UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

The Origin and Spread of Locative Determiner Omission in the Balkan Linguistic Area

Permalink

Author

Prendergast, Eric Heath

Publication Date

2017

Peer reviewed|Thesis/dissertation

The Origin and Spread of Locative Determiner Omission in the Balkan Linguistic Area

by

Eric Heath Prendergast

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Line Mikkelsen, Chair
Professor Lev Michael
Professor Ronelle Alexander

Spring 2017

The Origin and Spread of Locative Determiner Omission in the Balkan Linguistic Area

Abstract

The Origin and Spread of Locative Determiner Omission in the Balkan Linguistic Area

by

Eric Heath Prendergast

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Line Mikkelsen, Chair

This dissertation analyzes an unusual grammatical pattern that I call locative determiner omission, which is found in several languages belonging to the Slavic, Romance, and Albanian families, but which does not appear to have been directly inherited from any individual genetic ancestor of these languages. Locative determiner omission involves the omission of a definite article in the context of a locative prepositional phrase, and stands out as a feature of the Balkan linguistic area for which there are few, if any crosslinguistic parallels. This investigation of the origin and diachronic spread of locative determiner omission serves the particular goal of revealing how the social context of language contact could have resulted in a pattern of grammatical borrowing without lexical borrowing, yielding a present distribution in which locative determiner omission appears in several Balkan languages no longer in direct contact with one another. A detailed structural and historical analysis of locative determiner omission in Albanian, Romanian, Aromanian, and Macedonian is used as a basis for comparison with other Balkan languages. The analysis pays particular attention to the sequence of grammatical changes necessary for the outcome of locative determiner omission in each language, and the specific sociocultural configurations between speaker communities at relevant historical stages that allowed for the spread of locative determiner omission without direct lexical borrowing. This makes it possible to establish that locative determiner omission arose from a period of early contact between proto-Albanian and Late Latin, resulting in the generalization of the structure across all branches of Balkan Romance. During a later period, contact between Aromanian and individual dialects of Albanian and Macedonian resulted in the transfer of this feature, which then spread throughout most, but not all dialects of these latter languages.

A methodological contribution of this dissertation is the demonstration that in-depth study of a grammatical feature that is suspected to have developed through language contact can yield important insights into the historical and social dynamics of a linguistic area that cannot be determined through synchronic observation of broad similarities alone. Even in the absence of documentation, careful reconstruction of the structural accommodations required for the adoption of a grammatical innovation can reveal new information about the process of language contact. This is particularly true for features that are not uniformly distributed across a linguistic area, as is the case with locative determiner omission in the Balkans. As a consequence, my proposal argues for an approach to linguistic areas that views them as an outcome of localized, layered clusters of convergence.

To my parents, Corinne Heath and Edward Prendergast, with all my love.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my heartfelt gratitude to my dissertation committee, Line Mikkelsen, Lev Michael, and Ronelle Alexander. My advisor, Line Mikkelsen, has provided me with steady guidance and profound insight throughout the circuitous development of this dissertation project. Her unwavering support and her excitement in seeing my ideas develop have been essential to bringing this research to fruition. I also must thank Lev Michael for his extraordinary intellectual precision. His deep understanding of language contact and sociocultural linguistic theory enabled him to ask fundamental questions that led the work forward, and his expert approach to fieldwork gave me the best possible model from which to learn. My immense gratitude also extends to Ronelle Alexander, whose mastery of Slavic and Balkan linguistics and attention to detail was tremendously valuable for establishing rigor in my work.

I would also like to thank other scholars in linguistics and Balkan studies who contributed to my intellectual growth and answered my many, many questions. I would particularly like to acknowledge Victor Friedman of the University of Chicago; Brian Joseph of the Ohio State University; Cristina Kramer of the University of Toronto; Grace Fielder of the University of Arizona; Elena Petroska of Indiana University; Stanislava-Staša Tofoska and Marjan Markoviḱ of Ss. Cyril and Methodius University; Afrodita Totsili of Euro-Balkan University; and Shpëtim Elezi of the University of Prishtina.

This project required long periods of work away from home, and it would not have been possible without the good humor, practical experience, and moral support of trusted friends and colleagues working in the Balkans. In particular, I would like to thank Andrew Dombrowski, Maxim Makartsev, Amanda Greber, Chelsi West, and Max Wahlström, all of whose work has always been and will remain an inspiration to me.

Finally, I would like to express my deepest gratitude to my parents, Edward Prendergast and Corinne Heath, who have shown me so much love and who have taken utter delight in my intellectual endeavors. None of this work would have been possible without them. And to John Jacob Bauer, who sat with me through the long, late nights of writing and revising, right to the end, my love and all my thanks.

Chapter 1

Introduction

This dissertation examines the development and distribution of a grammatical phenomenon in the Balkan linguistic area called locative determiner omission (LDO). LDO involves the omission of a definite article in the context of a locative prepositional phrase. This pattern of omission warrants investigation, in that it is a crosslinguistically unusual grammatical means of expressing definite reference, which is shared among several Balkan languages, but which is not universal to the Balkan linguistic area. The goal of this dissertation is to identify how LDO developed in each language that shows it and to explain its puzzling synchronic distribution by identifying the source of LDO and its path of spread through Balkan speaker communities. An important contribution of this dissertation is a detailed, diachronic, and structural analysis of the synchronic phenomenon of LDO through fieldwork, elicitation, and observation in the four languages in the Balkan area that show LDO. This dissertation makes a coordinated argument for LDO as a grammatical feature developed through multiple instances of language contact in the Balkans over time. I will support this argument with a three-tiered approach. In chapters 2-4, I pursue a structural analysis of the syntax and semantics of LDO in the modern Balkan languages of Albanian, Romanian, Aromanian, and Macedonian. This lays the groundwork for comparison among the languages and through their stages of grammatical change over time. In chapter 5, I analyze the diachronic development of LDO and compare its path of development in the language branches of Albanian, Balkan Romance, and Balkan Slavic. In chapter 6, I discuss the literature on the Balkan linguistic area and theories of language contact in support of my argument for how, when, and why LDO became a grammatical feature in some Balkan languages, but not others. In chapter 7, I conclude by synthesizing the contributions of this three-tiered approach and discuss how this method can be applied to other grammatical features shared through language contact.

The history of contact between and among Balkan languages and the processes that led them to converge in grammatical structure are essential to my analysis of LDO as a phenomenon with a shared origin, rather than an independently developed and coincidental similarity in the Balkan linguistic area. Fundamental to my analysis is the proposition that linguistic areas may be identified by synchronic clusters of grammatical, phonological, or lexical similarity, but individual features must be confirmed as shared through the identification of particular circumstances of contact between speaker communities within the linguistic area. This is to rule out the possibility either of convergent, but independent development of a particular feature, or the possibility of shared inheritance. The Balkan linguistic area is primarily composed of languages from several branches of the Indo-European language family, making the latter potential particularly important to investigate.

Balkan languages belong to the Slavic (Macedonian, Bulgarian, the Torlak dialect region of southeast Serbian), Romance (Romanian, Aromanian, Megleno-Romanian, Istro-Romanian, Judezmo), Hellenic (modern Greek), and Albanian branches. Not all languages spoken on the Balkan peninsula show the shared features that mark a language as part of the Balkan linguistic area. Languages belonging to this area, set against a map of the region as a whole, are shown in Figure 1.1. This depiction involves some necessary simplifications, given that the Balkan linguistic area must be viewed as composed of local speech forms in close contact with one another. Some languages, such as Romanian, show extensive common Balkan features in all dialects. Other languages, such as Serbian, lack Balkan features in the standard form, but include dialects that must be considered part of the Balkan linguistic area. In addition, speaker populations of individual Balkan languages may extend over modern nation-state boundaries, or may be located at a significant remove from the core zone of contact.[1] This complexity demands caution in making generalizations about whether a language as a whole does or does not include a feature. Moreover, it also problematizes the notion of broad contact among 'languages,' rather than contact among local speech varieties.

Though Balkan languages are remarkable for the extent of their similarities, they do not all show LDO. In Figure 1.2, only those languages containing dialects that feature LDO are depicted.

This distribution is interesting in that it does not seem to reflect either genetic inheritance or areal spread. If one examines the phenomenon by comparing languages within the Balkan linguistic area with their relatives outside of it, the distribution of LDO among languages of the Albanian, Balkan Slavic, and Balkan Romance branches suggests that this feature must have developed through language contact. LDO does not appear in Slavic or Romance languages outside of the Balkan linguistic area. The feature has few parallels cross-linguistically. Its attestation in four members of a known linguistic area strongly suggests that it is an areal innovation. But if LDO is an areal feature, it did not spread by a process of diffusion from a central point. This would be the expected pattern of spread for an innovation, from center to periphery over time. In contrast, the distribution of LDO is uneven throughout the Balkan linguistic area. Not all languages with LDO are geographically contiguous to one another. While Aromanian, Albanian, and Macedonian share territory and are spoken in zones of multilingual interface, they are collectively separated from Romanian, which features LDO throughout. Thus the mechanism by which LDO could have emerged as an areal feature is not immediately clear.

The goal of this dissertation is to explain this puzzling synchronic distribution by identifying the

---

[1]This map is only intended to provide a rough approximation of the situation of Balkan languages. Balkan languages are overlapping, co-territorial, and have changed in their range and distribution within the past century. It is misleading, and even inaccurate, to assign a language to a particular national territory, but it is also beyond the scope of this depiction to detail all cases of territorial co-existence (Friedman 2007:205). In some cases, this map depicts languages that have been standardized and employed as part of a nation-state. Romanian, Greek, and Bulgarian are depicted here as confined to national borders, but in fact they may be spoken in pockets beyond the boundaries of the modern states. The distinction between language and dialect does not conform neatly to the borders of the Balkan linguistic area. Thus, although the standard Serbian language is not considered part of the Balkan linguistic area, its southeastern dialects, referred to as Torlak, do show Balkan features and must be considered as part of the Balkan linguistic area. Torlak, Macedonian, and Bulgarian exist in a dialect continuum and are thus depicted with similar colors. In addition, this map omits some Balkan languages with highly localized distribution, such as Romani, Judezmo, Balkan Turkish, and Megleno-Romanian. Determining the extent of a Balkan language speaker population and the standardized language to which a dialect should belong is contentious and often politically charged (e.g. the classification of the Goran dialect as either Serbian or Bulgarian, Friedman 2002-3:164). For more detailed maps that depict local situations for individual languages, see e.g. Friedman (1993:301) for Macedonian, Browne (1993:383) for Torlak, Gjinari and Shkurtaj (2003:118-119) for Albanian, and Schwandner-Sievers (1999:iii) for Aromanian.

Figure 1.1: Map of the Balkan languages sharing grammatical features

diachronic source of LDO and proposing a path of spread. This requires comparing languages of the Balkans to one another, and to their relatives outside the Balkan linguistic area. In some cases, tracing the spread of LDO requires the reconstruction of periods of contact for which documentation is not available by coordinating and comparing with other shared lexical or grammatical features. The attestation of Aromanian, Macedonian, Albanian, and Romanian before the 19th century is sporadic, with longs gaps between the documentation of known ancestors and documentation of the modern languages. Detailed accounts of population movements that preserve data about the language spoken by a given population are also sparse. It is clear that migrations of speaker communities played an important role in the spread of LDO, but the available historical, social, and linguistic information must be synthesized to reach this conclusion. My analysis depends on the use of historical documentation situated in the previous linguistic literature on diachronic interactions within the Balkan linguistic area. This analysis also depends on detailed structural analysis of the synchronic phenomenon of LDO based on data gathered through fieldwork, elicitation, and

Figure 1.2: Map of the Balkan languages that show locative determiner omission

observation in Aromanian, Macedonian, Albanian, and Romanian. An important contribution of this dissertation is to go beyond the mere observation of a grammatical similarity between these languages by gathering together this information to make a coordinated argument for LDO as a grammatical feature developed through multiple instances of language contact in the Balkans over time.

Hitherto, a systematic comparison of LDO among speech varieties of Albanian, Balkan Romance, and Balkan Slavic has not been attempted in the literature. Research into the phenomenon has been undertaken in individual languages, often in concert with attempts to account for grammatical changes resulting in the collapse of inherited morphological case distinctions from Indo-European. For Albanian, work on the development and the loss of the locative case was an important basis for the research I present here, since the conditioning factors that once governed the locative and its eventual loss play a vital role in the introduction of LDO into Albanian (Ajeti 1985, Demiraj 1988, 1993, 2002, Riza 2002, Gjinari & Shkurtaj 2003, Karagjozi 2005, Topalli 2009,

2011, Elezi 2010, Mulaku 2012). Boretzky (1968) provides a particularly thorough analysis of the expression of definiteness in Albanian prepositional phrases. In Macedonian, discussion of determiner omission in prepositional phrases occurs primarily in the literature on contact between Aromanian and Macedonian, particularly by Gołąb (1959, 1984, 1997) and Markoviḱ (2007, 2011, 2012, 2013). Michov (1904) first observed the parallels in the omission of the definite article between Balkan Romance, Albanian, and Balkan Slavic, a phenomenon also discussed by Schaller (1975) as a 'secondary Balkanism' and by Joseph (1999) as one of several parallels between Albanian and Romanian. Reichenkron (1962) also observes grammatical properties of prepositions in early Romanian that show the deep grammatical integration of LDO into early Balkan Romance. My dissertation draws on this work and introduces new data in order to synthesize findings from all of the languages in which LDO appears into a single analysis of its origin and spread.

