

## EXPLORATIONS IN LINGUISTIC TYPOLOGY

GENERAL EDITORS: Alexandra Y. Aikhenvald and R. M. W. Dixon  
*Research Centre for Linguistic Typology, La Trobe University*

This series focuses on aspects of language that are of current theoretical interest and for which there has not previously or recently been any full-scale cross-linguistic study. Its books are for typologists, fieldworkers, and theory developers, and designed for use in advanced seminars and courses.

### PUBLISHED

#### 1 Adjective Classes

*edited by*

R. M. W. Dixon and Alexandra Y. Aikhenvald

#### 2 Serial Verb Constructions

*edited by*

Alexandra Y. Aikhenvald and R. M. W. Dixon

#### 3 Complementation

*edited by*

R. M. W. Dixon and Alexandra Y. Aikhenvald

#### 4 Grammars in Contact

*edited by*

Alexandra Y. Aikhenvald and R. M. W. Dixon

### PUBLISHED IN ASSOCIATION WITH THE SERIES

Areal Diffusion and Genetic Inheritance:

Problems in Comparative Linguistics

*edited by*

Alexandra Y. Aikhenvald and R. M. W. Dixon

OXFORD

# Grammars in Contact

## *A Cross-Linguistic Typology*

*edited by*

ALEXANDRA Y. AIKHENVALD and

R. M. W. DIXON

Research Centre for Linguistic Typology

La Trobe University

OXFORD  
UNIVERSITY PRESS

- Matras, Y. 1998. 'Utterance modifiers and universals of grammatical borrowing', *Linguistics* 36: 281–331.
- Nadkarni, M. V. 1975. 'Bilingualism and syntactic change in Konkani', *Language* 51: 172–86.
- Rayfield, J. R. 1970. *The language of a bilingual community*. The Hague: Mouton.
- Reed, C. E. 1947. 'The question of aspect in Pennsylvania German', *Germanic Review* 20: 5–12.
- Salmons, J. 1990. 'Bilingual discourse marking: code-switching, borrowing, and convergence in some German-American dialects', *Linguistics* 28: 453–80.
- Sauder, B. 1955. *Der Nachbar an de Schtroas*. Toronto: Pennsylvania Folklore Society of Ontario.
- Thomason, S. G. 1986. 'Contact-induced language change: possibilities and probabilities', pp. 261–84 of *Akten des 2 Essener Kolloquiums zu Kreolsprachen und Sprachkontakte*, edited by W. Enninger and T. Stolz. Bochum: Studienverlag Dr. N Brockmeyer.
- and Kaufman, T. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley and Los Angeles: University of California Press.
- Weinreich, U. 1953. *Languages in contact*. New York: Linguistic Circle of New York Publication No. 2.

## Balkanizing the Balkan Sprachbund A Closer Look at Grammatical Permeability and Feature Distribution<sup>1</sup>

VICTOR A. FRIEDMAN

### 1 Introduction

The Balkan sprachbund<sup>2</sup> was the first area of contact-induced language change to be identified as such (Leake 1814: 380; Kopitar 1829: 86). Owing perhaps to this pedigree, it has been described as 'the most studied' and 'most famous' example of its kind, but, owing to the known antiquity and complexity of multilingualism in the region, it has also been described as 'notoriously messy' (Thomason and Kaufman 1988: 95). Hamp (1989) noted that one could speak of a plurality of *Sprachbünde* (for which I would now use sprachbunds) in former Yugoslavia, while Masica (2001) provides a useful summing up of the question of defining a linguistic area, an issue with which he has been occupied for many years (Masica 1976). Aikhenvald (Chapter 1) raises the fundamental question of the relationship between sociolinguistic history and typological structure to the nature and progress of contact-induced change. Moreover, many recent typological and other studies (e.g. Haspelmath 1998; van der Auwera 1998; Lindstedt 2000) treat complex Balkan

<sup>1</sup> Much of the research and thinking for this chapter was done while I was a Fellow at the Research Center for Linguistic Typology at La Trobe University, August–November 2005. I am grateful to Bob Dixon and Sasha Aikhenvald for the opportunity to work in such a peaceful, productive, and supportive atmosphere. Uncredited examples are from fieldwork. All English translations are my own.

<sup>2</sup> Among the proposed glosses for sprachbund are 'linguistic league', 'linguistic area', 'convergence area', and 'diffusion area', but here I shall treat sprachbund as a loanword into English, like French genre, so it will be neither capitalized nor italicized.

phenomena as unitary, ticking off points or drawing lines in such a way that the facts 'on the ground' disappear from view.<sup>3</sup> It is in this context of relative causation and typological study that I wish to examine some fundamental assumptions about the Balkan languages and argue that when viewed at close range, some apparent commonalities of the Balkan sprachbund reveal different cleavages that reflect differences in shared history, and that the relative susceptibility of contact-induced change in a given system shows language-ideological (sociolinguistic) as well as typological factors at work. At the same time, keeping these details in mind makes for better-informed typological studies that reflect more clearly existing linguistic situations.

This is not to say, with Andriotis and Kourmoulis (1968: 30), that the Balkan sprachbund is 'une fiction qui n'est perceptible que de très loin' and that the commonalities are 'tout à fait inorganiques et superficielles'. To the contrary, Balkan linguistic diversity occurs within the context of a set of structural similarities that comprise a framework of contact-induced change. Moreover, we must distinguish between 'superficial' and 'surface'. As Joseph (2001) has persuasively argued, surface realizations constitute the locus of language contact, and explanations that appeal to typological aspects of universal grammar (including so-called formalist 'explanations') tell us nothing about language contact (Aikhenvald makes a similar point in Chapter 1). Also, surface realizations are by no means 'inorganic'; they represent convergences that are evidence of the multilingualism that we know existed for centuries and even millennia. Nonetheless, while rejecting the notion that the Balkan sprachbund is a fiction, we must place the differences in the context of the similarities. In this chapter, then, I shall, as it were, Balkanize the Balkans by examining certain cleavages—particularly with respect to future marking and referentiality—and discuss the nature of their areality. In so doing, I hope not only to produce a more nuanced picture of the most famous sprachbund, but also to argue that, like all language change, degrees of convergence can take place with varying speeds. At the same time, however, based on our available documentation, processes that may have been set in motion, or at the very least begun to be reinforced during the middle ages, achieved their

