argues that this double accusative construction was transitional between the loss of the instrumental and the rise of the preposition *so* 'with' to replace it in this function.

325 https://worldstories.org.uk/reader/the-three-brothers/albanian/458.

326 Interestingly, Blass & Debrunner 1961: §159 say this construction "need not be a pure Hebraism" because of its occurrence in Modern Greek.

Such is the case also with a 'small clause' construction with an accusative as the attribute of a direct object with verbs of considering, taking, making, etc., as in:

(7.269) a. Τα κορίτσια ήθελαν να τον πάρουν άνδρα (Grk)
the girls.NOM wanted.3PL DMS him.ACC take.3PL man.ACC
'The girls wanted to take him (as) husband' (lit., ' … take him husband')

b. Ai e ka marrë grua (Alb)
he.NOM her.ACC has.taken.3SG woman.ACC
'He has taken her as his wife'

c. Toj ja zede žena (Mac)
he.NOM her.ACC took.3SG woman.ACC
'He took her (as his) wife'

d. U are kãftatã nveast (Aro)
her.ACC has.sought.3SG bride.ACC
'He sought her as his bride'

e. Ja posaka [za] nevesta (Mac)
her.ACC sought.3SG [for] bride
'He sought her [for] his bride'

While the surface similarity here in the various languages is striking, this predicate accusative construction occurs in Ancient Greek, Latin, OCS (and elsewhere in Old Slavic languages), and Sanskrit, so it may not have any contact-related significance. Koneski 1981: 107 connects these constructions in Macedonian with the loss of case in Balkan Slavic.

However, Sandfeld points to one very particular instantiation of a double accusative with 'what' as a direct object and an accusative personal pronoun with the verb 'have' or 'want' in the meaning 'What does so-and-so have to do with X?' in Albanian, Aromanian, and Greek:

(7.270) a. Tsi lu ai? (Aro)
what him.ACC have.2SG
'What is your issue with him?' (lit., 'what do-you-have him?')

b. Τι με θέλεις; (Grk)
what me.ACC want.2SG
'What do you want with me?' (lit., 'what me you-want')

c. Ç' e do atë njeri? (Alb)
what him.ACC.WK want.2SG this.ACC man.ACC
'What do you want to do regarding this man?'
(lit., 'what him you-want this man')

d. Abe što me gnjaviš (Mac)
hey what me.ACC bother.2SG
'Hey, what are you bothering me for?' (lit., 'what me you.bother')

Given the specificity of the pattern and the meaning, it is likely that this convergence represents a calque based on one of the languages, though which one was the source is not clear.

7.9 Prepositional Syntax

Prepositions, or more accurately adpositions, figure prominently in the syntax of the individual Balkan languages and there are numerous commonalities to be noted among the languages with regard to adpositions. Asenova 2002: 97–104 gives an overview of isosyntagms with prepositions that begins with a display of the main prepositions and serves as the basis for the discussion here in Table 7.9, to which we have added Judezmo, Romani, and WRT.

Not all prepositional similarities are related to language contact. Some have resulted from internal developments of a sort found in many languages outside of the Balkans and thus not particularly unusual. For instance, the occurrence of composite prepositions such as Greek μαζί με 'together with,' Alb *nëpër* 'among' (*në* 'in' + *për* 'for'), Mac *pomegju* 'between' (*po* 'by' + *megju* 'between'), Aro *piningã* 'between' (*pi* 'on' + *ningã* 'along, near') is striking, but can be understood as structural innovations that any language can develop, as shown by forms like English *into* and French *devant* 'in front of' (from *de* 'from' + *avant* 'before'). Here it is also worth noting that in both Romani and Turkish, case still functions for some of these usages. The postpositional usages of Turkish and, sometimes, Romani and elsewhere reinforce the utility of *adposition* as the terms of choice.