I propose that LDO first emerged from an early period of interaction between proto-Albanian and Balkan Late Latin, from roughly the 2nd century BCE to the 5th century CE (Mallory and Adams 1997:9). During this period, proto-Albanian was heavily influenced by Latin. Systematic and deeply integrated lexical borrowing from Latin into Albanian attests to a long-term period of multilingualism between proto-Albanian and Balkan Latin. Balkan Romance appears to be the outcome of Albanian speakers shifting into Latin and carrying important elements of Albanian grammatical structure into the resulting speech form. The Albanian locative case, with its specific prepositional distribution and semantic restrictions, is a plausible model for the introduction of LDO in Balkan Romance. Since LDO can be found in all variants of Balkan Romance, even varieties at a great distance from the main body of Romanian speakers, one can surmise that LDO must have developed in Balkan Romance before its separation into variants such as Romanian and Aromanian, at some point during the undocumented period. By the time of the first attestation of Balkan Romance in the 16th century, LDO is already fully integrated into the grammatical system of this branch of Balkan languages.

The development of LDO in Albanian and Macedonian must be dated to a later period of contact, after these languages had already undergone extensive convergence with Aromanian over a sustained period of multilingualism. In contrast to Balkan Romance, the first attestation of Albanian in the 16th century still features a locative case, rather than the LDO structure that becomes generalized throughout Albanian by the 20th century. Simplification and the development of syncretisms in the Albanian case system between the 16th and 19th centuries provide the preconditions for the adoption of LDO into the Albanian grammatical system. While the loss of the Albanian locative can be accounted for by internal phonological developments producing parallel changes elsewhere in the Albanian case system, the replacement of the locative with LDO cannot be explained by these same Albanian-internal changes. Fully integrated LDO is first documented in Albanian in the 19th century, among dialects in Albanian communities that coexisted with extensive communities of Aromanian speakers. Aromanian multilingualism and assimilation into Albanian led to the borrowing of LDO into Albanian grammatical structure. This process is unattested for northern Geg dialects; rather, documentation for the process of Aromanian influence on Albanian LDO comes from southern Tosk dialects, where LDO emerges following Aromanian assimilation. The locative case is last to be lost and LDO is last adopted in Albanian dialects spoken in areas where the Aromanian community remained distinct, into the early 20th century. By the time of the mid-20th century, LDO was systematic across the main body of Albanian dialects depicted in Figure 1.2. In contrast, LDO never developed in Albanian dialects separated from the main body of the Albanian speaking community; these dialects did not sustain longterm contact with Balkan Romance

or undergo assimilation of Balkan Romance speakers. Thus the development of LDO in Albanian must be viewed as an outcome of contact with Balkan Romance and the shift of Balkan Romance speakers into Albanian, although inter-Albanian contact is the best explanation for the spread of LDO throughout the contiguous body of Albanian speakers.

Slavic languages do not arrive in the Balkans until the Slavic migration of the late 6th century. Only after a long period of multilingualism with Balkan Romance and Albanian from the 12th century to the 15th century does a distinct form of Balkan Slavic develop that shows common innovations with other Balkan languages. This distinguished Macedonian and Bulgarian from non-Balkan South Slavic languages, such as standard Serbian, Bosnian, or Croatian. Among the innovations common to Balkan Slavic are the structural preconditions for the development of LDO such as the collapse of goal/location semantics, loss of synthetic nominal case, and the post-posed definite article. However, only Macedonian, the variety of Balkan Slavic in close contact with Aromanian, shows the subsequent development of LDO in the 19th and 20th centuries. Where multilingualism with Aromanian and assimilation of Aromanian speakers into the Slavic speech community was not a major factor, LDO does not develop. This accounts for the lack of LDO in Bulgarian.

LDO is thus the result of the same processes of language contact that led to the development of other shared features of the Balkan linguistic area. It has been argued that these commonalities arose from a sustained period of contact between speaker communities. The particular form of this contact was mutual multilingualism; these speaker communities were multilingual in one another's languages, and the sustained dynamic of multilingualism led to a convergence toward comparable grammatical and phraseological structures. The practice of regularly switching between languages in mutual accommodation reinforced changes toward common structures that were easier to map from one language to another, thus easing intercommunal language learning, understanding, and codeswitching. LDO appears to be yet another of these contact-induced innovations of the Balkan languages, an innovation that easily maps in terms of surface structure between those languages that show it due to the grammatical structures they already shared.

But it is not enough to state that LDO arose from contact among Balkan languages. LDO is not shared by all languages of the Balkans, and so a theory of language contact in which the existence of a shared multilingual speaker community is enough to cause general convergence across all languages of the area is insufficient to account for the specific distribution of LDO. It also does not account for the development of this particular grammatical innovation, considering that it is a cross-linguistically unusual pattern for expressing definite reference. The concept of the Balkan linguistic area is a necessary component of the explanation for the synchronic distribution of LDO, but it is not a sufficient one. This dissertation presents a more detailed approach that takes into account diachrony, speech community relations, and the particular mechanisms of contact. I argue for contact at particular stages of grammatical change, among specific communities, in my pursuit for an explanation of the synchronic distribution of LDO.

Rather than taking linguistic areas to be uniform zones of diffusion, my proposal views the development of linguistic areas as a process of localized, layered clusters of convergence. A linguistic area can be seen as a 'cluster of clusters' (Hamp 1989:44), and localized diffusion can be seen as a spectrum of differential bindings (Hamp 1989:47; Joseph 2010:628-29). Iterated individual interactions between individual language communities at the local level lead to mutual accommodation, which prompts feature convergence. But the features that undergo convergence depend on the local speech forms and their specific situation of contact. Some configurations of contact may be
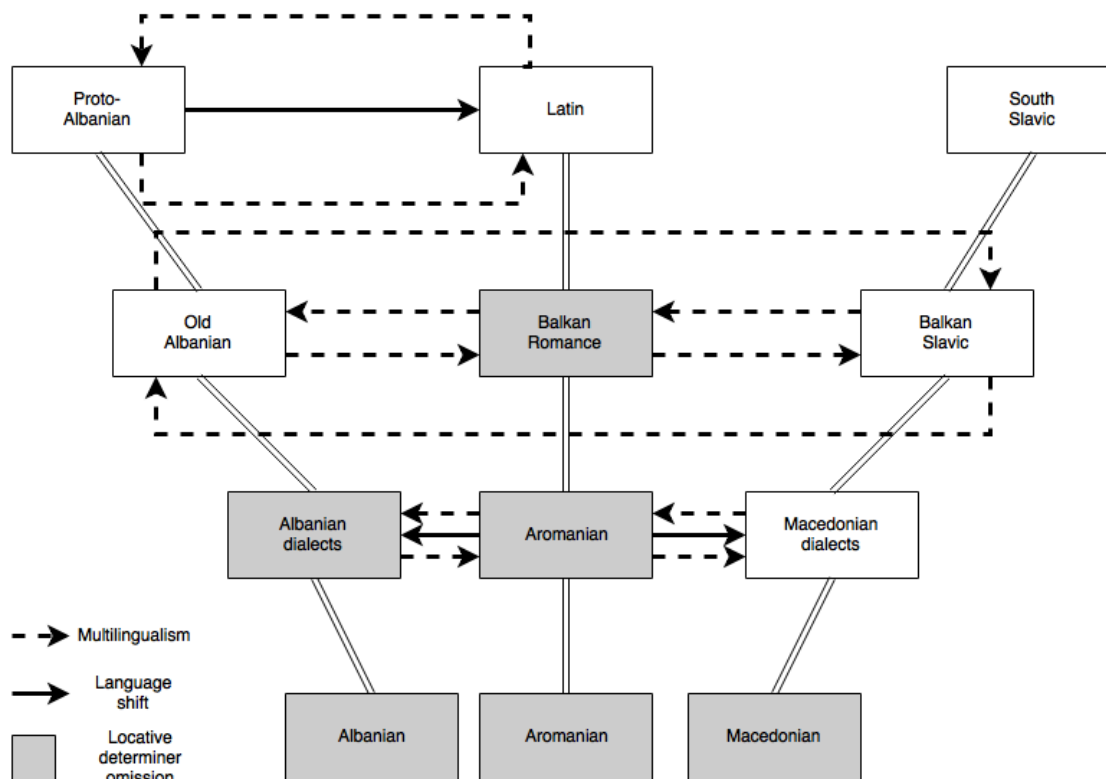
Figure 1.3: Diagram of the path of transmission for locative determiner omission

repeated so often that they give the appearance of a widespread convergence toward a single feature, across the whole area, as with the loss of the infinitive in the Balkans. This is especially common if the innovation allows greater transparency in the mapping of meanings to morphemes, which eases the process of codeswitching and favors acquisition by non-native speakers. Some innovations, however, require a pre-existing cluster of similarity in order to spread, depending on the greater density of intercommunal contact and accommodation. This, I argue, is the case for LDO, which spread through sequential, local contacts between individual speech communities, and required convergence in other preconditioning features before LDO itself could spread.

This detailed study of a single shared morphosyntactic feature is intended to solidly establish LDO as one of the features of the Balkan language union. But it is also meant to illustrate how the development of a shared feature requires more than just a drift in the direction of convergence. My proposal for the path of development and transmission of LDO is presented in the diagram in Figure 1.3. As illustrated in the diagram, multiple periods of multilingual contact and language shift were necessary to yield the final synchronic outcome of LDO.

LDO did not just 'seep in' to Albanian and Macedonian from Aromanian over long-term multilingual contact. Rather, periods of borrowing overlap with one another over time, depending on changes in social relations. Each of these different stages was necessary for the changes that occurred in the next stage, ultimately yielding LDO. In the first stage, illustrated in the figure, a period of shift among Albanian speakers into Late Latin lay the base for LDO in Balkan Romance, in that Balkan Romance developed a post-posed definite article, a goal/location merger in prepositional semantics, and a grammatical condition for the special treatment of definiteness in locative prepo-

sitional phrases. In the next stage, during a later period, mutual multilingualism among Balkan Romance, Albanian, and Balkan Slavic drove the development of key grammatical commonalities that would allow LDO to easily map across similar surface syntactic structures. Subsequently, a shift of Aromanian speakers into Albanian and Macedonian was necessary to spread the specific configuration of definiteness marking coded by LDO. This spread of LDO depended on other contact-induced changes like the loss of case and the introduction of a definite article, which were necessary to enable the mappability of morphemic structures between these languages in contact. In addition, the surface-level expression of definiteness with a single, transparent morpheme and the discourse-oriented nature of definiteness marking allowed the structure to be easily transferred from Balkan Romance into Albanian and Macedonian. Other Balkan languages for which structural equivalence was not possible due to insufficient convergence toward a single grammatical form (Turkish, Romani, and Greek), or where multilingualism with Balkan Romance did not last long enough and widespread language shift of Aromanian speakers did not take place (Bulgarian) do not show LDO. This yields the final stage in the figure, where LDO exists specifically in Albanian, Macedonian, and Balkan Romance languages (as represented by Aromanian), but not in other speech forms of Balkan linguistic area.

This dissertation makes a contribution to the study of contact linguistics by illustrating how and why a grammatical structure would be borrowed. Levels of relatively equal intercommunal prestige promote multilingualism and put pressure on the morphosyntactic structure of Balkan Slavic, Balkan Romance, and Albanian toward variants in which the nature of the relation between meaning and morph is relatively one-to-one. This mapping relation makes codeswitching easier and also makes it easier to carry surface-oriented structures from a source language into a target language. A linguistic area such as the Balkan linguistic area is constituted of many iterations of this process, with feature spread occurring in local clusters of contact that overlap into a broader area to such an extent that it is hard to determine which features are core and which are peripheral. Only by unpiecing individual interactions between languages and situating them in the diachronic circumstances of speaker community interaction is it possible to determine how a feature came to be shared in a linguistic area. LDO is a particularly good illustration of this process, because it is the result of many clusterings of language contact over time and space. This resulted in an uneven synchronic distribution of the feature, but remarkable similarity among the languages that have it.

I will begin by presenting synchronic data on LDO from languages of the Balkans. In chapter 2, I introduce the data from Albanian, whose grammatical properties allow the most vivid demonstration of LDO and its triggering conditions. In chapter 3, I present comparative data from Balkan Romance and show that Balkan Romance languages can be compared with this structure in Albanian. In chapter 4, I discuss LDO in Balkan Slavic and show that it exists in Macedonian, but not Bulgarian, and that its distribution in Macedonian is particularly weighted toward southwestern dialects that underwent the most intensive contact-induced grammatical change resulting from mutual multilingualism and eventual language shift of Aromanian speakers into Macedonian. In chapter 5, I then look at LDO from a diachronic perspective, beginning with its earliest documentation in Balkan Romance, and going on to compare it with documentation in Albanian and Macedonian. In chapter 6, having presented the data, I then review the literature and discuss the situation of contact in which these languages have existed. I use this review to support my argument of how LDO arose through contact, how it spread through the Balkan languages, and why this particular structure was adopted. Finally, in chapter 7, I analyze the development and spread of LDO as an areal feature in the context of research in contact linguistics and the Balkan linguistic area, and in chapter 8, I

conclude with a summary of my arguments.

Chapter 2

Locative Determiner Omission in Albanian

From among the languages of the Balkans that show a pattern of determiner omission in prepositional phrases, Albanian serves as the best example for a detailed illustration. This is because determiner omission is thoroughly integrated into the Albanian grammatical system, across all contiguously spoken dialects. Definite determiners, whose absence is the hallmark of locative determiner omission, are easy to identify in Albanian, and some noun phrase modifiers show further agreement for definiteness. The interaction of definite determiners with noun morphology is predictable and their presence shows a stable relation with definite reference in the noun. Moreover, the distribution of definite determiners in syntactic positions other than prepositional phrases is consistent. The distribution of definite determiners within prepositional phrases however, is exceptional. In particular, certain configurations of case selection, prepositional semantics, and noun phrase syntax require the *absence* of a definite article to indicate definite reference. These restricted syntactic and semantic domains do not result in a neutralization of the grammatical distinction between definite and indefinite reference; rather, zero-marking on the noun indicates that the noun is to be interpreted as definite. This system for marking definite reference diverges strikingly from the consistent pattern found in all other syntactic domains in Albanian. Thus, the detailed analysis of the morphology, syntax, and semantics of the Albanian noun phrase in this chapter establishes the phenomenon of LDO and provides points of comparison for other Balkan languages when examining LDO as a shared feature of the Balkan linguistic area.