<sup>3</sup> A problem with van der Auwera's (1998) methodology, for example, is the dependence on standard languages and the treatment of features as unitary. In his approach, Bulgarian comes out as 'more' Balkan than Macedonian owing to the presence of stressed schwa. Leaving to one side the inadequacy of his list (e.g. the Balkan conditional in Macedonian is more grammaticalized than in Bulgarian), there are two realities that his approach fails to capture. One is the fact that the majority of Macedonian dialects do have stressed schwa (assuming that one even wants to accept this as a Balkanism) while a minority of Bulgarian dialects (Teteven-Erkech, some Rhodopian) do not. The other is that object doubling is fully grammaticalized in (West) Macedonian, whereas it is a pragmatic feature in all of Bulgarian. What is needed is not a map of language names, but of regions, as well as distinctions for degree of grammaticalization.

current state during the *Pax Ottomanica* (and it is telling that this same period is referred to in Bulgarian as *turskoto igo* 'the Turkish yoke'). Moreover, the effects of the end of that historical period have shown a combination of mutability and resiliency, which, at this early stage, can only be hinted at.

## 2 The Balkans and Balkanization

A number of studies have argued that the Balkans as a concept were constructed during the late nineteenth and early twentieth centuries by West European discourse as a negative 'Other' against which it could define itself positively in much the same fashion as it constructed the Orientalist discourse identified by Edward Said (Bakić-Hayden 1995; Todorova 1997). Thus, for example, the use of the term *Balkanization* to mean 'break up into tiny entities' does not date from the nineteenth century, when the Balkan nation-states were separating from the Ottoman Empire, but rather from the end of the First World War, when the Austro-Hungarian, German, and Russian empires were resisting the national self-determination that led to the creation of Czechoslovakia, Poland, and the Baltic republics. This construing of the Balkans was part of the process by which Western Europe constituted itself as 'Europe', a process that in some ways continues to this day.

It is arguable that the linguistic construction of the Balkans as a significant contact zone, i.e. as different from the orderly genetic differentiation posited for Western Europe, was part of this same process of West European differentiation. The most extreme formulation of this was Schleicher's (1850: 143), when he described Balkan Romance, Balkan Slavic, and Albanian as forming a group of different languages 'agree[ing] only in the fact that they are the most corrupt in their families'.<sup>4</sup> It can be argued that a mirror image of Schleicher's view was Whorf's (1938) S[tandard]A[verage]E[uropean], which was basically a linguistic equivalent of this same West European Germano-Romance creation of Europe. Relating to this vein of occidentoeurocentrism, just as the Balkans have been deconstructed by modern historians, so, too, some Balkan linguists, e.g. Reiter (1994), and more recently the EURO-TYP project, have argued that the Balkans do not constitute a special case but rather a part of SAE, albeit a periphery of the core (!) (cf. Haspelmath 1998). Nonetheless, the political separation of south-eastern Europe from the early modern upheavals that led to the creation of 'Europe' qua Europe—the so-called

<sup>4</sup> 'Es ist eine bemerkenswerte Erscheinung, dass um die untere Donau und weiter nach Südwesten sich eine Gruppe aneinandergränzender Sprachen zusammengefunden hat, die bei stammhafter Verschiedenheit nur darin übereinstimmen, dass sie die verdorbensten ihrer Familie sind!'

*Pax Ottomanica*—had linguistic consequences. While some of these have been erased in the ensuing century or so of ‘national liberation’ others remain, and this, too, is of interest to our topic. In this sense, I will be discussing the linguistic Balkanization of the Balkans, i.e. the creation of a relatively unified linguistic area owing to centuries of multilingual contact—the very opposite of political ‘Balkanization’.