Adpositional developments that do have some involvement with contact, however, often fit more comfortably into sections of chapters other than syntax proper. For instance, the use of prepositional marking in place of older cases, as with the emergence of *na* in Balkan Slavic for various older genitive, dative, and locative functions, while tied to the contact-related replacement of case, is treated in §6.1.1, under the rubric of the morphosyntax of nominal case. Moreover, there is considerable borrowing of adpositions, as documented in §4.3.3.2, and in the usage of particular adpositions and the spread of some prepositional combinations, there is strong evidence of contact effects as well, as discussed in §4.3.3.2 and §4.3.10.2.[327]

Here, a few remaining details are covered that have to do with aspects of the syntax of prepositions in contact situations in the Balkans.

Table 7.9 *Prepositional isosyntagms*

Balkan Slavic	v(o),văv / na	za	ot/od	s(o), săs	do
Balkan Romance	în / la, a	pentru	de	cu	pînă [la]
Albanian	në	për	nga / prej	me	gjer [në] / deri
Greek	σε	για	από	με	ως
Judezmo	en/a	para	de	kon	asta, kadar
Romani	LOC, to, ko, adv.	DAT	ABL, TAR-O	INSTR	dži ko
WRT	LOC (DAT)	için	ABL	ile	kadar
DEFINITION	'in/at/on'	'for'	'from/of'	'with'	'up to'

327 Prepositional usage is traditionally considered to be an aspect of syntax, as in Smyth 1920 on Ancient Greek. Cf. Asenova 2002: 97–104, who treats these phenomena as morphosyntax. Here such usages are treated as a matter of lexical semantics (§4.3.3.2).

7.9.1 Adposition Order

In §4.3.3.2, instances of adpositional borrowing are detailed, and a key syntactic fact is observed: for the most part, items that are postpositional in Turkish become prepositions when borrowed into Balkan languages. This tendency can be explained by the typological fact that the Balkan languages other than Turkish are generally head-initial languages, so that the head of an adpositional phrase, the adposition itself, would typically be positioned before its object.[328] Importantly, as noted in §4.3.3.2, there are some exceptions to this tendency, such as postpositional *gibi* 'like' in Ottoman-era Edirne Greek, and Aromanian postpositional *carshí* 'opposite,' both from Turkish (*gibi / karşı*, respectively), and this latter form can occur as postpositional *karši* in colloquial Macedonian, although normally it would be prepositional. Romani dialects with Turkish conjugation (§6.2.1.1.1) can even treat native prepositions as postpositions on a Turkish model, e.g., Sliven Nange *štar zisendar palal* 'after four days' (lit., 'four days.ABL after'), where *palal* 'after' is native and the ablative on 'days' calques the Turkish case usage. These differences speak to the degree of integration of these loans, but also to the flexibility of various languages when it comes to absorbing both the form and the syntax of borrowed items.

7.9.2 Definiteness Omission

Various sources on the Balkan languages give an interesting-sounding and seemingly very precise parallel between Albanian and Romanian involving the preposition 'with' (Alb *me*, Rmn *cu*) that is worth discussing here because it is much cited but also because, despite that, it is not at all clear how much Balkanological significance there is to it.[329] Moreover, Prendergast 2017 has provided an exhaustive account of what he calls determiner omission in Balkan Slavic, Balkan Romance, and Albanian with locative prepositions.

The parallel (or possibly convergence) has to do with definiteness marking in prepositional phrases. The general pattern with prepositions in these languages is that definiteness is typically not marked on certain simple (i.e., unmodified) prepositional objects; compare for 'in the house' Rmn *la casă*, with indefinite *casă*, not *la casa, with definite *casa*, and Alb *në shtëpi*, with indefinite *shtëpi*, not *në shtëpinë, with definite accusative *shtëpinë*; cf. Mac *vo bolnica* 'in [the] hospital,' *na fakultet* 'at the department,' with no overt definite marker.[330] Cf. also Rmn *fără băiat* with indefinite *băiat* (not *fără băiatul, with definite *băiatul*), Alb *pa djalë*

328 A similar change befell the one inherited postposition in Balkan Slavic: *radi* 'for the sake of, because of' was postpositional in OCS, but in those Slavic languages where it is preserved outside the Balkans, e.g., BCMS and Russian, it can be both preposed and postposed, depending on the expression (e.g., BCMS *radi mene* 'for my sake' but *Boga radi* 'for God's sake!,' Russ *čego radi* 'what for?' but *radi Boga* 'for God's sake!'). In Balkan Slavic today, *radi* is not current, but *zaradi* 'because of' and *poradi* 'on account of' indicate the shift to preposition at some point.