The path of development of LDO in Albanian is also the most complicated in the Balkans, resulting from the transformation of an original nominal locative case into a new structure involving the regular omission of an overt definite article. A thorough account of locative determiner omission in synchronic varieties of Albanian allows broader conclusions to be drawn about the development of this phenomenon and its spread through the Balkan languages. Thus, although Albanian is not necessarily the source of the grammatical pattern of LDO, it is the best language with which to begin the analysis. The level of detail in this analysis of Albanian allows for a basis of reference when analyzing LDO in Balkan Romance or Balkan Slavic. To the extent that LDO features repeat themselves in Balkan Romance or Balkan Slavic, I will illustrate them in less detail and instead refer to the analysis presented in this section.

In §2.1, I lay out the grammatical structure of the noun and determiner phrase in Albanian. The interaction of nominal morphological features, definite articles, indefinite articles, and noun phrase syntax all serve as an important background for understanding how definite reference is marked in prepositional phrases. I begin with an analysis of Albanian case morphology, showing how it

interacts with definite articles and noun phrase syntax to yield a standard pattern for expressing a definite referent, whether specific or generic. This three-part analysis begins with §2.1, where I focus on case, gender, and number in Albanian nouns and show how these features are coded in the morphology of definite articles. Next, in §2.1, I introduce the grammar of modifiers in the noun phrase such as adjectives or possessive pronouns. The presence of nominal modifying expressions is an important factor determining the distribution of indefinite and definite articles. In §2.1, I conclude the analysis of the standard grammatical marking of definiteness by analyzing the syntactic distribution of definite articles, indefinite articles, and bare nouns.

From this morphosyntactic analysis, I turn to the semantic features of definiteness, indefiniteness, and specificity in Albanian as they relate to overt morphological markers and determiners. In §2.2 I compare the distribution of bare nouns to that of nouns with definite articles, and in §2.2 I contrast this with the distribution of the indefinite article. In §2.2, I draw conclusions about the denotational structure of Albanian nouns. Thus §2.1 and §2.2 together show how definite reference is expected to be coded by Albanian grammar.

In §2.3, having laid out the syntactic and semantic structure of noun phrases and determiners, I introduce the exceptional behavior of prepositional phrases. As detailed in §2.3, the case that a preposition selects is an important factor determining whether a definite article appears in a noun phrase, with accusative prepositions in particular selecting for the omission of definite determiners. This omission is interpreted as a zero marking for definite reference in the noun. In §2.3, I show that this is a grammatically conditioned phenomenon, subject neither to lexical restrictions nor morphological properties of the noun. Rather, the determining condition for LDO is whether the noun takes any syntactic modifiers: unmodified nouns must omit the definite determiner, modified nouns require it. In §2.3, I also introduce a locative semantic condition on the prepositions, which license a definite interpretation for unmodified nouns only when they code location or movement toward a goal. The sum of these case-selection, syntactic, and semantic conditioning factors together with determiner omission are what I label as locative determiner omission (LDO). Identifying the exact conditions under which determiners are omitted is important for establishing that this feature is shared among several Balkan languages, including Albanian.

2.1 The noun phrase and determiner phrase in Albanian: Case, morphology, and syntax

Definite reference in Albanian is indicated by definite articles post-posed to the first declinable constituent of the noun phrase. Because of their syntactic position, definite articles interact in complex ways with nominal case and the syntax of nominal modifiers. Thus, the Albanian case system involves extensive syncretism and morphological interactions between nominal case morphology and article declension. When case, gender, and number morphology allow, the Albanian definite article is always found in second position of a noun phrase, regardless of the presence of modifiers or the internal syntax of the noun phrase.

The expression of case in Albanian nouns and determiner phrases

The constituent structure of Albanian determiner phrases is important for understanding the expression of definiteness in the language. Data presented in this section was derived from fieldwork,

unless otherwise noted. Albanian determiner phrases are composed of determiners designating definite or indefinite reference, nouns, and optional noun modifiers. Nouns are distinguished by gender, number, and case. Nouns can be either masculine or feminine,[1] with nominal gender mostly predictable from the ending of the noun in the singular. Nouns ending in vowels in the singular are generally feminine, nouns ending in a consonant in the singular are generally masculine. Nouns have singular and plural stems, which are frequently irregular. Some plural stems are formed with the addition of a vowel, as in *qytet* 'city' > *qytete* 'cities.' Other nouns derive their plural stems through umlaut, as in *dash* 'ram' > *desh* 'rams,' or suppletion, as in *djalë* 'boy' > *djem* 'boys.' There is also a set of nouns with a zero plural, such that the singular and plural stems are identical, as in *mësues* 'teacher' > *mësues* 'teachers.' Albanian nouns decline for nominative, accusative, dative/genitive,[2] and ablative case, with great syncretism between the case forms. Case endings are distinct for masculine (Table 2.1) and feminine (Table 2.2) and attach to the singular or plural stem.

Table 2.1: Albanian masculine nominal declension for *qytet* 'city'

	Sg. m.	Pl. m.
Nominative	qytet	qytete
Accusative	qytet	qytete
Dative	qyteti	qyteteve
Genitive	i qyteti	i qyteteve
Ablative	qyteti	qytetesh

Table 2.2: Albanian feminine nominal declension for *vajzë* 'girl'

	Sg. f.	Pl. f.
Nominative	vajzë	vajza
Accusative	vajzë	vajza
Dative	vajze	vajzave
Genitive	i vajze	i vajzave
Ablative	vajze	vajzash

Albanian distinguishes definite referents from indefinite referents through the use of articles. Nouns marked as indefinite in Albanian indicate that the entity to which the noun refers is as of yet unknown to the listener. The referent may be known to the speaker, or the referent may be non-specific or generic. In any case, an indefinite referent is indicated with an indefinite article that accompanies the noun and marks the referent as unidentified (Newmark et al. 1982:150). The

---

[1]There is a residual neuter that behaves like the masculine in the indefinite and the feminine in the definite (Mann 1977:103). It has mostly disappeared from the modern language.

[2]These cases are morphologically the same. They are distinguished by phrasal properties, the syntactic positions in which they appear, and the kind of agreement they trigger in other constituents. Genitive-case nouns must take a particle of concord.

indefinite article is an invariant *një* for the singular (homophonous with *një* 'one') and *disa* for the plural. The indefinite article precedes the noun, as in (1a-d) (Demiraj 1993:122).[3]

(1) a. një vajzë
         a   girl.F
         'a girl'

     b. një qytet
         a   city.M
         'a city'

     c. disa   vajza
         some girls.F
         'some girls'

     d. disa   qytete
         some cities.F
         'some cities'

While the indefinite article always occurs first in the noun phrase, the definite article is positioned internally to the noun phrase, as a suffix rather than an independent word. The definite article must be post-posed to the first declinable constituent in the noun phrase. This is generally the noun, as shown in Tables 2.3-2.4. Like nouns, the definite article declines for gender, number, and case. In the singular, the forms of the definite article for masculine nouns (Table 2.3) are markedly different from the forms of the definite article for feminine nouns (Table 2.4). The definite article distinguishes gender in the singular, but not the plural. There is some variability in the form of the definite article, which is predictable from the phonological structure of the noun stem. There is no overt marker for definiteness in the dative/genitive plural.[4]

---

[3]Some masculine nouns, such as *qytet* 'city,' become feminine in the plural.

[4]At one time, the definite article for the plural was *-t* for all cases, including the dative/genitive and the ablative, but the definite article for these cases is elided in the modern language. Thus the dative/genitive plural fails to distinguish definite from indefinite referents, except by the use of an indefinite plural article. The ablative still distinguishes definiteness from indefiniteness by means of a distinct indefinite ablative case form (see Tables 2.1-2.2).

Table 2.3: Albanian masculine definite nominal declension for *qytet* 'city'

	Sg. m. definite	Pl. m. definite
Nominative	qytet-i	qytete-t
Accusative	qytet-in	qytete-t
Dative	qytet-it	qyteteve
Genitive	i qytet-it	i qyteteve
Ablative	qytet-it	qyteteve

Table 2.4: Albanian feminine definite nominal declension for *vajzë* 'girl'

	Sg. f. definite	Pl. f. definite
Nominative	vajz-a	vajza-t
Accusative	vajz-ën	vajza-t
Dative	vajz-ës	vajzave
Genitive	i vajz-ës	i vajzave
Ablative	vajz-ës	vajzave

It is worth emphasizing that the lack of a definite determiner for dative/genitive and ablative plural nouns is a morphological property. Regardless of the syntactic position in which a dative/genitive or ablative plural noun appears, it cannot take a definite article. The only permitted article for these particular combinations of number and case is the indefinite article. Thus the morphology of the noun itself is a determining condition for whether a definite article can appear in the noun phrase. While noun morphology does not distinguish definite from indefinite in the dative/genitive plural, indefinite nouns do show distinctive morphology from definite nouns in the ablative plural.

In general, the core argument cases of nominative and accusative are only distinguished by the morphology of the definite article; thus, the noun itself does not distinguish nominative from accusative case. In contrast, peripheral cases such as the dative/genitive, and ablative show distinctive morphology both in the noun and the definite article. Case and number in the noun affect the expression of definiteness: dative/genitive and ablative plural, in particular, do not allow an article for nouns with definite reference. Instead, definite is distinguished from indefinite for plural dative/genitive nouns according to whether an indefinite article is present or absent. For the ablative in particular, plural nouns with indefinite reference have a distinctive case form to differentiate them from plural nouns with definite reference, thus *vajzash* 'girls (abl. indefinite)' versus *vajzave* 'girls (abl. definite).' The ablative is otherwise indistinguishable from the dative/genitive in nominal case morphology. The Albanian system of case expression is determined by a complex interaction of gender, number, and definiteness that leads to numerous instances of syncretism and overlap in case morphology. This set of interactions also affects modifiers of the noun that make up the determiner phrase as a whole.

Case and definiteness agreement in noun phrase modifiers

The syntax of modifiers in the Albanian noun phrase involves a relation of agreement between the head noun and modifiers, including articles that accompany modifiers. This relationship is presented in detail in order to support later arguments regarding the development of case and definiteness markers in Albanian from an older system, in which a specific definite locative article occurred within prepositional phrases, to a newer system, in which the locative case was lost, to be replaced by determiner omission, constituting LDO. The morphology of agreement among nouns, definite determiners, nominal modifiers, and noun phrase particles plays an important role in tracking the development of LDO in Albanian, as will be analyzed in §5.2.

Definite or indefinite nouns can take a range of modifiers, including adjectives, possessive pronouns, genitive-case nouns, ablative-case partitive nouns, relative clauses, and prepositional phrases. Some of these modifiers also decline for agreement with the noun in grammatical features. Modifiers governed by the determiner phrase, including adjectives, normally follow the noun. Thus, the most common configuration of constituents in the Albanian noun phrase is Noun, (Definite Determiner), Adjective, followed by other modifiers such as prepositional phrases or relative clauses. In agreement, some adjectives decline for gender and number alone, as with *shqiptar* 'Albanian,' which shows the same form whether the noun is indefinite, as in (2a-b), or definite, as in (2c-d), and does not decline for case, cf. nominative-case (2a-d) with accusative-case (2e).

(2) a. një qytet   shqiptar
         a   city.M.NOM Albanian.M
         'an Albanian city'

     b. një vajzë   shqiptare
         a   girl.NOM Albanian.F
         'an Albanian girl'

     c. qytet-i   shqiptar
         city-M.NOM.DEF Albanian.M
         'the Albanian city'

     d. vajz-a   shqiptare
         girl-F.NOM.DEF Albanian.F
         'the Albanian girl'

     e. në qytet-in   shqiptar
         in city-M.ACC.DEF Albanian.M
         'in the Albanian city'

Some possessive pronouns also fall into this set of modifiers that decline for agreement, directly following the noun. When modified by a possessive pronoun, the noun takes a definite article, as in (3a-b). This determiner governs the whole of the noun phrase. As with adjectives, possessive pronouns follow the head noun, yielding the order Noun, Definite Determiner, Possessive Pronoun.

(3) a. shok-u   im
         friend-M.NOM.DEF my.M.NOM
         'my (male) friend'

b. shoq-ja   ime
friend-F.NOM.DEF my.F.NOM
'my (female) friend'

Another set of modifiers takes an obligatory element known as an 'adjectival particle' (Newmark et al. 1982:67) or a 'particle of concord' (Friedman 2004:25), which shows agreement with the head noun in case, number, gender, and, particularly, in definiteness. The role of the particle of concord in expressing agreement with definiteness in the noun is noteworthy. This agreement pattern reinforces the role of definite reference in the grammatical system of Albanian.

The particle of concord applies to a large set of adjectives, as in (4a-b), to some possessive pronouns, as in (4c-d), and to declining complementizers, as in (4e). There is extensive syncretism in the forms of the particle of concord, with only four overt phonological forms (*i*, *e*, *të*, and *së*) distributed among all combinations of case, number, gender, and definiteness. In the examples below, for clarity, the particle of concord will be glossed with all elements of the noun for which it is encoding agreement, even if it is syncretic in form with other particles of concord.