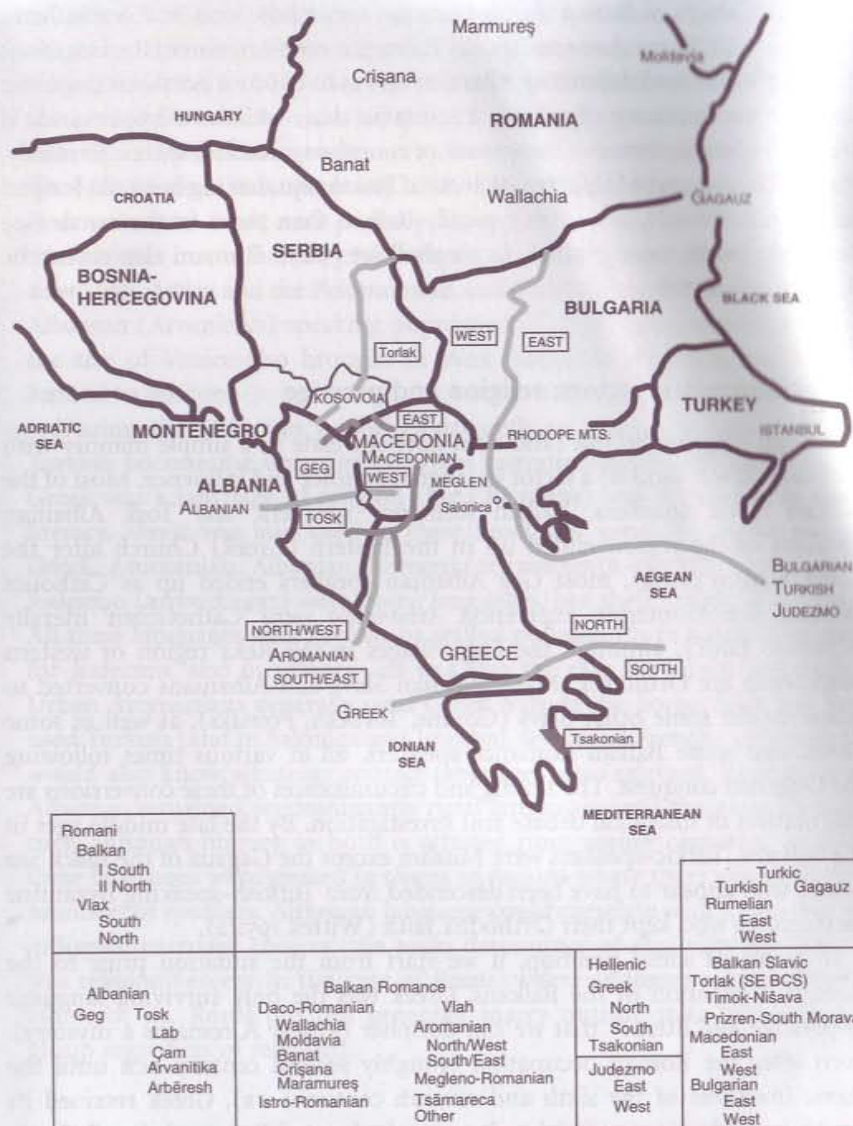
### 3 The Balkans

#### 3.1 *The Balkan languages*

It is certainly the case that anything *can* be borrowed, but it is equally the case that not everything *is* borrowed in a contact situation in which languages maintain separate identities, i.e. when there is balanced contact. In the case of the Balkans—and differentially among the Balkan languages—these differences reflect both geography (west vs. east, south-west vs. north-east) and history. This brings us to some issues of definition. Regardless of the definition of the Balkan peninsula or of what constitutes a language versus a dialect—both issues of burning political importance in south-eastern Europe—there is a general consensus that four Indo-European groups are represented in the Balkan sprachbund: Albanic, Hellenic, Romance, and Slavic. A fifth Indo-European group, Indic (as represented by Romani), occupies a more marginal position, but precisely this marginality renders it particularly important in the examination of contact-induced change. Finally, while Turkic has long figured as a factor in the Balkans, the relevant Turkic dialects have only recently been examined in the context of mutual interaction with the other Balkan languages (see Map 1 for an overview of languages, dialects, and distribution). Of these six groups, we have documentation of forms that we can treat as representative of ancestral languages or at least much older relatives in the case of Hellenic, Indic, Romance, Slavic, and Turkic.

#### 3.2 *The Balkans as a place*

With regard to geography, the east, south, and west sides of the Balkan peninsula are unproblematically defined by water: starting from the north-east and moving around to the north-west we have the Black Sea, the Sea of Marmara, the Aegean, the Mediterranean, the Ionian, and the Adriatic. There is no boundary—be it geographical or political—that unproblematically defines the northern limit of the Balkans. Geopolitically, the accepted definition used to be the northern borders of former Yugoslavia and Romania, but among Balkan linguists it was generally accepted that only the southernmost



MAP 1 Languages of the Balkan sprachbund

Note: Unlike some other language contact situations, the various Balkan languages are overlapping and co-territorial. It is therefore misleading and even inaccurate to assign or limit a specific language to a specific territory, and this is even more the case if the situation of a century ago is taken into account. Thus, each of the seven groups has speakers in each of the Balkan nation-states. While the language of a given eponymous nation-state dominates most of the territory within its borders, there are local areas in all of them where other languages dominate. This is especially true near international borders. The East/West isogloss bundles for Bulgarian, Judezmo, and Turkish are close enough to be represented together. The Aromanian isogloss bundle roughly follows the Macedonian after intersecting with it.

(Torlak) dialects of former Serbo-Croatian (now Bosnian/Croatian/Serbian, henceforth BCS) are characteristically Balkan. A combination of the isoglosses defining Torlak and delimiting Albanian serves to define a northern linguistic limit on the territory of former Yugoslavia that, while not impervious, is relatively clear in terms of the spread of morphosyntactic features. Similarly, the Wallachian and Moldavian dialects of Romanian, having been the longest under Turkish rule, are more typically Balkan than those to the north (see Belyavski-Frank 2003: 235–81). As we shall see (§6.1), Romani also enters the definitional picture.

#### 4 Sociolinguistic factors: religion and prestige

Although religion did not (and does not) correlate in a simple manner with language, it was (and is) a factor in language choice and influence. Most of the Balkan Slavic speakers, Balkan Romance speakers, and Tosk Albanian speakers of the region ended up in the Eastern (Greek) Church after the Great Schism (1054), most Geg Albanian speakers ended up as Catholics (whence the Montegrin expression *Arbanaška vjera* 'Catholicism' literally 'Albanian faith'), although the Geg villages in the Reka region of western Macedonia are Orthodox. Many Bosnian Slavs and Albanians converted to Islam, as did some other Slavs (Gorans, Torbesh, Pomaks), as well as some Greek and some Balkan Romance speakers, all at various times following the Ottoman conquest. The timing and circumstances of these conversions are still matters of historical debate and investigation. By the late middle ages in the Balkans, Turkic-speakers were Muslim except the Gagauz of the Black Sea Coast, who appear to have been descended from Turkish-speaking Byzantine mercenaries who kept their Orthodox faith (Wittek 1951/2).

In terms of social position, if we start from the situation prior to the Roman occupation of the Balkans, Greek was the only surviving language of prestige and literacy that we can decipher (Linear A remains a mystery). Even after the Roman occupation (roughly second century BCE until the Slavic invasions of the sixth and seventh centuries CE), Greek retained its position as a language of high culture, and educated Romans learned Greek, although Latin now also had prestige as the language of empire. The Slavic invasions radically altered the linguistic balance on the peninsula. The evidence of toponymy combined with survivals of Slavic linguistic islands in southernmost Greece into the sixteenth century indicate that the entire peninsula was linguistically inundated. It is presumed that Greek survived only in the coastal cities and towns, while Latin (or Romance) and Common

Albanian were pushed up into the mountains; and that, during this period, Slavic speakers and Romance/Albanian speakers occupied complementary agricultural and pastoral economic niches, respectively. It was also during this period that the other languages of the Balkans presumably disappeared, although they had been under pressure from Latin and Greek for centuries. The Slavic invasions probably also spilt Balkan Romance into Daco-Romanian and Vlah, and possibly Albanian into Geg and Tosk, although the evidence is contradictory.<sup>5</sup> During the late middle ages Albanian extended down into Attica and the Peloponnese, and the region still has remnants of its Albanian (Arvanitika) speaking population. The Fourth Crusade (1204) and the rise of Venice also brought in West Romance influence, but this was limited to lexicon.