329 See below, just after (7.271) for a listing of said sources.

330 Note that these phrases could also mean 'in a house,' 'in a hospital,' etc.

(not *pa djalën, with definite *djalën*), for 'without the boy.' In Macedonian, the article is permissible with *bez* 'without' but would be omitted for pragmatic reasons in context. Note that if a modifier is added, then definiteness is required if appropriate, e.g., Alb *në shtëpinë e bardhë* 'in the white house.' Within Romance, Nedelcu 2013b: 460 writes that "the non-realization of the article in the context of a preposition is specific to Romanian," which, in fact, is not so, at least as far as the Balkans are concerned, so that one might be inclined to think of contact as the source of this apparent innovation. Moreover, according to some sources, Romanian and Albanian agree on 'with' being an exception to this pattern, taking simple objects that are always definite:

(7.271) a. cu prietenul 'with the.friend' / *cu prieten (INDF) (Rmn)
Plec cu trenul 'I.leave by (the.)train' / *cu tren (INDF)
b. me shqiptarët 'with the.Albanians' / *me shqiptarë (INDF) (Alb)
me hoxhën 'with the.priest' / *me hoxhë (INDF)

Michov 1908 observed such facts but Boretzky 1968: 134 seems to be the first to explore them carefully, aided by access to more material. The suggestion of a parallel was picked up by Schaller 1975: 169–170, who was cautious about its significance, and then by Banfi, who was more positive, including it in a listing of "Tratti linguistici interbalcanici: 'I balcanismi'" ('inter-Balkanic linguistic traits: Balkanisms') under the heading of syntax ("Sintassi"). Joseph 1999a, drawing on Banfi's formulation, suggests that it may be a syntactic reflex of the old Albanian-Romanian connection that has been much remarked upon in the literature as to the lexicon (see §1.2.1.4 and §4.2.1.1) and to which Hamp 1989a attributes an accentual parallel between the two languages (see §5.2).

The facts, however, are far more complex than the simple formulation given above and in Banfi (and thus Joseph) would suggest. Boretzky, for instance, notes numerous cases in Albanian folktales and other sources of definite objects with prepositions *për*, *në*, and others, and examines differences between modified and unmodified objects, where the type of modification (adjectival or genitival) seems to play a role. Ultimately, the context and speaker's (or writer's) intent seems to be paramount, and one can only speak at best of a tendency towards definiteness of objects of 'with' and indefiniteness of objects of other prepositions in the two languages, but no direct parallelism. This means that there might have been a prepositional restriction involving definiteness and 'with' in whatever prehistoric speech community underlay the intriguing and seemingly old Albanian-Romanian parallels, but the exact details are not fully recoverable, and certainly each language has gone its own way as far as the nexus of definiteness and prepositions is concerned.[331] However, the difference between Romanian and the rest of

331 As a parallel to the elusiveness of enunciating hard and fast rules here, one can note differences in definiteness marking seen in the comparison of British English *to hospital* / *in hospital* and American English *to the hospital* / *in the hospital*, with the further complication that American English allows a British-like article-less locution for an institution as prepositional object in *to school* / *in school*. Thus there is probably a combination of individual speaker preferences, context, meaning, and lexical idiosyncrasies at work in these cases in English and in the Balkans.

Romance in this regard is striking and further study may reveal a contact component to this divergence.[332] Moreover, Balkan Slavic may prove a rich area in which to explore the question of definiteness in prepositional phrases; Mladenova 2007: 175, for instance, notes that an expression with an indefinite, such as *dokraj selo* 'up-to-end village,' and one with a definite, e.g., *do kraja na seloto* 'up-to end-the of village-the,' for 'up to the end of the village' show a "subtle semantic difference" that would be perceived "mostly as a stylistic one"; cf. Koneski 1967: 232–236, 507–534 for Macedonian, which has a number of prescribed semantic differences connected to indefinite versus definite prepositional phrases.