(4)   a.   në një libër **të**   madhë
         in   a   book M.ACC.IND big.M
         'in a big book'

       b.   në libr-in   **e**   madhë
         in   book-M.ACC.DEF M.ACC.DEF big.M
         'in the big book'

       c.   libra-t   **e**   mi
         books-M.NOM.DEF M.PL.NOM.DEF my.M.PL
         'my books'

       d.   prej   librave   **të**   mi
         from books.M.ABL M.PL.ABL my.M.PL
         'from my books'

       e.   libr-i,   **i**   cil-i   për herë **të**   parë është botuar   në
         book-M.NOM.DEF M.NOM.DEF which-M.NOM.DEF for time F.ACC.IND first.F is   published in
         Kosovë
         Kosovo.F.ACC
         'the book, which has been published in Kosovo for the first time'

Even though the particle of concord is treated as a separate word, this particle always accompanies the adjective (4a-b), pronoun (4c-d), or complementizer (4e) to which it is attached. Thus the particle of concord forms an essential part of the constituent structure of the modifiers for which it is required (Friedman 2004:7). Not all lexical items belonging to the category of adjective, possessive pronoun, or complementizer require a particle of concord. When a nominal modifier does require a particle of concord, its presence is mandatory even if the nominal modifier is not syntactically adjoined to the noun or if there is no overt head noun. This is demonstrated by the predicate in (5) (Topalli 2009a:351). In cases where the adjective is not directly modifying the noun, as with *i ri* 'new,' the particle of concord takes the indefinite form, even if the corresponding noun is definite. Definiteness agreement only occurs for modifiers directly governed by the determiner phrase. Indefinite agreement is the default form for the particle of concord.

(5) Fustan-et   janë **të**   reja.
     dresses-F.NOM.DEF are **F.NOM.IND** new.F.PL
     'The dresses are new.'

The particle of concord accompanies not only adjectives and possessive pronouns, but also nouns in the genitive case acting as modifiers. Thus, the dative case is distinguished from the genitive, as outlined in §2.1, by the presence of a particle of concord that agrees with the head noun. This is demonstrated in (6), where the particle of concord accompanying the genitive noun *i vajz-ës* 'of the girl' agrees with the head noun *fustan-i* 'the dress.' In this way, when the modifying noun is in the genitive case, its particle of concord agrees with the head noun, which is nominative case in (6). Case is doubly expressed on genitive nouns, an example of Suffixaufnahme strictly grammaticalized into an adjectival article that is perhaps unique among documented languages (Plank 1995:74-76). Besides the presence or absence of a particle of concord, the genitive and dative cases are identical in declension (Boretzky 1968:132).

(6) fustan-i   i   vajz-ës
     dress-M.NOM.DEF M.NOM.DEF girl-F.GEN.DEF
     'the girl's dress'

Sometimes, where the noun fails to distinguish case, gender, or number in its overt morphology, the particle of concord will make the properties of the noun explicit. For example, indefinite nouns do not show distinctive forms for accusative and nominative. The indefinite nominative particle of concord, however, is distinct from the indefinite accusative particle of concord.

(7) a. një libër   **i**   ri
        a   book.M.NOM **M.NOM.IND** new.M
        'a new book'
    b. në një libër   **të**   ri
        in a book.M.ACC **M.ACC.IND** new.M
        'in a new book'

Thus, while the noun *libër* 'book' with indefinite reference shows the same form whether it is in a position that requires nominative or accusative case, the particle of concord differentiates nominative *i* (7a) from accusative *të* (7b). Given its complex pattern of agreement with all morphological features of the noun phrase, including definiteness, the particle of concord is an essential component of noun phrase structure, and its form can be used as a diagnostic for case, number, gender, and definiteness in the absence of overt morphology on the noun.

The post-posed definite determiner: A phrasal affix

In all examples cited thus far, modifiers in the determiner phrase have followed the head noun. This gives the impression that the definite determiner always attaches to the noun, and may in fact be part of nominal morphology. Indeed, here and in other grammars of Albanian, data is generally presented in a way that implies there is an indefinite declension and a definite declension for every noun. This is because the definite determiner is tightly bound to nominal morphology, sometimes triggering changes in the noun stem (*libër* 'book' > *libr-i* 'the book' with the syncope

of the unstressed vowel *ë*) or changing in form in response to the structure of the noun stem (*shok* 'friend' > *shok-u* 'the friend' with the alternation of *-u* for normal *-i* in the presence of a velar consonant). Nonetheless, the definite determiner is a phrasal affix, and it is not exclusively a suffix to the noun. Under the right syntactic conditions, the definite determiner will appear on a modifier, rather than on the noun itself, in a noun phrase with definite reference.

The nature of the definite determiner as a phrasal affix rather than an inherent part of nominal morphology can be illustrated with nominal modifiers as a diagnostic. Under some circumstances, an adjective in the noun phrase can precede the head noun (Newmark et al. 1982:194). This syntactic variation in noun phrase structure makes clear that the definite article acts as a phrasal affix. When an adjective occurs as the first element in a noun phrase, the definite article attaches to the adjective and the noun appears as its bare singular or plural stem form. Thus, the Albanian definite article is not a fixed element of noun declension, but rather a post-posed article whose position is determined by the phrasal structure of the noun and its modifiers. The article always occurs in second position, following the first declinable constituent in the noun phrase.

(8) më   i   madh-**i**   qytet turistik bregdetar i   Shqipëri-së
    most M.NOM.DEF big-**M.NOM.DEF** city   tourist.M coastal.M M.NOM.DEF Albania-F.GEN.DEF
    'the biggest coastal tourist city of Albania'

In (8), the definite article declining for masculine, nominative, and singular attaches to the adjective *i madhë* 'big,' since this adjective is the first declinable constituent in the noun phrase. It does not attach to *më* 'more/most,' because this constituent is not declinable, nor does it attach to the noun *qytet*, because this constituent is not the first in the noun phrase. The position of the definite article is syntactically determined. This assignment of the definite article to a position post-posed to the first declinable element in the noun phrase is a famous characteristic of the Balkan linguistic area that Albanian shares with other languages in the region, such as Macedonian, Romanian, Bulgarian, and Aromanian (Miklosich 1862; Seliščev 1925; Sandfeld 1930; Boretzky 1968:131; Hock and Joseph 2009:405).

Other declinable elements, such as possessive pronouns, can also take the definite article, as in (9). Here, the possessive pronoun is doubly marked both with a definite article (the suffix *-i*) and for definite agreement with a nominal head (the particle of concord *i*), which may be absent in the sentence. The definite article attaches to the only available head in the noun phrase, which is the possessive pronoun.

(9) Është i   im-**i**
    is   M.NOM.DEF my.**M.NOM.DEF**
    'It's mine.'

There is some variability in the use of the definite article with nouns modified by a possessive pronoun. Either the noun may be treated as the first declinable constituent in the phrase, as in (10a), or the noun and its possessive modifier may be treated as a single constituent, as with *shok im* 'my friend' (10b), in which case the definite article attaches to the noun and its possessive modifier as a unit.

(10) a. një shok-**u**   im
          a   friend-**M.NOM.DEF** my.M.NOM
          'a friend of mine'

    b. një shok   im-**i**
       a   friend.M.NOM my-**M.NOM.DEF**
       'a friend of mine'

Only when the combination of case and number prevents a definite article from ever occurring, as described in §2.1, does the definite article fail to appear. In (11), one would expect a definite article to be post-posed either to *miqve* 'friends' or *të mi* 'my,' as we see in (10a-b). Since *miqve* is a plural, ablative-case noun, however, a definite article is not licensed, no matter the syntactic structure. This is a morphological condition, and its application here illustrates that it generalizes across all syntactic domains without exception.

(11)   a. sipas   disa miqve(*-t)   të   mi
           according-to some friends.M.ABL.PL(*-DEF) M.ABL.PL my.M.ABL.PL
           'according to some of my friends'
    b. sipas   disa miqve   të   mi(*-të)
           according-to some friends.M.ABL.PL M.ABL.PL my.M.ABL.PL(*-DEF)
           'according to some of my friends'

Example (11) is a particularly important illustration of the fact that the lack of a distinct definite article for genitive-, dative-, and ablative-case plural nouns is not the result of article omission. There is no circumstance in which a definite article is compatible with a plural noun in these cases, even when the presence of other modifiers would allow the definite article to attach separately from the head noun, or the syntax of the phrase normally requires a definite noun. The lack of a definite article for this combination of case and number generalizes across all syntactic domains. This is a separate phenomenon from the omission of the accusative-case definite determiner that occurs in the domain of locative prepositional phrases, as well be discussed in §2.3. Although subtle, the distinction between the loss of the genitive/dative/ablative definite article across all syntactic domains and the loss of the locative definite article within a specific syntactic domain has important consequences when viewed in the context of the diachronic changes in Albanian grammar. Analyzing these changes separately allows us to see the origin of LDO in language contact, rather than internally motivated grammatical developments.

    These examples serve to show that the definite article is a distinct phrasal affix in Albanian, governed by syntactic conditions that require it to be post-posed to the first declinable constituent in the noun phrase. The expression of definiteness in the Albanian noun phrase involves not only the post-posed definite article, but also a marker for agreement in definiteness between modifiers and the head noun, as represented by the particle of concord. Under select conditions, nouns may trigger agreement in grammatical features for which they lack overt morphology, and so the particle of concord may show case and definiteness agreement even where overt morphology appears on the noun. Indefiniteness and definiteness are both morphologically encoded, but the indefinite article appears at the beginning of the noun phrase, as in English. The definite article, in contrast, appears in a different place in the noun phrase and interacts with nominal declension. These definite articles are subject not only to syntactic conditions, of course, but also to morphological conditions that interact with syntax to determine whether an article appears with the noun, and where it appears.

    This section has analyzed the grammatical coding of definite and indefinite reference in Albanian noun phrase syntax and morphology. How and when definite reference is grammatically

coded also depends on the semantic structure of nominal reference in Albanian. There is a standard set of these semantic conditions on the expression of definite reference that apply to the majority of Albanian determiner phrases. There is also an exceptional set of semantic conditions that only apply in the domain of certain prepositional phrases, resulting in LDO. Understanding the standard conditions helps to show how prepositional phrases trigger the exceptional behavior of determiner omission.

2.2   Definiteness in Albanian: Syntax and semantics

Thus far, this analysis has treated definiteness as a grammatical category without directly addressing its meaning. Distinguishing identifiable or unique referents known to both the speaker and listener from newly introduced referents in a discourse is the primary function of the grammatical and semantic feature of 'definiteness.' The application of definiteness to noun referents shows similarities across languages, in that definiteness appears to be a property of reference that human languages of diverse origins show a variety of grammatical means for marking. One of the common constructions for marking definiteness, often grammaticalized from a demonstrative, is the determiner. The exact circumstances under which definiteness is marked must be demonstrated for each individual language, with some languages treating particular categories of referents as grammatically definite when other languages would not mark them as definite. In this regard, Albanian and English show many similarities. The presence of definite and indefinite articles in Albanian often corresponds with their presence or absence in English translation. But there are details of article distribution in Albanian that differ from English in patterned ways. In addition, Albanian imposes certain syntactic and lexical restrictions on the appearance of the definite determiner that may not apply in English. Identifying these definiteness distribution patterns for nouns in argument positions is a necessary precondition for demonstrating how definiteness marking in prepositional phrases diverges both from the language-internal expectations of Albanian's system of definiteness marking as well as from cross-linguistic expectations. The semantic structure of definite reference is analyzed here in order to help identify LDO as a divergent grammatical pattern with its own consistent triggering conditions.

Albanian definite articles appear with a wider distribution than their English equivalents. In particular, the definite article accompanies generic nouns and proper nouns, including nouns with unique reference to people or places. The availability of an indefinite article results in a strongly non-specific reading for bare nouns in Albanian. Moreover, bare nouns in Albanian must be lexically governed, and thus are not licensed in argument positions such as the subject position of a verb. In contrast, they are licensed in the object position, which is a position lexically governed by the verb. As a result, it should be expected that bare nouns are interpreted as non-specific and indefinite in all positions, and that they are licensed only in lexically governed positions (which would not include prepositions). This is of great relevance to the analysis of determiner omission in prepositional phrases, where the grammatical pattern runs counter to expectation and bare nouns are interpreted as definite.

The syntactic and semantic distribution of definite articles and bare nouns

In general, bare nouns without a definite or indefinite article are much less common in Albanian than in English. The Albanian definite article is used more extensively, including for nouns with inherently definite, unique reference. Thus, definite reference is strongly associated with the presence of a definite article. When Albanian nouns appear bare, without any article, they are subject to a set of grammatical and semantic restrictions. This results in a strong association between bare nouns and predicational readings in Albanian.

   One example of a domain where bare nouns can be found is the copula construction. As the complement of a copula, a singular noun must be bare, as in (12a). It may only take an indefinite article if it is modified by a qualifying feature, such as the adjective *i mirë* 'good' in (12b) (Newmark et al. 1982:150). Modifiers that license the use of an indefinite article in a predicational clause are adjectives (12b), possessive pronouns (12c-d), genitive-case modifiers (12e), prepositional modifiers (12f), and relative clauses (12g). As will be demonstrated below, this is not the only case in which this set of modifying constituents acts collectively to determine the presence or absence of an article in the determiner phrase. This set of modifiers, syntactically subordinate to the noun, may condition the appearance of an article in cases where it would otherwise be omitted.

(12)   a. Ai ishte (*një) student.
           he was (*a)   student
           'He was a student.'

        b. Ai ishte (një) student **i   mirë**.
           he was (a)   student **M.NOM.IND good.M**
           'He was a good student.'

        c. Ai ishte një shok-u   **im**.
           he was a   friend-M.NOM.DEF **my.M.NOM**
           'He was a friend of mine.'

        d. Ai ishte një shok   **im-i**.
           he was a   friend **my-M.NOM.DEF**
           'He was a friend of mine.'

        e. Ai ishte një shok **i   Petrit-it**.
           he was a   friend **M.NOM.IND Peter-M.GEN.DEF**
           'He was a friend of Petrit's.'

        f. Ai ishte një student **me kohë të   pjesshme**.
           he was a   student **with time F.ACC.IND partial.F**
           'He was a part-time student.'

        g. Ai ishte një student **që kam   takuar në Tiranë**.
           he was a   student **that I.have met   in Tirana**
           'He was a student that I met in Tirana.'