During the Ottoman period (fourteenth to early twentieth centuries), Turkish became the language of urban communication and sophistication. Greek was a language of religion (for Christians) and commerce as well as literacy. Slavic was also used in these functions, generally to the north of Greek. Aromanian, Albanian, Romani (arrived tenth–twelfth centuries), and Judezmo (arrived 1492) were home languages, but their distributions varied. All these languages were spoken by settled populations in towns, and except for Judezmo, also in villages—as was also the case for Greek and Turkish. Urban Aromanians generally used Greek outside the home, Jews and Roms used Turkish (and in Salonica and Istanbul, Jews used French), although they would also know whatever contact languages were relevant. Romanian and Albanian remained predominantly rural languages until the nineteenth century, although literacy in both is attested from earlier centuries, and both these languages were spoken in towns in regions where there were significant numbers of speakers. Although language could correlate with status and thus influence marriage choices, the main determiner of eligibility for marriage was religion (except in the case of Roms, where endogamy correlated with ethnicity, i.e. Roms did not generally marry outside their ethnolinguistic group regardless of religion).

#### 5 Balkanisms

In discussing the properties of the Balkan languages that arguably result from language contact (Balkanisms), giving a list runs the risk of oversimplification,

<sup>5</sup> Tosk rhotacism, which is consistently realized in Latin loanwords, is not consistent in Slavic ones. This may be due to non-Tosk (Balkan Romance) intermediaries, the differential spread of rhotacism on Tosk territory, or a post-Slavic date of the Geg–Tosk split.

while space limits full elaboration. For our purposes here, it will suffice to note some of the most salient. For fuller discussion see Friedman and Joseph (forthcoming) and the handbooks listed in Friedman (2000). Although the Balkan languages show convergence at every linguistic level it is the morphosyntactic convergences that first captured European attention, and these are still the most remarkable. The first observed Balkanisms (Kopitar 1829: 106) are among the simplest at first glance and more complex when viewed carefully: the postposed definite article, the replacement of the infinitive with an analytic subjunctive, and the marking of futurity with a verb or particle derived from a verb meaning 'want'. To these Miklosich (1861: 6–8) added genitive-dative merger and doubled object pronouns, while Seliščev (1925) added the use of resumptive clitic pronouns to mark dative and accusative substantival objects, datives as possessives, and the general loss of case forms (among others). (Miklosich also adduced directional/locational mergers such as 'whither/where', the use of 'on ten' to mean '-teen', repetition for quantity, intensity, and/or distributivity, and the presence of schwa as Balkan features, but these are not of interest to us here.) Sandfeld (1930), the first book-length treatment of Balkan linguistics, did not appear to have expanded significantly on Seliščev's list if one simply looks at the section headings of chapter 4 'Concordances générales en dehors de lexique' (163–216), but in fact his third chapter, 'Concordances entre différentes langues balkaniques en dehors de lexique' (100–62), while less comprehensive than later studies, contains the first observations of a number of Balkanisms that have subsequently been expanded upon: future in the past used as conditional (105); interrogatives as relatives (107); preservation of the aorist/perfect opposition (105); zero partitives and accusatives of motion (109–11); perfect and pluperfect in 'have' (132, 149); indefinite article (130); double determination (122); intransitive perfects in 'be' (132, 149); preservation of the vocative (146–7); analytic comparison of adjectives (156, 161, 164); independent (optative) use of the analytic subjunctive (180).

## 6 Balkan futurity and referentiality

Two of Kopitar's 'first' Balkanisms, the 'want' future and the postposed definite article, serve as excellent starting points for illustrating the complexity hidden behind the commonly cited generalizations.

### 6.1 The Balkan future

In the case of the 'want' future, the choice of auxiliary verb is known elsewhere in Europe as is the fact that it reduced to an invariant marker

(which actually is not entirely the case in the Balkans; see Kramer 1994 for details), but the fact that these developments took place in the Balkan languages at approximately the same time in a multilingual geopolitical environment, triumphed over competing models that were successful in neighbouring regions and/or related languages, and can be seen to have spread from Balkan centres of innovation, sets the Balkans apart from adjacent languages.

With regard to futurity, the attested ancestral languages had no morphological category (Slavic) or an inflected morphological one (Sanskrit, Latin, Greek). In the late middle ages in the Balkans, the verbs 'want' and 'have' competed as auxiliaries in Greek, with 'have' being predominant until the twelfth century, and only gradually giving way to 'want' during the late middle ages. The construction reduces to particle + subjunctive marker + finite verb (*the na graphō* 'I will write') in the thirteenth century and to a simple future marker (*tha graphō*) in the sixteenth, which was completely generalized by the eighteenth, except in Italo-Greek, which still has a 'have' future (Sandfeld 1930: 184).

In Latin, too, 'have' and 'want' futures competed, the former first attested in Cicero, the latter in Plautus. The 'have' construction dominated all of Romance except in the Balkans.<sup>6</sup> In Daco-Romanian, an invariant particle *o* plus subjunctive is characteristic of the south (Wallachia), whereas further north there is a conjugating auxiliary or a conjugating 'have' future. In Aromanian the future is marked only with the particle *va* (from 'want'), while in Megleno-Romanian the future marker merged with the subjunctive marker, except in Tsărnareka, which has distinctive *ās* (from *va + sâ*).

In Old Church Slavonic, 'have' and 'want' competed with the perfective of 'be', verbs meaning 'begin', and the plain perfective. Constructions with the perfective of 'be' and plain perfectives became grammaticalized everywhere except Macedonian, Bulgarian, and the Štokavian and Čakavian dialects of BCS, but only in the Torlak dialects did 'want' reduce to an invariant particle. It is worth noting that the 'will' future was carried northward out of the Torlak area into the rest of Štokavian by migrating populations during the Ottoman period.