One contact-related aspect to this is suggested by Boretzky 1968: 139–140. Noting that in texts from Leskovik, a village in southern Albania near the Greek border, there is a greater incidence of definiteness in prepositional objects than elsewhere in Albanian, Boretzky speculates that Greek, with its greater freedom in article use than in Albanian, may have exerted influence on this dialect, though he admits that other dialects in contact with Greek do not necessarily show any such effects.

Prendergast 2017 examines what he calls determiner omission in locative prepositional phrases (his LDO) in Macedonian, Albanian, Aromanian, and Romanian. Based on this phenomenon, he argues that a linguistic area is best viewed as the iterative layering of contact between local speech forms, rather than diffusion of features across languages as a whole. He thus describes (pp. 149–152) the development of LDO in the following way, worth quoting at length as it offers a comprehensive model as well of Balkan contact (and see also §8.0 for further discussion and a particularly revealing diagram):

> Many other grammatical convergences had to occur before LDO could spread, making this grammatical feature … a representative of particularly dense layering of contact situations and accommodations. [LDO] as an areal feature depends on the numerous previous periods of contact and convergence that resulted in shared Balkan grammatical features. … [In particular] the development of the post-posed definite article [made it possible] to show a distinct means of expressing definite reference through the omission of an article. … [And, the] collapse in the distinction between goal and location also provides the semantic domain for triggering LDO … [allowing for speakers to] model the grammatical pattern from one language to another. These features first developed during contact between Albanian and Late Latin, but after the introduction of Slavic into the Balkans during the 6th and 7th centuries, they were reinforced by a dynamic of mutual multilingualism among Albanian, Balkan Romance, and Balkan Slavic (Lindstedt 2000: 235). … [As other features emerged,] Balkan Slavic thus developed a post-posed definite article, a loss of noun case, and a collapse in the semantic distinction between location and goal. However, only Balkan Romance showed LDO during this period. … LDO did not develop [more widely] until later, at more local levels of interaction centered on southern Albania and southeastern Macedonia. As the patterns of multilingual interaction began to change with the introduction of nationalism into the Balkans and the break-up of the Ottoman Empire in the 19th

332 It could of course be the rest of Romance that diverged from Romanian, since there seems to be a Western European dimension to the use of the definite article with prepositional objects.

> century, speaker dynamics changed as well. … At a local level, Aromanian speakers began to assimilate into Albanian and Macedonian dialects. At this time, some Albanian dialects already showed loss of the locative case, but not the full development of LDO. The origin of LDO in northern Geg is unclear, but may have been the outcome of earlier assimilation of Balkan Romance speakers into Albanian. The development of LDO in southern Tosk Albanian dialects can be more clearly tracked as an outcome of Aromanian assimilation. Macedonian dialects were also subject to Aromanian assimilation, and LDO appears most consistently in Macedonian dialects that show the greatest effect of Aromanian on their grammatical system. In this way, several iterated periods of contact were necessary for the development of LDO.

Prendergast's conclusion is significant, as he argues for contact-related effects surrounding the interplay of definiteness and prepositional usage in the Balkans, both in the distant past and in more recent time; moreover, that is precisely the idea of layered multilingual contact that is relevant here.

7.9.3 Comitative Coordination

The preposition 'with' figures in another possible, but again somewhat inconclusive, syntactic parallel within the Balkans. In various Balkan languages, it is possible for a prepositional phrase headed by 'with' – Alb *me*, Grk με, Mac *so*, Aro *cu*, etc. – to combine with a subject to determine plural verb agreement, as if it were a simple conjunction meaning 'and.' This has been discussed by Kazazis & Pentheroudakis 1984 for Greek, who give examples like (7.272a), and similar examples can be found for Albanian (7.272b) and Macedonian (7.272c):