   In contrast to the predicate position of a copula construction, the subject position of a verb does not permit a bare noun in Albanian, which indicates that bare nouns must be lexically governed by the verb. This can be illustrated by an example where a generic noun without a unique, identifiable referent occurs as the subject of a verb, a lexically ungoverned position. Even though the

reading for *luan* 'lion' in (13) is generic, it must occur with a definite article in order to form a grammatical sentence. Thus, the definite article here allows for a generic interpretation in which the speaker designates all potential referents that would fall under the category of 'lion' (Newmark et al.1982:159; Xhaferi 2011:67-68). This use is identical to the use of the generic definite article in English. In this case, it is the category of 'lion' that is marked as identifiable to the speaker and the listener by the definite article, rather than any individual lion. A bare noun cannot be used to express this same kind of reference in the subject position of a verb, as shown by (13b).

(13) a. **Luan-i**   është një kafshë   e   fortë.
          lion-M.NOM.DEF is   IND animal.F.NOM F.NOM.IND strong.F

          'The lion is a strong animal.'

      b. *Luan   është një kafshë   e   fortë.
          lion.M.NOM is   IND animal.F.NOM F.NOM.IND strong.F

The restriction on bare nouns in subject position also holds for nouns denoting abstract concepts such as *dashuri* 'love,' which would occur as a bare subject without a definite article in English. Thus, Albanian does not show the English distinction between abstract and concrete nouns with generic reference, requiring both sets of nouns to take the definite article in subject position, as shown by (14a), contrasted with the ungrammatical (14b).

(14) a. **Dashuri-a**   është e   mrekullueshme.
          love-F.NOM.DEF is   F.NOM.IND wonderful.F

          'Love is wonderful.'

      b. *Dashuri është e   mrekullueshme.
          love.F.NOM is   F.NOM.IND wonderful.F

          (Intended) 'Love is wonderful.'

Thus, Albanian requires a broader use of the definite article than English. While English allows bare plurals and nouns denoting abstract concepts to occur without a definite article in subject position, Albanian requires them to be marked as definite with an article. In contrast, bare nouns in both singular and plural are licensed as objects, since these positions are lexically governed. Bare nouns denoting mass quantities in object position, as in *uthull* 'vinegar' (15), are common and parallel English in structure and interpretation.

(15) Për të bërë **uthull**   molle   së   pari duhet të   copëtojmë
      for SBJN make **vinegar.M.ACC** apple.M.ABL F.DAT.DEF first should SBJN we.chop

      moll-ën   në copeza   të   vogla.
      apple-F.ACC.DEF in pieces.F.ACC F.ACC.PL.IND small.F.PL

      'To make apple cider vinegar, we should first chop up the apple into small pieces.'

Bare count nouns in the plural also appear in object position and are treated similarly in Albanian as in English. This is exhibited by the use of *lule* 'flowers' in (16).

(16) Petrit-i   i   dërgoi An-ës   **lule**.
      Peter-M.NOM.DEF 3SG.DAT he.sent Ana-F.DAT.DEF **flowers.F.ACC**

      'Peter sent Ana flowers.

Another important difference between Albanian and English is particularly useful for diagnosing any unexpected omission of the definite article: the use of definite articles with proper nouns. In (16), it may be noted that the proper nouns Ana and Petrit take definite articles with the associated marking of nominative case for *Petrit-i* and dative case for *An-ës*. Proper nouns in Albanian must be accompanied by definite articles if they are used as arguments, as in (17a) (Demiraj 1993:122). The bare form of a proper noun is not grammatical in most argument positions, as in (17b). Definite articles are required both for native Albanian names and for foreign names; e.g., Bill Clinton is rendered *Bill Klinton-i*. With two-part names, such as a first and last name, the entire name is treated as a single noun to which the definite article attaches. Indefinite forms of names serving as arguments only appear if they are in apposition with another name that will take the definite article, such as the case of *Petrit Merzak-u* in (17b-c), where the surname *Merzak* takes the definite article.

(17)   a.   Kjo   është për Petrit-**in**.
           this.F.NOM is   for Petrit-**M.ACC.DEF**
           'This is for Peter.'

        b.   *Kjo   është për Petrit.
           this.F.NOM is   for Peter
           (Intended) 'This is for Peter.'

        c.   Kjo   është për Petrit Murzak-**un**.
           this.F.NOM is   for Petrit Merzak-**M.ACC.DEF**
           'This is for Peter Merzaku.'

        d.   I-a   dha   Petrit Murzak-**ut**.
           3SG.DAT-3SG.ACC he.gave Petrit Murzak-**M.DAT.DEF**
           'He gave it to Peter Murzaku.'

In the coding of definite reference for proper names, Albanian shows a sharp contrast with English. While the more extensive use of definite articles with nouns showing generic reference is a matter of syntax and lexical governance, the use of definite articles with Albanian proper nouns is a matter of semantics. Since proper nouns inherently refer to unique, identifiable referents, Albanian requires the definite article and treats these names as if they were common nouns. As one would expect based on the behavior of common nouns, proper names do appear without a definite article when they are serving as predicates, as in the small clause of (18a) where the noun *Petrit* acts as the predicate of the pronoun *e* 'him,' or when they are serving in some other non-argument role, such as the vocative expression of (18b).[5]

(18)   a.   E   quajnë   **Petrit**.
           3SG.ACC they.call **Petrit**
           'They call him Petrit.'

---

[5]The use of the indefinite form of the name as a vocative has an exception: female names whose indefinite form would end in the schwa -*ë* and male names whose stems are monosyllabic retain the definite article in the vocative. Thus *Drit-a shkoi* 'Drita left,' but *O Drita, ela!* 'Hey Drita, come here!' rather than the expected **O Dritë, ela!*. This rule is phonologically conditioned and the commonality between female and male names is likely related to the fact that etymological final -*ë* [ə] is unpronounced in all but far southern dialects of Albanian. The constraint is thus that names in the vocative must contain at least two syllables, explaining the use of definite forms in the above cases.

b. O **Petrit**, ela!
VOC **Petrit**, come.here
'Hey Peter, come here!'

Not only proper nouns referring to persons, but also those referring to places must take the definite article in the subject and object positions of the verb, as in (19a) and (19b), respectively. The use of an indefinite article in (19c) shows how the bare form of the proper noun *Tiranë* 'Tirana' appears. Thus, any proper noun denoting a unique, identifiable referent must mark that definite reference with an article.

(19) a. **Tiran-a**   është një qytet   i   madh.
          Tirana-F.NOM.DEF is   a   city.M.NOM M.NOM.IND big.M
          'Tirana is a big city.'

      b. **Sekretar-i**   amerikan i   Shtet-it   viziton Tiran-ën.
          secretary-M.NOM.DEF American.M M.NOM.DEF state-M.GEN.DEF visits   Tiran-M.ACC.DEF
          'The American Secretary of State will visit Tirana.'

      c. Për **një Tiranë**   të   pastër.
          for **a Tirana.F.ACC** F.ACC.IND clean.F
          'For a clean Tirana.'

These examples demonstrate that the definite article in Albanian is used extensively both for specific, identifiable referents and for generic referents, whereas omission of the definite article only occurs in syntactically limited positions and can only denote a generic meaning. The definite article, then, is strongly associated with definite reference in argument positions, to the extent that even proper nouns must take a definite article, since a proper noun inherently refers to a specific, identifiable individual. Albanian may also express a generic interpretation in object positions with a definite noun in the plural (Kallulli 2001:217).

(20) I   dua **mace-t**   e   vogla.
      3PL.ACC I.love **cats-F.DEF.ACC** F.PL.DEF.ACC small.F.PL
      'I love small cats' or 'I love the small cats.'   (Kallulli 2001:217)

The distributional properties of the bare form make a generic interpretation available for definite plural nouns, regardless of position. Even contexts that license bare nouns, such as object position for count plurals or nouns denoting substances, also allow the use of a definite article. In sum, the use of the definite article in Albanian is considerably broader than English, and bare nouns appear less consistently.

The syntactic and semantic distribution of indefinite articles

Thus far, the appearance of bare nouns has been contrasted with the use of the definite article. Bare nouns also show a semantic contrast with nouns governed by the indefinite article. Here, again, a comparison between the interpretations of bare nouns and nouns governed by an indefinite article demonstrates that the bare noun strongly associates with non-identifiable reference in Albanian. Identifiable reference, in contrast, requires an article, whether this is a definite or an indefinite

article. The distinction is illustrated by count nouns in the singular, which can appear in bare form as the objects of verbs, with semantic limitations. The following examples illustrate the nuances in interpretation between bare singular count nouns and indefinite singular count nouns in object position.

(21) a. An-a   donte të blente   **fustan**.
          Ana-F.NOM.DEF wanted SBJN she.bought dress.M.ACC

          'Ana wanted to buy a dress' or 'Ana wanted to go dress-shopping.'

      b. An-a   donte të blente   **një fustan**.
          Ana-F.NOM.DEF wanted SBJN she.bought **a dress.M.ACC**.

          'Ana wanted to buy a dress' or 'Ana wanted to buy some (brand of) dress, i.e. a Versace dress or a Dior dress'

      c. An-a   donte t-a   blente   **një fustan**.
          Ana-F.NOM.DEF wanted SBJN-3SG.ACC she.bought **a dress.M.ACC**

          'Ana wanted to buy a (particular) dress.'

                                                              (Kallulli 2001:231-234)

In (21a), the verb *blente* 'she bought' takes the bare count noun *fustan* 'dress' as the verbal complement. This leads to a nonspecific interpretation, indicating that Ana wishes to buy something of the semantic *kind* dress, rather than any specific type of dress or any particular dress. Thus *blente fustan* 'buy dress' denotes an activity, rather than a telic action. The kind interpretation of (21a) contrasts with the type interpretations of (21b-c). These two sentences both denote the desire to purchase a particular *entity*. The identity of the entity is differentiated by the presence or absence of a direct-object doubling pronominal clitic, -*a*.[6] In (21b), the sentence could either denote a transparent interpretation in which Ana has a particular dress in mind that she wishes to buy, or an opaque interpretation in which she does not have a particular dress in mind, but there is some kind of subselection from the broader class of 'dress.' For example, (21b) could be used to express the sense that Ana wishes to buy a dress of a particular brand, though she has no individual dress in mind. In contrast, (21c) shows doubling of the indefinite object by a pronominal clitic *a* 'him/her/it,' which merges with the subjunctive particle. This object-doubling structure requires a wide-scope indefinite interpretation in which there is a particular dress that Ana wants to buy. This sentence is also marked for information structure in the discourse: the use of the doubling clitic suggests that *një fustan* 'a dress' will be a continuing topic of conversation.

The fact that (21a) denotes an activity while (21b-c) denote telic actions may give the impression that (21a) is better treated as a kind of noun incorporation of *fustan* into the verb *blente*. But noun incorporation does not account for all available structures. The bare noun *fustan* may be separated from the verb while retaining the same truth-conditional meaning for the sentence as a whole, as in (22), where *fustan* undergoes focus fronting. Thus, as Kallulli (2001:231-234) argues, it is clear that bare nouns in object positions form independent constituents that can serve as arguments.

---

[6]Albanian, like other languages of the Balkans, frequently makes use of pronominal object clitics that agree with the overt object, leading to duplicate overt expression of the object. Under some circumstances, this object-doubling clitic is required by the grammar; under other circumstances, it is facultative and depends on the discourse context in which the sentence appears. The complexities of Balkan object doubling are beyond the scope of this analysis, but have been discussed extensively in the literature on Albanian and other languages, see Kallulli and Tasmowski (2008), Franks (2009), Friedman (2006b), Kallulli (2001), Leafgren (2002), *inter alia*.

(22) **Fustan**   doja   të bleja.
dress.M.ACC I.wanted SBJN I.bought

'It was a dress that I wanted to buy' or 'It was dress-shopping that I wanted to do.' (Kallulli 2001:231-234)

These examples reinforce the point that bare nouns, without an article, are consistently interpreted generically and non-specifically. The presence of an indefinite or definite article allows for a specific interpretation, but does not require it. While definite and indefinite articles are widely available and frequently required, bare nouns are far more restricted in their distribution. In most argument positions, a specific referent requires an article of some kind, and a specific, definite referent is only possible with a definite article. This is further illustrated by (23). Though they function as arguments, bare nouns in object position have restrictions on their scope, and thus the bare noun *biçikletë* 'bicycle' can only take narrow scope as in (23a). A wide-scope existential interpretation as in (23b), where the speaker has in mind a specific bicycle that the speaker does not want, is impossible with a bare nominal argument.

(23) Nuk dua   **biçikletë**.
not I.want **bicycle.F.ACC**

'I don't want a bicycle.'

  a. > It is not the case that I want a bicycle

  b. > *There is a bicycle that I don't want.

(Kallulli 2001:237)

The interpretational restriction on bare nouns is reinforced by the data in (24), where *shumë fëmijë* 'many children' must take scope over *biçikletë*, such that *biçikletë* has a distributive reading, despite being a singular noun, as in (24a). The bare noun cannot take wide scope over the subject, as in (24b).

(24) Shumë fëmijë   blenë   **biçikletë**   dje.
many children.F.NOM they.bought **bicycle.F.ACC** yesterday

'Many children bought a bicycle yesterday.'

  a. > There were many children and for each child, that child bought a bicycle yesterday.

  b. > *There were many children and they all, together, bought one bicycle.

(Kallulli 2001:238)

Given all this data together, we can provide general guidelines for how definite articles, indefinite articles, or the omission of an article should be interpreted in regard to nominal reference in Albanian. The conclusions are summarized in Table 2.5. These represent the standard rules for the semantic interpretation of Albanian nouns, whether bare or accompanied by a definite or indefinite article. Reference to a unique, identifiable referent requires a definite article for all kinds of nouns, whether common or proper. Generic reference may require a definite article, depending on the syntactic position of the noun. In either case, the distribution of the definite article in Albanian article positions is more extensive than in English. Definite reference is very strongly tied to the presence of an overt definite article in Albanian. We should expect bare nouns, without a definite or indefinite article, to denote only highly non-specific referents that are not identifiable to the listener.