In Albanian, 'have' and 'want' futures still appear to be in competition in the oldest full texts (sixteenth century). Although superficial accounts generally characterize the 'want' future using invariant *do* as Tosk and a 'have' future using conjugated *kam* plus infinitive (*me + participle*) as Geg, in fact the realities are much more complex (Friedman forthcoming). On the one hand, the *do* future has completely taken over Southern Geg, the Central

<sup>6</sup> Although a 'want' future is reported for Southern Italian, the 'have' future dominates (Čelyševa 2001: 144).

Geg of Upper Reka in north-western Macedonia, the North-Western Geg of Shkrel, the left bank of the River Buna near Shkodër, and the central northern Geg of Puka. It is the predominant type in the North-Eastern Geg dialects of Has and Luma in Albania adjacent to Upper Reka, as well as in the northern Geg in Mirdita, south of Puka. In the foothills to the north-east of Shkodër and in Kelmend, Plav, and Gucî *do* competes with *kam*, and there is still a conjugated 'will' future:

- (1) Jam i lik e duo me dek  
 I.am PC:M ill and want:1sg with die:PARTIC  
 I am ill and will die (Shkurtaj 1975: 55)

The Arbëresh dialects of southern Italy, which separated from the main body of Tosk in the fourteenth and fifteenth centuries, have a generalized 'have' future with invariant *ka* or *ket* (although *do* also occurs), while in Geg territory, the 'will' future continues to spread at the expense of 'have'.

In Romani, there are two main ways of forming the future. In northern and central dialects (outside the Balkans), the so-called 'long form' of the present tense in *-a* is used, while the short form functions as the present, e.g. *ker-a-v* (root-stem vowel-1sg) 'I do'—*kerava* 'I will do'. In the Balkan dialects, the future is almost always formed by means of a particle based on the root *kam* 'want', normally plus the short form, and the long form functions as a simple indicative present, e.g. *kerava* 'I do'—*ka kerav* 'I will do'.<sup>7</sup> The so-called Vlax dialects show a complex distribution. Those that are coterritorial with Torlak BCS or east of that and south of the Danube are like the Balkan dialects. Vlax dialects outside the Balkans behave like northern and central dialects, while Vlax dialects in Romania, Bosnia, and northern Serbia have the two futures in competition (Boretzky and Iglja 2004: 2. 63 and 244). The extent of the future marker *ka* and its cognates in Romani is a marker of Balkan influence, and its complete absence from the northern and central dialects, its absence or attenuation in Northern Vlax dialects, and dominance in the Southern Vlax and Balkan dialects suggest the origin of the form after the diaspora of the later medieval period, which is to say the same Ottoman period when the future was grammaticalized in the rest of the Balkans. (Cf. also Boretzky 1996 on infinitive renewal outside the Balkans.)

Macedonian and Bulgarian, as well as the Romani, Turkish, and Aromanian dialects with which they are in contact, however, use a negative existential/possessive for the ordinary negated future as illustrated in Table 1, a grammaticalized remnant of competing 'have' constructions.

<sup>7</sup> Exceptions are the result of attrition of the type seen in Megleno-Romanian.

TABLE 1. Grammaticalized negative futures using possessive/existentials

Macedonian	nema		da	odime
Bulgarian	njama		da	hodime
Aromanian	noare		s'	neadzim
	<i>not.has</i>		SP	<i>go.1pl:PRES</i>
Romani	na-e	amen	te	dža[s]
	<i>not-is</i>	<i>1pl:ACC</i>	<i>SUBJ</i>	<i>go.1pl:PRES</i>
WRT	yok-tur		gid-elim	
	<i>not-is</i>		<i>go-OPT.1PL</i>	
English	we won't go			

To sum up, the 'will' future in the Balkans is an example of mutual reinforcement and feature selection that began to take shape in the late middle ages but did not reach its current situation until the early modern period, and in some areas, e.g. parts of Albania and Romani dialects, the process is still ongoing. Western Macedonia and adjacent parts of Albania emerge as a centre of innovation. In looking at the grammaticalization of future constructions, we see, on the one hand, that 'will' is still spreading at the expense of 'have', 'have' is not altogether vanquished, and reduction to invariant clitic is not altogether complete, especially in Romania, BCS, and Bulgaria (for the future in the past, where the auxiliary conjugates).

## 6.2 Referentiality

Definiteness and deixis involve reference, as does indicating topicality by means of object reduplication.<sup>8</sup> In the Balkans these features interact in different ways in different regions with object reduplication as the least grammaticalized but most widespread and deictic distinction being the least permeable to contact-induced change.

**6.2.1 The definite article** In the case of the postposed article, the presence of similar phenomena in Scandinavian and in north Russian dialects and the absence of postposing in Greek and Romani (Balkan Turkish marks definiteness but not with an article) do not by themselves vitiate the Balkan contact nature of the phenomenon in Albanian, Balkan Slavic, and Balkan Romance. (See Hamp 1982 on the possible antiquity of this phenomenon in the Balkans.) For both Balkan Romance and Balkan Slavic, we know that the

<sup>8</sup> The use of resumptive clitic pronouns to mark dative and accusative nominal and pronominal objects is traditionally referred to as *reduplication* or *object reduplication* (French *redoublement de l'objet*) in Balkan linguistic literature, and this is the usage we follow here.

development took place after the arrival of speakers in the Balkans. In Balkan Slavic the Timok-Nišava dialects of north-eastern Torlakia have a three-way deictic distinction in the definite article (see §6.2.2) independent of the same phenomenon in western Macedonia, whereas the deictic distinction in Goran postposed articles represents a continuation of western Macedonian. The postposed articles are lacking in all the other Slavic dialects of Kosovo (as well as the adjacent South Morava region of Torlak in southern Serbia proper).

6.2.2 *Deictic pronouns* The postposed definite article can be seen in the context of deixis and topicality, this latter expressed by the use of clitic pronouns to mark certain substantival direct and indirect objects (see §6.2.3). Deixis is more resistant to contact phenomena than definite referentiality, although we do have contact simplifications at the local level (see below). It also appears that certain aspects of deixis and referentiality are salient as emblems of identity (see §7).