(7.272) a. Η Ζντούπω με τον Κλεινία χορέψανε ωραία στο πάρτι (Grk)
the Z.NOM with the K.ACC danced.3PL well at.the party
'Z. and K. danced well at the party' (lit., 'Z with K they.danced … ')

b. Unë me gruajën time po arritojmë së shpejti (Alb)
I. with wife. mine PROG arrive.1PL PC.F.DAT soon
NOM ACC
'My wife and I will be arriving soon' (lit., 'I with my wife we will arrive … ')

c. Toj so Viktor otide / otidoa (Mac)
he.NOM with V. left.3SG left.3PL
'He left with Victor / He and Victor left' (lit., 'he with V. he.left / they.left')

Friedman 1993d: 289 notes regarding the distinction between singular and plural in (7.272c) that "The use of singular agreement is favoured by the younger generation of Skopje speakers and is considered correct by strict normativists; the use of plural agreement is favored by some older speakers – who consider it more literary – and in the southwest."

This "comitative coordination" construction is found elsewhere in Slavic, e.g., Polish (Dyła 1988) and Russian (McNally 1993) and is cross-linguistically common (cf. Aissen 1989 regarding Tzotzil) even if not well instantiated in Indo-European. It is found, moreover, in Ancient Greek, although it is rare. Smyth 1920:

§953 states that "a subject in the singular, followed by a clause containing the preposition μετά 'with,' rarely takes a plural verb," and he gives the following example from Xenophon:

(7.273) Ἀλκιβιάδης μετὰ Μαντιθέου … ἀπέδρασαν
A.NOM with M.GEN … ran.off.3PL
'Alcibiades and Mantitheus escaped' (*Hellenica* 1.1.10)

There is no way to be sure of the age of the construction in Albanian, and it is possible too, of course, that the Modern Greek construction does not continue the Ancient Greek one directly. Moreover, given the uncertainty discussed in §4.3.3.2 concerning the historical derivation of Modern Greek με 'with' and the relationship between it and Albanian *me*, i.e., whether they are cognate forms or instead one is a borrowing from the other, one might speculate that the comitative construction at least in Albanian is a case of the syntax of an element being borrowed along with the element itself (as seen in §7.9.1 with postpositions). A further contact-related dimension to this construction comes from the geography noted above as to where it is favored in Macedonian. Given the presence of this construction in all branches of Slavic, the current resemblance between the Modern Greek and Macedonian has many possible routes of formation, but such detailed speculation is beyond the bounds of the present study. At issue are questions such as: Did the *with = and* construction disappear from pre-Modern Greek and then reappear in Modern Greek? If it did not disappear, what evidence is there regarding its frequency? On the Slavic side, one could argue that this construction in Macedonian is inherited from Common Slavic, but there is a question of the role of Russian Church Slavonic in the nineteenth and early twentieth centuries. Crucial here would be a comparison of standard language corpora and dialect corpora, i.e., regionally identified colloquial corpora, coordinated with whatever texts are available from the late middle ages until the nineteenth century or so. All one can do is signal possibilities without fully resolving them.

7.9.4 Various Prepositional Calques

Sandfeld 1930 mentions a few scattered convergences involving prepositions that are noteworthy but do not generalize to any broader patterns. Asenova 2002: 97–104 provides a summary of all the relevant phenomena for standard Albanian, Bulgarian, Greek, and Romanian. They could in principle be subsumed under a rubric of calques, and as such they would be appropriate to Chapter 4 on lexical convergences; indeed, some are discussed in §4.3.3.2 and §4.3.10.2. However, in these cases, the calquing involves aspects of syntax as well, extending to such features as case government and combinatorics for lexical items more generally. Accordingly, treatment here under prepositional syntax is equally appropriate.