Table 2.5: The distribution of definites, indefinites, and bare nouns in Albanian

	Specific	Generic	Lexically ungoverned	Predicational
Definite singular (*qytet-i*)	✓	✓	✓	When modified
Indefinite singular (*një qytet*)	✓	✗	✓	When modified
Definite plural (*qytete-t*)	✓	✓	✓	When modified
Bare singular (*qytet*)	✗	✓	✗	✓
Bare plural (*qytete*)	✗	✓	✗	✓

Albanian nouns: kind- or predicate-denoting?

It is worth specifying exactly how this analysis views the semantic reference of a bare noun in Albanian. This allows us to distinguish bare nouns in other contexts from the bare nouns that appear in LDO constructions, where we must speak of zero marking for definite reference. The distribution of definites, indefinites, and bare nouns outlined in §2.2 parallels that of Romance languages such as Italian, in which bare nouns must be lexically governed and must have an indefinite interpretation (Longobardi 1994). The conclusion that one would draw within Longobardi's framework, as well as within approaches such as that of Chierchia (1998) is that noun phrases (NPs) denote predicates of type $< e, t >$ and that bare nouns serving as arguments project a null determiner (D) to shift them to type e. The subject-object asymmetry in the use of null-determiners derives from lexical government. While objects are lexically governed by the verb, subjects are not. Since these null determiners must be lexically governed, in parallel with other null elements, the distributional facts outlined in Table 5 can be derived from distributional restrictions on a type-shifting null D.

Kallulli (2001), however, argues that Albanian bare NPs are predicates even in argument position. Her reasoning for this rests on the ambiguous interpretational possibilities of Albanian indefinites with *një*. Since they are ambiguous between a specific and nonspecific reading, Kallulli maintains that they must be underspecified as to their type: either $< e, t >$ or e. She holds that verbs such as *blente* 'buy' in (21) can undergo typeshifting from $< e, < e, t >>$ to $<< e, t >, < e, t >>$ and thus take predicate complements.

$$\lambda P \lambda x \exists y [P(y) \wedge BUY(x, y)]$$

This then leads to a non-specific interpretation for both indefinite and definites under certain circumstances. She derives the difference between clitic-doubled and non-clitic doubled indefinites such as those found in (21b-c), for example, from the differing types of the direct object *një fustan*. When no doubling clitic is present, the interpretation of *një fustan* is ambiguous. But because doubling clitics can only select arguments, the direct object of (21c) with the doubling clitic *a* 'him/her/it' must be specific.

The drawback of this approach is that while it captures the distributional restrictions on indefinites with object doubling clitics, but fails to account for the significant parallels between Albanian and other Romance languages. Chierchia described such languages as NP[–arg, +pred], in that bare NPs seem to be available only as predicates, and any usage of them as arguments is subject to the requirement of lexical governance. Kallulli's approach also appears to contradict the theoretical generalizations of Szabolcsi (1987) and Stowell (1989) that Ds serve as typeshifters from $< e, t >$

to e that allow NPs to function as arguments, to limited descriptive gain. If one were to choose an approach like Chierchia's semantic representation instead and categorize Albanian as an NP[–arg, +pred] language, then one could capture the distribution of bare singulars and plurals in a licensing condition: null-D must be licensed by a lexical head or the functional head of a Focus Phrase (assuming a Focus Phrase projection like that proposed in Rizzi (1997)). Generic readings of definites are obtained with the intensional abstraction over worlds/situations that applies to overt D when the null-D that normally derives kind interpretations is not licensed.

The remainder of the argument in this thesis will work from the supposition that Albanian is an NP[–arg, +pred] language in which bare NPs gain their kind interpretation from a null D that must be lexically governed. When a null D is not licensed, or when other semantic constraints demand the presence of an overt article, definites may have generic interpretations derived by the means of Chierchia's typeshifting operations. The type of the NP itself remains $< e, t >$. As a practical result, it would expected that the configurations of referential type and article distribution in Table 2.6 would apply across the Albanian grammatical system.

Table 2.6: The standard distribution of referent type and article usage in Albanian

Referent type	Associated article
Indefinite non-specific	null or indefinite[a]
Indefinite specific	indefinite
Definite non-specific	definite
Definite specific	definite

[a] subject to a lexical government requirement

To summarize, Albanian determiner phrases with definite reference always show a definite article post-posed to the first constituent of the noun phrase (§2.1). Definite reference is grammatically reinforced by agreement with nominal modifiers (§2.1). Only under certain combinations of case and number is the definite article prohibited (§2.1). Semantically, bare nouns without a definite or indefinite determiner are highly restricted in both meaning (§2.2) and syntactic position (§2.2). From a semantic-type perspective, bare nouns must be interpreted as predicates and have their kind interpretation licensed by lexical government (§2.2). Having laid out these morphological, syntactic, and semantic conditions, we may now return to the object of analysis for the remainder of this dissertation: determiner omission in prepositional phrases. As I have demonstrated, Albanian requires the definite article in a wide variety of argument positions for all cases of definite reference, including proper nouns with unique reference. Bare nouns, in contrast, are strongly associated with an non-specific, indefinite interpretation. However, the distribution of determiners in prepositional phrases violates the generalizations of Table 2.6 in significant ways that require an explanation.

2.3 Definiteness in Albanian: Exceptional marking in prepositional phrases

The analysis in §2.2 covers the semantic distribution and corresponding overt expression of definiteness through definite and indefinite determiners for most argument positions. My analysis has not, however, addressed prepositional phrases. For this syntactic domain, it becomes more difficult to characterize the link between definite reference and Albanian determiners. Albanian prepositional phrases condition an exceptional distribution of definite and indefinite articles, a distribution that does not match the distribution found in other syntactic positions, particularly argument positions of the verb. Only prepositions that satisfy a combination of selectional and semantic restrictions trigger divergent marking of definiteness in Albanian noun phrases. The combination of the noun case for which a preposition selects and the semantic contribution of that preposition to the utterance as a whole determines how definiteness is marked on the following noun phrase.

Prepositions that select the nominative, dative, or ablative case require a definite article to express definite reference, as expected from the analysis in §2.2 of definite and indefinite articles in other argument positions. Thus in §2.3 I demonstrate that definite determiners, in principle, are available in the syntactic domain of all prepositions. Prepositions that select the accusative case, however, only require a definite article for nouns that take modifiers such as adjectives. When an accusative-selecting preposition takes an unmodified noun, definite reference is indicated by the omission of any article. In §2.3, I show that using a definite article in a context where omission is expected is ungrammatical. In §2.3, I demonstrate that this applies to all nouns, whether common or proper, without lexical exceptions. In §2.3, I analyze the meaning of the accusative-selecting preposition as a further determining condition for determiner omission. I show that if the accusative-selecting preposition denotes a location or a goal of movement, determiner omission applies. If the preposition expresses a more abstract relation, however, determiner omission does not apply and a definite article must appear as expected. Thus, determiner omission occurs only within the domain of accusative-selecting prepositions, with locative meaning, governing unmodified nouns. The triggering of determiner omission is not lexically determined by the preposition, as in some other languages. Rather, the case-selection and semantic properties of the preposition constitute the conditions for determiner omission. Such conditions on the appearance of the definite article do not apply to any other argument positions in the Albanian language, making this a unique pattern of expressing definite reference within the broader grammatical system. This constellation of semantic, morphological, and syntactic factors together produce LDO, and repeat themselves in other languages. Establishing the triggering conditions for LDO allows us to distinguish languages that have it from languages that do not, and to track its development over time and space.

The availability of definite determiners in prepositional phrases

Case selection is one of the key factors that produce determiner omission. The case-selection potential of Albanian prepositions is very broad: Albanian prepositions can select nominative, dative, ablative, or accusative case. Of these, nominative-, dative-, and ablative- selecting prepositions show the same definite determiner distribution as found in argument positions of the verb.

Prepositions that select the nominative case can designate a goal, a source, a location, or a more abstract relation such as the agent of a passive verb. In these nominative prepositional phrases,

definiteness is marked by a definite article, and indefiniteness by an indefinite article, as in other argument positions in Albanian. As illustrated by the examples, definite articles can appear within the domain of at least some prepositions, and articles code definite or indefinite reference just as would be expected for other syntactic domains.

(25) a. për të nxjerrë ujë   të   pastër **nga pus-i**
          for SBJN draw   water.F.ACC F.ACC.IND clean.F.SG **from well-M.NOM.DEF**

          'To draw clean water from the well.'

     b. duke nxjerrë ujë   **nga një pus**
          'PRS.PRT drawing water.F.ACC **from a well.M.NOM**

          'drawing water from a well'

In (25a), the noun *pus-i* 'the well' designates a definite referent and takes the nominative, masculine, definite article *-i*. This contrasts with (25b), where *pus* 'well' designates an indefinite referent in the prepositional phrase *nga një pus* 'from a well.' Although it is not possible to tell the difference between nominative and accusative case with unmodified indefinite nouns, the use of the nominative definite article for definite nouns makes clear that the preposition *nga* 'from' takes nominative case. Here, *nga* denotes a source, but it can also denote an agent, as in (26).

(26) Libr-i   është shkruar **nga një shkrimtar**.
      book-M.NOM.DEF is   written **by an author.M.NOM**

      'The book was written by an author.'

Example (27) shows that the requirement for a definite article to designate definite reference also applies to proper nouns in the domain of nominative-selecting prepositions, such as the personal name *Petrit-i* 'Peter' with the preposition *te* 'at.'

(27) te Petrit-**i**
      at Peter-**M.NOM.DEF**

      'at Peter's'

For the preposition *te*, a definite article is in fact required unless there is a modifier. Thus the interpretation of the referent can be either definite or indefinite for nouns without modifiers, as in (28a), whereas an indefinite article is required for indefinite reference in the presence of nominal modifiers, as in (28b) (Newmark et al. 1982:291).[7] This complication in interpretation still does not affect the morphological requirement for a definite article.

(28) a. te lum-**i**
          by river-**M.NOM.DEF**

          'by the river'

     b. te **një lumë**   ku   ka lule
          by **a river.M.NOM** where has flowers.F.ACC

          'by a river where there are flowers'

---

[7]In non-standard registers, one can observe constructions like *te një lumë* 'by a river' with a noun in the nominative without modifiers governed by the indefinite article. Whether the register is standard or non-standard, the preposition *te* takes the nominative case and requires a definite article to express definite reference for the noun.

Another set of prepositions select the ablative case. This is equivalent to the dative case for definite and singular indefinite nouns, as in (29a-b). That the ablative case is selected only becomes apparent with indefinite, plural nouns, for which there is an ablative-case ending distinct from the dative, as in (29c). In spoken Albanian, this distinct ablative plural indefinite declension may be lost, in which case it is simply substituted with the dative (Newmark et al. 1982:139). As with nominative-selecting prepositions, definite reference is marked with a definite article, and indefinite reference with an indefinite article, following the general pattern of Albanian nominal arguments. Definite articles cannot appear with plural ablative nouns, even if the noun has definite reference as in (29d), but as analyzed in §2.1, this is a morphological property of the plural dative and ablative cases. The condition applies to all dative and ablative plural nouns, regardless of the domain in which they appear.

(29) a. para   lum-**it**
          before river-**M.ABL.DEF**
          'before the river'

      b. para   **një lumi**
          before **a river.M.ABL.IND**
          'before a river'

      c. para   **disa ditësh**
          before **some days.F.ABL.IND**
          'a few days before'

      d. para   **shtëpive**
          before **houses.F.ABL**
          'in front of the houses'

The exceptional behavior of accusative-case selecting prepositions

A significant set of prepositions in Albanian select the accusative case. Accusative-selecting prepositions are the syntactic domain for which determiner omission is a relevant grammatical factor. Within this domain, the accusative definite article only appears with modified nouns. Unmodified nouns with definite reference must appear bare, without a definite article, while nouns with indefinite reference require an indefinite article. Thus, unlike the rest of the Albanian grammatical system, bare Albanian nouns within the domain of accusative prepositions are interpreted as definite, with an identifiable or unique referent.

The simple accusative prepositions are *në* 'in, on,' *me* 'with,' *pa* 'without,' *më* 'on (date),' *mbi* 'on, over,' *nën* 'under,' *ndër* 'toward, among,' *për* 'for.'[8] Generally, these prepositions denote a location, a direction of movement, or a goal. Albanian, like other Balkan languages, does not distinguish between location ('where') and movement towards a goal ('whither'), using a single interrogative for both concepts, as in (30a–b), and correspondingly, a single preposition to describe both static location and movement toward a goal, as in (30c–d). Thus, motion toward a goal and location at a goal can both be collapsed into single linguistic encoding of 'relation to a location,'

---

[8]These prepositions may be compounded to express more complex meanings; they still take the accusative case when compounded.

which I refer to as *locative*, for which accusative-selecting prepositions are the most common means of expression.

(30) a. Ku   po   shkon ti?
          where PROG go   you.NOM
          'Where are you going?'

     b. Ku   je?
          where you.are
          'Where are you?'

     c. Hipi   në çati-**në**   e   shtëpi-së
          jumped on roof-**F.ACC.DEF** F.ACC.DEF house-F.GEN.DEF
          'He jumped onto **the** roof of the house.'

     d. Ishte në çati-**në**   e   motel-it
          was   on roof-**F.ACC.DEF** F.ACC.DEF motel-M.GEN.DEF
          'It was on **the** roof of the motel.'

The case of the noun selected by a preposition like *në* 'in' is not necessarily visible on the noun itself. As outlined in §2.1, overt morphology distinguishing nominative from accusative only appears on definite determiners or the particle of concord that accompanies modifiers. When the noun selected by the accusative preposition takes a modifier, as with *çati-në e shtëpi-së* 'the roof of the house' or *çati-në e motel-it* 'the roof of the motel,' the accusative definite article appears as expected. Modifiers can be genitive nominals, as in (30a–b), adjectives (31a), possessive pronouns (31b), adverbs (31c), relative clauses (31d), and prepositional phrases (31e). This recalls the set of modifiers that trigger the presence of the indefinite article with predicate nominals, as noted in §2.2. These nominal-modifying expressions form a coherent class affecting the distribution of the article in determiner phrases.