Deictic systems in the Balkans are of two kinds: two way and three way. Moreover, Romani combines the usual proximal/distal (-a/o) with general/specific (d/k-) so that every demonstrative must express both oppositions (see Matras 2002: 103–6). In terms of attested ancestral languages, we know that Greek, Latin, Sanskrit, and Old Church Slavonic all had three-way deictic systems, as did earliest attested Turkic. In the case of Romance, the simplification to a two-term system covered the entire area. In Balkan Slavic, most of Bulgarian has a simplified two-way deictic opposition, but unlike Romance and Albanian, it is the distal rather than the proximal that is functionally marked in that the unmarked member is the one normally serving as the third person pronoun.<sup>9</sup> In Lower Gora, however, the two-way system parallels exactly the Albanian, which is presumed to be the source of the simplification (Mladenović 2001: 356). Typical forms are illustrated in Table 2.

6.2.3 *Object reduplication* The use of a clitic pronoun agreeing in gender, number, and case with an accusative or dative object—usually definite—is an oft-cited Balkanism that is grammaticalized to differing degrees in different languages and dialects. Example (2) from Macedonian is typical:

- (2) mu go davam moliv-ot na momče-to  
 3sg:N:DAT 3sg:M:ACC give:1sg pencil-DEF:M to boy-DEF:N  
 I give the pencil to the boy

<sup>9</sup> We are referring here to functional markedness, the unmarked member being the one chosen in neutral contexts. The dialects of the central Rhodopes and the transitional dialects of the extreme west of Bulgaria have three-way deictic systems in the pronouns and the article. The former is a unique isolated development using the older Slavic proximal deictic *s-* rather than the later, originally oppositional deictic, *-v-*, while the latter is a continuation of Torlak in this respect.

TABLE 2. Balkan deictic systems

	Proximal	neutral/[3sg]	distal/[neutral]
Albanian	këtë	atë (ACC)	—
Daco-Romanian	acest (ăst)	acel	—
Megleno-Romanian	tsistu	tsäl	—
Aromanian	aist	atsel	—
Judezmo	esti	akel	—
Greek	(e)toútos	autós	(e)keínos
Lower Gora	ovaj	taj	—
BCS	ovaj	taj	onaj
Macedonian	ovoj	toj	onoj
Bulgarian	—	tozi	onzi
Turkish	bu	[şu]	[o]
Romani	akava—adava		okova—odova <sup>a</sup>

<sup>a</sup> The original proximate/distal opposition marked by *d/k* in the root is obligatorily coupled with a non-specific/specific distinction usually marked by *d/k* in the root yielding a four-terms system in almost all Romani dialects (see Matras 2002: 103–12 for details).

Such reduplication is prescribed for all definite direct and all indirect objects in standard Macedonian, and indeed, this is generally the rule in the western dialects. In the eastern dialects, as in Bulgarian, reduplication is facultative, although there are contexts where, colloquially, it will be expected. Literary Bulgarian is at the opposite extreme, and reduplication is almost completely proscribed in formal contexts, a tendency that also influences standard Greek, Romanian, and Albanian usage, albeit to varying degrees. In Bulgarian, Greek, and Romanian, the exclusion of reduplication from formal contexts derives in part from the fact that the older prestige languages did not have it. Thus, for example, the Modern Greek and Bulgarian translations of the Gospels completely lack reduplication. (Both Albanian and Romanian require reduplication in some contexts.) In all the Balkan languages, object reduplication occurs colloquially, normally when the object is a topic, which is usually the case with definite and indirect objects. Many accounts confuse definiteness with topicality, however, claiming that indefinite objects cannot be reduplicated. As examples (3) and (4) show, however, even non-specific indefinites can trigger reduplication if they are topics. Example (3) is from a non-normalized colloquial Macedonian text and (4) is from a corpus of spoken Bulgarian, thus representing languages at the two extremes of grammaticalization of reduplication:



- (3) Star čoek da go pregrnuvaš vo son [ ... ] boles ke  
 old person SUBJ him:ACC embrace:2sg in sleep [ ... ] illness FUT  
 te fati  
 you:ACC grab:3sg  
 If you dream of embracing an old person [ ... ] you'll get sick  
 (Koneski 1967: 262)
- (4) Banan ne običam da go jam  
 banana not like:1sg subj it:m:acc eat:1sg  
 I don't like to eat banana[s] (Leafgren 2002: 176)

Except for Gora, none of the Slavic dialects of Kosovo have definite articles, but object reduplication can be used with referential topics, as seen in this example from Sretečka Župa, to the north-east of Gora:

- (5) pa de će vidjet dukat — zgazi — dok gi  
 and where FUT see:3pl ducat step.:IMP until them:ACC  
 zbrale sve dukati  
 gathered:3pl all ducats  
 and where[ever] they would see a ducat [they would] step on it until  
 they had gathered all the ducats. (Pavlović 1939: 289)

Such examples are common in southern Kosovo but appear to be less frequent in Timok-Nišava, apparently under the influence of the standard language.

Albanian, Greek, and Macedonian illustrate the three basic types of object reduplication in the Balkans. As already indicated, in Greek, as in Bulgarian, reduplication is almost always facultative, although it will be normal in some contexts. In Macedonian, it is fully grammaticalized, while in Albanian it is less grammaticalized, being obligatory under more limited conditions than in Macedonian. Examples (6) and (7) are illustrative

- (6) (a) Pap-a Ø vizitoi madje Tiranë-n (Albanian)  
 Pope-DEF Ø visited even Tirana-DEF:ACC
- (b) O Papas Ø episkeftike akoma ke ta Tirana (Greek)  
 the Pope Ø visited even and the:ACC Tirana
- (c) Papa-ta go poseti duri i Tirana (Macedonian)  
 Pope-DEF it visited even and Tirana  
 The Pope visited even Tirana—Tirana is focus  
 (after Kallulli 1999: 32)