Thus, for example, Sandfeld 1930: 118 notes that in Aromanian, the preposition *la* meaning 'at the house of' (French *chez*) can occur with the nominative case, e.g., *la io* 'at my place (*chez moi*),' unlike its syntax in Romanian, where it governs

accusative case (*la mine* 'at me,' not **la eu*). In Macedonian, as in the rest of Slavic, the equivalent preposition takes an oblique case, e.g., Mac *kaj mene* 'at my place.' The Aromanian can be explained by reference to the syntax of Albanian *tek* 'to; at,' which regularly takes a nominative complement, e.g., *tek unë* 'at my house (*chez moi*).' The etymology of *tek* as a compressed correlative phrase **to-* ... *k*^{w}- 'there … where,' with a PIE demonstrative stem correlating with a WH-word (Hamp 1999; Maynard & Joseph 2000: 26), shows that the Albanian pattern is original to Albanian, so that the Aromanian usage is a result of contact with Albanian, with the process being either calquing on the part of Aromanian speakers modeling their usage after the Albanian pattern or transfer from Albanian speakers shifting to using Aromanian. Moreover, in the Montenegrin dialect of Mrkovići the preposition *ge* (from earlier *gdje*) can be used with a nominative as in *ge kuća* 'at home,' which is clearly a calque on Albanian usage (Morozova 2019). In these cases, the syntax of the target lexical item, in this instance case government, is affected by language contact.

Similarly, in Macedonian, Aromanian, and Albanian, in a construction that is generationally and geographically marked, the quantifier 'all' often occurs between the preposition meaning 'with' and the prepositional object, without adding a universal quantifier meaning (Sandfeld 1930: 120). Thus one finds *sose* in (7.274a), *cu tut* in (7.274b), and *me gjithë* in (7.274c):

(7.274) a. Pojde sose ofčarot (Mac)
arrived.3SG with.all shepherd
'He arrived with the shepherd'

b. le băgă n casă cu tut căpitanlu (Aro)
them put.3SG in house with all leader.DEF
'He put them in the house with the leader (of the gang)'

c. me gjithë të shoqën (Alb)
with all PC wife.DEF.ACC
'with his wife'

Because of the absence of this construction in Bulgarian and in this form in Romanian, Sandfeld sees it as originating in Albanian and spreading into the other languages via calquing on the part of Macedonian and Aromanian speakers. Transfer from Albanian speakers shifting to the other languages would seem to be a possible mechanism as well.[333] Whatever the process involved, the important lesson here is that the internal syntax of the prepositional phrase is disrupted by an element that comes to intervene between the preposition and its complement, even though the governor and governed in this instance generally form a tight bond that is not usually interrupted.

We can also note here the use of doubled prepositions. Markovikj 2015 discusses calquing of doubled prepositions from Aromanian into Ohrid Macedonian, whence it has spread to Skopje and the speech of educated standard speakers there, perhaps

333 In Romanian, *cu tot* 'with all' does occur, according to Sandfeld, but not with 'all' intercalated between the preposition and its object; rather, it is repeated after the object, e.g., *cu cal cu tot* 'with the horse' (lit., 'with horse with all').

via a mechanism of diffusion akin to the "parachuting," but in reverse, inherent in the "gravity model."[334] While all Slavic languages have some doubled prepositions with spatial meanings – e.g., in most of Slavic, *iz* 'from' + *pod* 'under' gives a compound preposition meaning 'out from under' and spelled variously *izpod, ispod, iz-pod* – nonspatial uses are much more limited, and there are also limitations on compound spatial uses. In Ohrid Macedonian, however, such doubled spatial uses go beyond what would occur elsewhere, e.g., *izlegov od v kupatilo* 'I.came.out from in[side] (the) bathroom' (Markovikj 2015: 297). As Markovikj argues, the expansion of doubled preposition usage is a result of calquing on Aromanian usage, where the phenomenon is typical for Balkan Romance. Examples (7.275ab) from Markovikj (p. 302) show how Ohrid Macedonian matches Aromanian:[335]

(7.275) a. nj' u d'edãrã di la lukur (Aro)
mi go dadoja od na rabota (Mac (Ohrid))
me.DAT it.ACC gave.3PL.AOR from at work
'They gave it to me from at work'

b. aist peshch njã l'ãps'eshti ti la iatru (Aro)
ovie ribi mi trebet za na lekar (Mac (Ohrid))
these fish.PL me.DAT need.3SG for at doctor
'I need these fish for at the doctor's'