(31) a. Ishte   në tavolinë-**n**   **tjetër**
          he.was at table-**F.ACC.DEF** **other**
          'He was at **the** other table.'

     b. Ishte   në tavolinë-**n**   **e**   **tyre**
          he.was at table-**F.ACC.DEF** **F.ACC.DEF** **their**
          'He was at **their** table.'

     c. Ishte   në tavolinë-**n**   **afër**
          she.was at table-**M.ACC.DEF** **near**
          'She was at **the** table nearby.'

     d. U   ul   në tavolinë-**n**   **që ndodhej   pranë derë-s**
          REFL he.sat.down at table-**F.ACC.DEF** **that was.located next.to door-F.ABL.DEF**
          'He sat down at **the** table near the door.'

     e. Janë   ulur në tavolinë-**n**   **në kuzhine**
          they.were sat   at table-**F.ACC.DEF** **in kitchen.ACC**
          'They sat down at **the** table in the kitchen.'

The condition of modification thus extends across a broad set of syntactic sub-constituents of the noun phrase. This same group of modifiers also triggers the appearance of the indefinite article in the predicate position of a copular construction, as noted in §2.2. It can be surmised that these modifiers form a coherent set that affect the realization of Albanian determiners.

In the case of accusative-selecting locative prepositions, definite reference is expressed by an accusative definite article as long as the noun is modified. But (31e) features an example of the exceptional behavior of unmodified nouns governed by accusative prepositions. Though *tavolinë-n* 'the table' takes the definite article to express definite reference, *kuzhine* 'kitchen,' which is also governed by the accusative-selecting preposition *në*, has no definite article. Nonetheless, the correct translation of *në tavolinë-n në kuzhine* is 'at the table in the kitchen.' Both the noun *tavolinë* and the noun *kuzhine* stand for definite, identifiable referents in this context. But the lack of a definite accusative article on *kuzhine* is not an accidental gap. In this context, the noun *kuzhine* does not take a definite article, and cannot, grammatically. This can be contrasted with the prepositional phrase *pranë derë-s* 'near the door' in (31d), where the ablative definite article indicates that the noun *derë* 'door' has definite reference. While the ablative-selecting preposition *pranë* requires a definite article in this context, the accusative-selecting preposition *në* requires the omission of the definite article.

This is a general rule for Albanian nouns governed by accusative prepositions. If they are modified by another constituent, then they take a definite article to express definite reference. Thus, all instances of *tavolinë* in (31) take the accusative definite article, since the noun phrase in each example sentence contains a modifying expression of some kind. But if the noun governed by the accusative preposition has no phrasal modifiers, then definite reference is expressed by the *absence* of a definite article, as with *kuzhine* in (31e). The use of a definite article for unmodified nouns in the domain of accusative prepositions is identified by speakers as uniformly ungrammatical.

(32) a. Hipi   në **një** makinë
          he.climbed in **a**   car.ACC
          'He got into **a** car.'
      b. Hipi   në **makinë**
          he.climbed in **car.**ACC
          'He got in **the** car.'
      c. *Hipi   në makinë-n
          he.climbed in car-F.ACC.DEF
          (Intended) 'He got in the car.'
      d. Hipi   në makinë-**n**   zyrtare
          she.climbed in car-**F.ACC.DEF** official.F
          'She got into **the** official car.'

When the noun *makinë* 'car' has indefinite reference and is governed by an accusative preposition, the noun must appear with an indefinite article *një*, as in (32a). If the indefinite article is removed, as in (32b), speakers identify the referent of the noun as definite (Newmark et al. 1982:150). It is not that the difference between definite and indefinite reference is neutralized within the domain of accusative prepositions. Rather, the absence of a definite article in (32b) is the signal itself of the definiteness of the referent. A definite article is not licensed without a nominal modifier, as

in (32c), even when definite reference is intended. Only in the presence of a modifier is the definite article licensed, in which case it can be seen that the preposition is selecting the accusative case, as in (32d). In the subsequent analysis, I will argue that this is grammatical rule in Albanian that is determined by the prepositional environment in which these nouns appear, rather than a lexically or morphologically conditioned phenomenon.

Determiner omission with accusative prepositions: Neither morphological nor lexical

As analyzed in §2.1 and §2.1, there are combinations of case and number in Albanian that yield a morphological neutralization of the distinction between definite and indefinite. Taking a broader view of definite determiners cross-linguistically, many languages allow lexical exceptions to what is otherwise a general rule that definite articles are required to indicate definite reference. In the case of Albanian, neither nominal morphology nor lexical rules can account for the omission of determiners in the domain of accusative prepositions. The syntactic position of the definite article as a post-posed element of the noun phrase means that it is not possible for the preposition to undergo phonological merger with the definite article—no property of the prepositions captures the missing marker of definite reference. Thus LDO is neither a morphologically, nor a lexically determined phenomenon.

Close attention to the analysis of Albanian case in §2.1 would yield the conclusion that this is not the only instance in Albanian where the presence of an indefinite article signals indefinite reference, while the absence of a definite article signals definite reference. Plural nouns in the dative, genitive, and ablative cases cannot take a definite article. Thus, the only way to determine whether the reference of a dative/genitive or ablative plural noun is definite or indefinite is by the presence or absence of the indefinite article. This looks similar to determiner omission in accusative prepositional phrases.

(33)   a. Ia   dhashë disa miqve   të   mi
            them.it I.gave some friends.M.DAT M.DAT.PL.IND my.M.PL
            'I gave it to some friends of mine.'
       b. Letr-a   miqve   të   mi
            letter-F.NOM.DEF friends.M.DAT M.DAT.PL.DEF my.M.PL
            'The letter to my friends.'

While the dative recipient in (33a) is indefinite, as indicated by the indefinite article *disa*, the recipient is definite in (33b), despite the absence of a definite article.

This is not, however, the same phenomenon of determiner omission as found with nouns governed by accusative-selecting prepositions. Plural nouns in dative/genitive or ablative case cannot take a definite article *in general*. It does not matter whether they are modified or unmodified, nor does the particular syntactic position in which they occur affect the distribution of the definite article. Rather, a definite article is simply morphologically incompatible with plural dative/genitive and ablative nouns. In contrast, there is no inherent conflict between accusative case and the presence of a definite article, as demonstrated in §2.3. Rather, it is in the specific environment of accusative-selecting prepositions that the accusative definite article is prohibited, and only when the noun has no modifiers.

It may be objected here that examples of definite determiner omission with nouns governed by accusative prepositions do not necessarily illustrate a grammatical rule. For example, English features a number of nouns that require the omission of the definite article when they are unmodified and governed by prepositions. 'School' is perhaps the most prominent example of this phenomenon: "He is in school right now," is more acceptable than "?He is in the school right now," although the latter is not strictly ungrammatical. While the absence of a definite article indicates that he is attending the institution of 'school,' the presence of the definite article emphasizes that he is in a specific school building. This set of nouns is a closed class and dialects of English show differences in which nouns require definite article omission. For example, English speakers in the United Kingdom customarily say "in hospital," as in "He was in hospital for 10 days," treating 'hospital' as an institutional noun like 'school.' In contrast, American English speakers use a definite article with this noun: "He was in the hospital for 10 days." This shows that the set of nouns that requires determiner omission in English is limited and idiomatic.

In contrast, determiner omissions with the specific set of accusative-selecting prepositions extends across every kind of noun in Albanian. Exceptions to determiner omission are not determined by the category of noun. Rather, they are determined by the syntax of the noun phrase (the presence of absence of nominal modifiers). Thus, both a concrete noun like *makinë* and an abstract noun like *dashuri* 'love' are restricted from taking a definite determiner when it is governed by an accusative preposition. This is despite the fact that both are required to take the definite article in other argument positions, as demonstrated in §2.2.

(34) a. Jam në **dashuri**
          I.am in love.ACC
          'I am in **love**.'

      b. *Jam në dashuri-në
          I.am in love.F.ACC.DEF
          (Intended) 'I am in love.'

      c. Pasion-i   në **një** dashuri fillon të   zhduket pas dymbëdhjetë
          passion-M.NOM.DEF in a   love.ACC begins SBJN disappear after twelve
          muajve.
          months.ABL
          'The passion in **a** love affair begins to disappear after twelve months.'

      d. Në **dashuri-në tënde** ke   çliruar   shpirt-in tim
          in love-F.ACC.DEF   your.F.ACC you.have liberated spirit-M.ACC.DEF my.M.ACC
          'In **your love** you have liberated my spirit'

      e. Dashuri-**a**   është e   mrekullueshme.
          love-F.NOM.DEF is   F.NOM.IND miraculous.F
          '**Love** is miraculous.'

Determiner omission is required for the unmodified noun *dashuri* 'love,' as illustrated by (34a-b). The expression of indefinite reference, as in (34c), still requires an indefinite article. And if the noun is modified, then a definite article is required, as in (34c), which is similar to what we would expect from the noun *dashuri* in other syntactic positions such as sentential subject, shown in (34d), reiterating example (14) that was analyzed in §2.2. It is, in particular, the conditions of

accusative-case selection for the preposition and lack of nominal modification for the noun that require determiner omission to indicate definite reference.

This condition applies not just to singular nouns, but also to plural nouns with definite reference.

(35) a. Të   gjithë kanë hipur   në **tavolina**.
          PL.NOM all   have climbed on **tables.F.ACC**

          'Everyone got up on **the** tables.'

      b. Hipën   në tavolina-**t**   e   qendr-ës   së   votim-it.
          they.climbed on tables-**F.ACC.DEF** F.ACC.PL.DEF center-F.DAT.DEF F.DAT.DEF vote-M.DAT.DEF

          'They climbed up on **the** tables of the polling station.'

Because the noun *tavolina* 'tables' in (35a) is unmodified, it cannot take a definite article, even though the reference is to a definite set of tables. In contrast, when the noun is modified as in (35b), the definite article is required.

The importance of modification as a syntactic condition can be illustrated with the complementizer *i cili* 'which.' This complementizer takes the form of an adjective, but with a null head that agrees in gender and number with the head of the relative clause.

(36) Burr-i   **i**   **cil-i**   humbi grua-n.
      man-M.NOM.DEF **M.NOM.DEF** **which-M.NOM.DEF** he.lost woman-F.ACC.DEF

      'The man who lost his wife.'

When the complementizer is governed by an accusative preposition such as *në*, it takes the accusative definite article. In effect, the modification condition applies to the whole of the complementizer phrase, even though the nominal head with which the complementizer agrees is null. Since the nominal head undergoes modification, the determiner phrase as a whole shows an accusative definite article.

(37) Cil-i   ka qenë koncert-i   më i   bukur   **në të**
      which-M.NOM.DEF has been concert-M.NOM.DEF more M.NOM.DEF beautiful.M **in M.ACC.IND**

      cil-in   ke   marrë pjesë?
      which-M.ACC.DEF you.have taken part.F.ACC

      'Which is the best concert you've taken part in?'

Thus, because of the condition of modification, the accusative definite article appears on the complementizer in the prepositional phrase *në të cil-in* 'in which.' This is despite the fact that the particle of concord, *të*, shows indefinite agreement with a null nominal head. It is the choice of whether or not the noun phrase contains a modifier of the head noun that determines whether determiner omission is triggered within the domain of an accusative preposition. If the nominal head is null, but the phrase contains a modifier, the determiner must be present and LDO does not apply.

Of note here is that the determiner is *not* absorbed into any element of the prepositional phrase. Contractions between prepositions and definite articles are common in many languages, such as French (preposition *à* + definite article *le* > *au*; Price 2003:20), German (preposition *zu* + definite article *der* > *zur*; Borgert and Nyhan 1976:211), or Greek (preposition *se* + definite article *tin* > *stin*; Bien et al. 1983:305-306). But here in Albanian, no morphological or phonological change

to the preposition occurs. Moreover, the definite article in Albanian does not precede the noun phrase and is never in direct phonological contact with the preposition. Instead, the definite article occurs in second position after the first declinable element of the noun phrase, as detailed in §2.1. Thus, the definite article cannot be contracted with the preposition. Rather, when LDO occurs, the definite article is simply omitted altogether.

Nouns that themselves denote locations offer the strongest evidence that determiner omission with accusative prepositions in Albanian is a grammatically conditioned phenomenon. As noted above in (19), proper nouns in Albanian are inherently definite, identifiable, and specific, and thus they require a definite article, as with *Maqedoni-a* 'Macedonia' in (38a). But as would be expected from the general condition of determiner omission with accusative-selecting prepositions, these location nouns appear in bare form after prepositions such as *në* 'in' in (38b). Only when they are modified does the definite article reappear, as in (38c)

(38) a. Maqedoni-**a**　　　　 është e　　 gatshme për zgjedhje-t.
　　　　 Macedonia-**F.NOM.DEF** is　　 F.NOM.IND ready.F　for elections-F.ACC.DEF
　　　　 'Macedonia is ready for the elections.'

　　　 b. Zgjedhje-t　　　　 në Maqedoni
　　　　 elections-F.NOM.DEF in **Macedonia.ACC**
　　　　 'The elections in Macedonia.'

　　　 c. Zgjedhje-t　　　　 në Maqedoni-**në**　　　 perëndimore
　　　　 elections-F.NOM.DEF in Macedonia-**F.ACC.DEF** western.F
　　　　 'The elections in western Macedonia.'

The preposition *në* 'in' is particularly good for illustrating determiner omission with accusative-selecting prepositions, since its meaning is entirely locative. That is, *në* only ever denotes a static location or a goal of movement. Other accusative prepositions, however, show more complicated behavior in regard to determiner omission. In (38a), it can be seen that the noun *zgjedhje* 'elections' takes the definite article *-t* in the phrase *për zgjedhje-t* 'for the elections.' The preposition *për* is an accusative-selecting preposition, which would lead to an expectation of determiner omission for *zgjedhje*. This illustrates that determiner omission is subject not just to a combination of syntactic conditions applying to the preposition and noun (case selection of the preposition and the presence of modifiers in the noun phrase), but also to a semantic condition that applies to the preposition. Specifically, determiner omission only occurs with accusative-selecting prepositions that designate a locative meaning.