- (7) (a) Madje Pap-a e vizitoi Tiranë-n (Albanian)  
 even Pope-DEF it:ACC visited Tirana-DEF:ACC
- (b) Akoma ke o Papaps [ta] episkeftike ta  
 even and the Pope [it:ACC] visited the:ACC  
 Tirana (Greek)  
 Tirana
- (c) Duri i Papa-ta go poseti Tirana (Macedonian)  
 even and Pope-DEF it visited Tirana  
 Even the Pope visited Tirana—Pope is focus, Tirana is topic  
 (after Kallulli 1999: 34)

In (6b) Tirana is the focus and Albanian does not permit reduplication, whereas in (7b) Tirana is the topic and reduplication is required. When substantives are topicalized coming before the verb, reduplication is required in Romanian as well as Albanian. In Macedonian, Tirana has definite reference and therefore requires reduplication, whereas in Greek Tirana is not reduplicated if it is the focus and is only facultatively reduplicated when it is the topic. The northern dialects of Aromanian reduplicate as in Macedonian, whereas the southern dialects reduplicate as in Greek. Judezmo in this respect behaves like Spanish, while Romani also behaves like Greek, i.e. reduplication occurs but is never required.<sup>10</sup>

While object reduplication in northern Aromanian reflects the obligatoriness of this phenomenon in western Macedonian, there has been mutual reinforcement in the effect of Aromanian patterns in terms of pattern copying in local (e.g., Ohrid) Macedonian, as seen in example (8):

- (8) (a) lu vidzui pi/al Marko (Aromanian)  
 him:ACC:M see:AOR:1sg to M.  
 I saw Marko
- (b) go vidov na Marko (Macedonian)  
 him:ACC:M see:AOR:1sg to M.  
 I saw Marko

The use of a directive or dative marker with an accusative object is characteristic of Balkan (and also some Western) Romance and is clearly a

<sup>10</sup> Romani does require reduplication in possessive constructions, but this is a different sort of phenomenon. Bulgarian requires reduplication with negative existentials and Greek requires it when the direct object is *ola* 'everything'.

borrowing in local western Macedonian. Moreover, it appears to be limited to animate or personal nouns, a Romance feature.

## 7 Conclusions

The current dialectal situation around Gora can be seen as the result of centuries of what we can characterize as *isolated contact*, i.e. a population of relatively mobile men (shepherds and migrant workers [*pečalbari*]) and sedentary women whose movements were determined by ties of kinship and marriage. Under such conditions, the dialects of Gora and neighbouring dialects in Kosovo and Macedonia illustrate an epidemiological model of feature spread (Enfield 2003: 366–9) or an ecological model of feature selection (Mufwene 2005: 106–35). The differential spread of the postposed definite article and the reduplication of non-pronominal objects indicates that they mark different allegiances as well as different types of narrative strategy. In Serbia, the definite article of Macedonian is perceived as one of its most distinctive features, and indeed the folk stereotype of Macedonian in Serbia is that it sounds like *ta-ta-ta, to-to-to* (repetitions of the feminine and neuter definite articles). This can be attributed in part, at least, to the fact that the majority of BCS dialects rely on case endings for decoding syntactic relations, and definite articles occur where most BCS speakers would expect those inflections. It is arguable, then, that the definite article has had emblematic status in Gora, given that its speakers are Muslims with ties to the Slavic-speaking Muslims to their south (cf. Storch, Chapter 3, and Enfield 2003: 267–8 on emblematic formal and structural features in a linguistic community). Reduplication, on the other hand, is more widespread but less grammaticalized, and appears to have become negatively valued in the south-east Serbian periphery during the course of the twentieth century. This same negative valuation occurs in the Albanian dialects of Kosovo, where the feature is more frequent but associated with Slavic.

Romani has been completely open to Balkan contact in future formation and the influence of Greek in using native material for a definite article, but more closed with regard to object reduplication and entirely closed in the maintenance of its deictic system. In general in the Balkans, modalities such as the future and the conditional are more subject to spread and grammaticalization, while in referentiality there is a gradation from topicalization by object reduplication through the grammaticalization of a definite article to the preservation of deictic systems. In this respect, it is worth noting that while Balkan and non-Balkan Romance followed similar paths, Balkan Romance did so in contact with Albanian, which in this respect influenced

the dialects of Lower Gora. The Bulgarian development seems more consistent with what were once adjacent East Slavic dialects, while Macedonian has had the conservatism of BCS. The parallels in 'will' and 'have' future competition in the languages spoken on Balkan Slavic territory, together with the continued spread of the 'will' future in Albanian, point to central Albania and western Macedonia—areas of the most intense and complex multilingual contact and also areas of intense social contact, with speakers of the various languages being both Christian and Muslim—as a major source of both innovations and conservatisms. To be sure, influences also came up from Greek in the south and travelled east from Bulgaria as well as north from north-eastern Macedonia into Torlakia, where, however, the three-way deictic distinction points to a stronger deictic system.

In this chapter we have seen changes that are structural, diffusional (including multidirectional), differential, and ideological (in the restriction of object reduplication as well as, perhaps, the use of definite articles). I have attempted to look at a larger picture without looking at too large a picture. On the one hand, one must be wary of the classic problem of missing the forest for the trees. On the other, however, it would be a gross oversimplification to say that the Schwarzwald and Kara Orman are merely manifestations of the same 'Black Forest'. They are both mixed, broad-leaf and coniferous, but their histories and current realities are very different.

Owing to their relatively contiguous extension over the Balkan peninsula, Slavic dialects, combined with the extent of Albanian and certain features in Romani, provide a measure of boundary definition that differs from what can be deduced from the other language groups. In this sense, Leake was not entirely mistaken in the position he assigned to Slavic in Balkan linguistic contact, for it is precisely on current South Slavic and adjacent territory that features spread and diminish.