This type of doubled preposition usage is now found in Standard Macedonian, where it would not have occurred fifty years ago, as in example (7.276) from Markovikj 2015: 297:

(7.276) Makedonija e podgotvena za vo NATO
Macedonia is prepared for in NATO
'Macedonia is prepared to enter NATO'

This same type of construction also occurs in Albanian, as in (7.277):

(7.277) I bënin gati fëmijët për në çerdhe
them.ACC make.PRS.3PL ready children.DEF.ACC.PL for at nest
'They get the children ready for nursery school'

The calquing of doubled prepositional constructions is a good demonstration that Balkan sprachbund processes continue even now. Although the usage in Ohrid Macedonian could be considerably older, the spread to Skopje and the standard is more recent and demonstrates that the interaction of local languages in the Balkans remains alive and well in the countries that have not engaged in repressive language policies. The Albanian parallel, which was taken from a news headline citing a Macedonian official, might be a calque on Macedonian or an older commonality with Aromanian.[336]

334 See §3.2.2.10 on this model of diffusion.

335 These usages do not occur in Bulgarian, which further supports Markovikj's hypothesis.

336 Cf. Asenova 2002: 203–204, who adduces similar parallels from OCS and standard Bulgarian. She labels the Albanian equivalents as less frequent. For the purposes here, the convergence between Skopje and Ohrid Macedonian is significant even if the construction is attested elsewhere. When

Asenova 2002: 97–104 points to a variety of prepositional parallels among the Balkan languages, such as the expansion of the functions of the preposition *za* 'for' in Balkan Slavic to more closely resemble usages of the corresponding prepositions in the other languages, a process she locates in the fourteenth to fifteenth centuries. Asenova also adduces convergences among the other prepositions in Table 7.9. We can also mention here the semantics of the Slavic inherited *iz* 'from out of [elative]' shifting to a perlative 'through, throughout' by analogy with the Turkish ablative (*-DAn*), which has perlative as well as ablative functions that are found precisely and only in Balkan Slavic *iz* (cf. §4.3.10.2).

Finally, we can note here that Sandfeld 1930: 204 draws attention to various convergences in the expression of being 'X-many years old,' seeing Greek as the ultimate source of the use of the preposition for 'from' or 'of.' That is, robustly in Romanian, one finds both *un copil de zece ani* 'a ten-year-old child' (lit., 'a child of ten years') and *copilul e de zece ani* 'the child is ten years old' (lit., 'child.DEF is of ten years'), and so also in Aromanian, though an expression with the verb 'have' is also possible: *căts an ai* (lit., 'how.many years have.you'), the construction found more widely across the rest of Romance. In Balkan Slavic, the preposition *ot/od* 'from' can be found dialectally with 'years,' though the more usual expression is with *na* 'of,' and in Albanian the ablative case is used (*jam njëzet vjetsh* 'I.am 20 years.ABL'). The use of *na* in Balkan Slavic is most likely simply the normal successor there to a construction with a genitive usage found in Old Church Slavonic in similar expressions, alongside a dative. For Slavic *ot*, Balkan Romance *de*/*di*, and even the Albanian ablative, Sandfeld looks to Greek as the starting point via calquing, a reasonable hypothesis because the use of the genitive is old in Greek, dating back at least to New Testament Greek, and because the prepositions that are found are appropriate renderings of a genitive; the Albanian ablative is perhaps harder to motivate in terms of the Greek genitive, although one can point to historical connections between genitive and ablative across the Indo-European family.[337] In this instance, then, under this account, the range of use of the prepositions *ot* and *de*/*di* is extended by the effects of language contact.

doubled prepositions occur in Bulgarian, which is rarely, many educated speakers consider them characteristic of uneducated speech.

337 For instance, genitive and ablative are syncretic in all noun classes in Sanskrit except the *a*-stems (and presumably so also in PIE), and within Greek some local uses of the genitive are best explained as having been taken over from the older ablative. In Slavic, genitive and ablative had already merged by the time of Common Slavic.