## References

- Adams, D. and Mallory, J. P. 'The Greek language', pp. 240–6 in *Encyclopedia of Indo-European culture*, edited by J. P. Mallory and D. Q. Adams. London: Fitzroy Dearborn.
- Andriotis, A. P. and Kourmoulis, G. 1968. 'Questions de la linguistique balkanique et l'apport de la langue grecque', pp. 21–30 of *Actes du premier congrès international des études balkaniques et sud-est européens*, Vol. 6, edited by I. Gălăbov, V. Georgiev, and J. Zaimov. Sofia: BAN.
- Bakić-Hayden, M. 1995. 'Nesting orientalisms: the case of former Yugoslavia', *Slavic Review* 54: 917–31.

- Belyavski-Frank, M. 2003. *The Balkan conditional in South Slavic*. Munich: Otto Sagner.
- Boretzky, N. 1996. 'The "new infinitive" in Romani', *Journal of the Gypsy Lore Society*, ser. 5, 6: 1–51.
- and Iglá, B. 2004. *Komentierter Dialektatlas des Romani*, 2 vols. Wiesbaden: Harrassowitz.
- Brixhe, C. and Panayotou, A. 1994. 'Le Thrace', pp. 179–204 of *Langues indo-européennes*, edited by F. Bader. Paris: CNRS Editions.
- Čelyševa, I. I. 2001. 'Dialekty Italii', pp. 90–146 of *Romanske jazyki*, edited by I. I. Čelyševa, B. P. Narumov, and O. I. Romanova. Moscow: Academia.
- Enfield, N. J. 2003. *Linguistic epidemiology: semantics and grammar of language contact in Mainland Southeast Asia*. London: Routledge.
- Friedman, V. A. 2000. 'After 170 years of Balkan linguistics: whither the millennium?', *Mediterranean Language Review* 12: 1–15.
- forthcoming. 'Albanian in the Balkan linguistic league: a reconsideration of theoretical implications', in *Proceedings: ninth international congress on southeast European studies*. Tirana: Albanian Academy of Arts and Sciences.
- and Joseph, B. J. Forthcoming. *The Balkan languages*. Cambridge: Cambridge University Press.
- Hamp, E. P. 1982. 'The oldest Albanian syntagma', *Balkansko ezikoznanie* 25: 77–9.
- 1989. 'Yugoslavia—a crossroads of Sprachbünde', *Zeitschrift für Balkanologie* 25, 1: 44–7.
- Haspelmath, Martin. 1998. 'How young is Standard Average European?', *Language Sciences* 20.3: 271–87.
- Joseph, B. D. 2001. 'Is Balkan comparative syntax possible?', pp. 17–43 of *Comparative syntax of Balkan languages*, edited by M. L. Rivero and A. Ralli. Oxford: Oxford University Press.
- Kallulli, D. 1999. The comparative syntax of Albanian: on the contribution of syntactic types to propositional interpretation. Ph.D. dissertation, University of Durham.
- Koneski, Blaže, 1969. *Gramatika na makedonskiot literaturni jazik*. Skopje: Kultura.
- Kopitar, J. 1829. 'Albanische, walachische und bulgarische Sprache', *Jahrbücher der Literatur* 46: 59–106.
- Kramer, C. E. 1994. 'The grammaticalization of the future auxiliary in the Balkan languages', *Indiana Slavic Studies* 7: 127–36.
- Leafgren, J. 2002. *Degrees of explicitness*. Pragmatics and beyond, new series, 102. Amsterdam: Benjamins.
- Leake, W. M. 1814. *Researches in Greece*. London: John Booth.
- Lindstedt, J. 2000. 'Linguistic Balkanization: contact-induced change by mutual reinforcement', pp. 231–46 of *Languages in contact*, edited by D. Gilbers, J. Nerbonne, and J. Schaeken. Studies in Slavic and General Linguistics 28. Amsterdam: Rodopi Press.
- Masica, C. 1976. *South Asia as a linguistic area*. Chicago: University of Chicago Press.

- Masica, C. 2001. 'The definition and significance of linguistic areas', pp. 205–68 of *The yearbook of South Asian languages and linguistics*, edited by P. Bhaskararao and K. Subbarao. New Delhi: Sage.
- Matras, Yaron. 2002. *Romani: a linguistic introduction*. Cambridge: Cambridge University Press.
- Miklosich, F. 1861. 'Die slavischen Elemente im Rumunischen', *Denkschriften der Kaiserlichen Akademie der Wissenschaften, Philosophisch-historische Klasse* 12: 1–70.
- Mladenović, Radivoje. 2001. *Govor Šarplaninske Župe Gora*. Srpski dijalektološki zbornik 48. Belgrade: SANU.
- Mufwene, S. 2005. *Créoles, écologie sociale, évolution linguistique*. Paris: L'Harmattan.
- Pavlović, Milivoj. 1939. *Govor Srečke Župe*. Srpski dijalektološki zbornik 8. Belgrade: Srpska Kraljevska Akademija.
- Reiter, N. 1994. *Grundzüge der Balkanologie: Ein Schritt in die Eurolinguistik*. Wiesbaden: Harrassowitz.
- Sandfeld, K. 1930. *Linguistique balkanique*. Paris: Klincksieck.
- Seliščev, A. 1925. 'Des traits linguistiques communs aux langues balkaniques: un balkanisme ancien en bulgare', *Révue des études slaves* 5: 38–57.
- Schleicher, A. 1850. *Die Sprachen Europas in systematischer Übersicht (Linguistische Untersuchungen von Dr. A. Schleicher II)*. Bonn: König.
- Shkurtaj, Gj. 1975. 'E folmja e Kelmendit', pp. 5–129 of *Dialektologjia shqiptare III*, edited by M. Domi. Tirana: Universiteti i Tiranës.
- Thomason, S. G. and Kaufman, T. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley and Los Angeles: University of California Press.
- Todorova, M. 1997. *Imagining the Balkans*. Oxford: Oxford University Press.
- van der Auwera, J. 1998. 'Revisiting the Balkan and Meso-American linguistic areas', *Language Sciences* 20.3: 259–70.
- Whorf, Benjamin Lee. 1938. 'Some verbal categories of Hopi', *Language* 14: 275–86.
- Wittek, P. 1951/2. 'Les Gagaouzes = les gens de Kaykäu', *Rocznik orientalistyczny* 17: 12–24.