The locative semantic condition on determiner omission with accusative prepositions

If accusative case selection were the only feature of the preposition that determined LDO, it would be expected that all accusative prepositions would trigger this pattern of determiner omission. But some prepositions fail to require determiner omission. The explanation for this lies in an additional restriction on LDO: the accusative preposition must have a locative meaning, designating a location or a goal of movement.

Most accusative-selecting prepositions have a primarily locative meaning, denoting either a static location or a goal of movement depending on the sentential context. These prepositions always trigger determiner omission with unmodified nouns.

(39) a. E mori   mbi shpinë(*-n)
          it he.took on   back(*-F.ACC.DEF)
          'He took it onto his back.'

      b. Po kërkonte mace-t   e   saja nën shtrat(*-in)
          PROG she.sought cats-F.ACC.DEF F.ACC.PL.DEF her.F.PL under bed(*-M.ACC.DEF)
          'She was looking for her cats under the bed.'

      c. Ishin   duke ecur nëpër pyll(*-in)
          they.were PROG walked through forest(*-M.ACC.DEF)
          'They were walking through the forest.'

      d. Mori kordhë-në   ndër   duar(*-t)
          he.took sword-F.ACC.DEF between hands(*-F.ACC.DEF)
          'He took the sword in his hands.'

Some of these accusative-selecting prepositions, however, also have extended meanings that do not directly refer to locative relations in space. For example, *mbi* 'on' also denotes the object of speech, thoughts, or feelings, equivalent to the English words 'about' or 'concerning' (Newmark et al. 1982:295). In this usage, an unmodified noun with definite reference *does* take the definite article.

(40) Ka   ardhur koh-a   për një debat   kombëtar **mbi arsim-in**.
      it.has come   time-F.NOM.DEF for a   debate.M.ACC national.M **on   education-M.ACC.DEF**
      'The time has come for a national debate concerning education.'

Only in this abstract usage of the accusative-selecting *mbi* 'on' does LDO fail to apply. Compare (40) with *mbi shpinë* 'on his back' in (39) above, where the locative meaning of *mbi* requires LDO with the noun *shpinë* 'back, spine.' The semantics of the preposition, not its lexical identity, determines whether LDO applies. Likewise, it must be noted that no lexical property of the noun *arsim* 'education' demands a definite article with an accusative-selecting preposition. When *arsim* is unmodified and governed by accusative-selecting *në* 'in,' it must appear bare, without a definite article. This is because *në* can only ever have a locative meaning.

(41) Investim-i   **në arsim**   është i   shenjtë.
      investment-M.DEF.NOM **in education.M.ACC** is   M.NOM.IND holy.M.SG
      'Investment in education is sacrosanct.'

The argument that LDO is not determined by the lexical properties of the preposition, but rather but its semantic interpretation in context, is supported by looking at other accusative-selection preposition, such as *me* 'with' and *pa* 'without.' These prepositions, which may designate a relationship of accompaniment (comitative) or the means by which an action is taken (instrumental), generally require a definite article, particularly with nouns that denote persons (Mëniku and Campos 2011:137).

(42) a. Bisedoj **me mësuese-n**.
          I.speak **with teacher-F.ACC.DEF**

          'I speak with the teacher.'

      b. *Bisedoj me mësuese.
          I.speak with teacher.F.ACC

          (Intended) 'I speak with the teacher.'

      c. Studentë-t   shkojnë në muze   **pa   mësuese-n**.
          students-M.NOM.DEF they.go to museum.F.ACC **without teacher-F.ACC.DEF**

          'The students go to the museum without their teacher.'

      d. *Studentë-t   shkojnë në muze   pa   mësuese.
          students-M.NOM.DEF they.go to museum.F.ACC without teacher.F.ACC

          (Intended) 'The students go to the museum without their teacher.'

                                                        (Mëniku and Campos 2011:137)

With inanimate nouns, Albanian shows a split between instrumental and comitiative meanings for the preposition *me*. While the comitative meaning expressing accompaniment, as in (43a), requires a definite article to express definite reference, the instrumental meaning expressing the means by which an action is performed allows for determiner omission, as in (43b) (Mëniku and Campos 2011:138).

(43) a. U   nis   **me top-in**
          REFL set.off **with ball-M.ACC.DEF**

          'He set off with the ball.'

      b. Po luan   **me top**
          PROG he.plays **with ball.M.ACC**

          'He is playing with the ball.'

                                                        (Mëniku and Campos 2011:138)

This is particularly clear for nouns that define the means by which someone travels to a location. In this case, the definite article is normally omitted, even if the noun has definite reference.

(44) Mund të shkosh në Durrës   **me tren**,   por është më e   shpejtë
      can   SBJN you.go to Durrës.M.ACC **with train.M.ACC**, but is   more F.NOM.IND fast

      me autobus.
      with bus.M.ACC

      'You can go to Durrës with the train, but it is faster with the bus.'

This split between semantic conditions, where the same accusative preposition triggers different behavior, also functions for the preposition *për* 'for.' This preposition is primarily used to indicate the recipient or benefactor of an action. In this case, the definite article is required for definite reference, even for unmodified nouns.

(45) a. Kjo   është **për Petrit-in**.
          this.F.NOM is   **for Peter-M.ACC.DEF**

          'This is for Peter.'

b. *Kjo   është për Petrit.
    this.F.NOM. is   for Peter.M.ACC

    (Intended) 'This is for Peter.'

Even though the preposition *për* 'for' in (45) selects the accusative case and even though *Petrit* 'Peter' is unmodified, this proper noun must take a definite article. Determiner omission fails to apply. Other non-locative uses of *për* also require a definite article for definite reference, as in (46a), where *për* is best translated with the English preposition 'about.' But when *për* is used to denote a goal of movement, as in (46b), the definite article is ungrammatical and the noun must appear as a bare accusative.

(46)   a. Kadare   shkroi   një novelë   të   famshme **për**
             Kadare.F.NOM.DEF he.wrote a   novel.F.ACC F.ACC.IND famous.F **for**
             Gjirokastr-ën.
             Gjirokastra-F.ACC.DEF

             'Kadare wrote a famous novel about Gjirokastra.'

    b. Autobus-i   **për Gjirokastër**   niset   për një orë.
       bus-M.NOM.DEF **for Gjirokastra.F.ACC** sets.off in   an hour.F.ACC

       'The bus for Gjirokastra leaves in an hour.'

In both example sentences in (46), the preposition *për* and the noun denoting a city, *Gjirokastra*, are used. But when *për* has a locative meaning, denoting that Gjirokastra is the endpoint for movement, the definite article is omitted. Use of a definite article in (46b) would be ungrammatical. The consistent differentiation between locative and non-locative meanings in prepositions such as *për* 'for' and *mbi* 'on' indicates that locative semantics are an additional condition on determiner omission. In some cases, the presence or absence of a determiner helps distinguish the intended meaning of the preposition. LDO functions to narrow the range of meaning of a given accusative-selecting preposition to only its locative aspect. This applies even to the accusative-selecting prepositions *me* 'with' and *pa* 'without,' which show a distinction between comitative meanings, where the article is required, and instrumental meanings, where it is omitted. The contextually determined semantics of the preposition, as well as its case selection, create the conditions for determiner omission to apply to noun phrases without modifiers.

2.4   Locative determiner omission in Albanian: Distinctive, but shared

From this data, some conclusions about determiner omission in Albanian can be drawn. Determiner omission occurs with unmodified nouns within the domain of accusative-selecting prepositions. For these prepositions, the absence of a definite determiner on the noun signals definite reference. This pattern in which zero-marking yields a definite interpretation differs sharply from the morphological marking of definite reference with a phrasally affixed definite determiner in other grammatical domains in Albanian.

There is a combination of three conditions that license LDO in Albanian. It is subject to syntactic limitations. In the presence of any nominal modifier, a determiner must appear to mark definite

reference, regardless of the preposition that governs the noun. It is also subject to a case restriction. Only accusative-selecting prepositions trigger determiner omission. Finally, it is subject to a semantic restriction. From among these prepositions, only those that carry a locative meaning, denoting static location or the goal of movement, require determiner omission with unmodified nouns. For example, the prepositions *me* 'with' and *pa* 'without' are exceptions among the accusative-selecting prepositions in that they distinguish between comitative and instrumental meanings with determiner omission. If these three syntactic, morphological, and semantic conditions are satisfied, LDO applies regardless of the noun or the preposition involved. Thus, LDO applies to all lexical items that fall within the domain of prepositions satisfying the grammatical and semantic conditions. LDO is a grammatical pattern in Albanian, not a case of exceptional marking of definiteness for a closed class of lexical items.

What is striking about this phenomenon is not that it is irregular. As shown by the analysis in this chapter, the conditions that trigger LDO are consistent and can be applied to wide variety of noun constituents. Rather, what distinguishes LDO and demands further examination is how regular it is in comparison to the rest of the Albanian system of expressing definite reference. It is an exceptional but consistent pattern of coding definite reference. LDO, with its particular configuration of determining factors, appears to be very rare among the world's languages.

The semantic category of definiteness is encoded with a variety of grammatical means across languages. Often, an independent word, a particle, or an inflection will indicate the definiteness of a nominal constituent. Verbs can also encode definiteness by showing variable agreement with subjects, objects or other constituents. What is rarer is for the *absence* of a grammatical feature to encode definiteness, particularly when the feature's presence otherwise indicates that a nominal referent is definite. The expression of definiteness with a determiner is common and widely distributed among the world's languages. Given a noun with a definite referent, many of these languages allow the determiner to be omitted under certain conditions. The particular pattern of determiner omission seen in Albanian, however, is a typological rarity due to its consistent application without lexical restrictions on the noun, across the whole range of nominal constituents.

That the distribution of the definite determiner in Albanian is peculiar was noted by Lyons (1999) in his typological survey of the definiteness. He asserts that while it is not unusual for languages with definite determiners to omit them in prepositional phrases, these phrases tend to be fixed. Phrases such as English *at school* or Spanish *en casa* 'at home' exhibit determiner omission for nouns that would otherwise require a determiner to express definiteness. But these phrases fall within a closed class of lexicalized expressions, as the unacceptability of **at bank* to express 'at the bank' or **on desk* to express 'on the desk' demonstrates. In addition, a language may have a preposition or set of prepositions that never allow definite articles, no matter the properties of the noun phrase. For example, the preposition *en* 'in, on' in French generally does not allow the noun it governs to take a definite article, even if the noun has definite reference (L'Hillier 1999:373, Price 2003:509-512). This behavior, however, is specific to the preposition *en*, and allows for idiomatic exceptions. Determiner omission in prepositional phrases, when it occurs, is a lexical phenomenon in most languages. Omission of determiners normally depends on specific nouns or specific prepositions, and it is limited to particular preposition/noun combinations. This contrasts with the omission of determiners in Albanian prepositional phrases, which is subject to strict grammatical and semantic conditions but applies to the full range of nouns, including proper nouns with an inherently unique referent. This phenomenon is "rather mysterious," as Lyons remarks (1999:51), because it appears to be grammatically, not lexically motivated. Yet Albanian

otherwise makes use of determiners to express the semantic category of definiteness. Only within prepositional phrases, particularly those with locative meaning, is the pattern reversed. This yields LDO as the zero-marking expression of definiteness.

Although I describe LDO as a typological rarity, it is not unique to Albanian. Similar grammatical patterns are also shared by languages with speaker communities that intermingle with the Albanian speaker community. Determiner omission in locative prepositional phrases must be analyzed not only as a distinct feature of Albanian, but also as an areal feature shared with Macedonian and Balkan Romance languages. The ways in which this feature is realized in these other languages, with which Albanian has had extensive contact over time, illuminate how this peculiar way of expressing definiteness developed in Albanian itself, and why it is distinct for the languages of this region as a whole.

Chapter 3

Locative Determiner Omission in Balkan Romance

I continue my analysis of LDO as a Balkan areal feature by demonstrating that it is also present in Balkan Romance languages. The form of LDO found in Balkan Romance is remarkably similar to that of Albanian. Balkan Romance nominal syntax parallels Albanian in notable aspects, which makes comparison relatively easy. In particular, Balkan Romance also features nominal case that interacts with a post-posed definite article. The triggering conditions for determiner omission—accusative-case selection in the preposition, locative semantics, and an unmodified noun—can be observed to trigger determiner omission in Balkan Romance as well. This applies across all languages of the Balkan Romance branch, including (Daco-)Romanian,[1] Aromanian, Megleno-Romanian, and Istro-Romanian. The approximate locations of the Balkan Romance speech communities are indicated in Figure 3.1, based on the findings of Harris (1988:22-24).

LDO is a pattern that repeats itself in all these Balkan languages with remarkable consistency. In §3.1, I begin by analyzing the similarities between LDO in Romanian and Albanian. In §3.2, I then discuss LDO in Aromanian, the other Balkan Romance language for which there is still a significant speaker community. This will form a basis for comparison with the more poorly documented Romance languages, Istro-Romanian and Megleno-Romanian, which also show LDO and which will be addressed in §5.1 and. It can be concluded that LDO is a feature of Balkan Romance languages that bears striking similarity to the same phenomenon of determiner omission in Albanian. Showing that the unique conditions of this pattern for coding definite reference apply to multiple languages in contact with one another will help to establish it as an areal feature and

---

[1]Due to some ambiguity in the usage of *Romanian* as a descriptor for a single codified language—spoken in the modern state of Romania—or as a term for all forms of Romance spoken in the Balkans—some of which have their own independent codifications as standard languages for literature and education—the prefix Daco- is used here to refer specifically to Romanian as spoken in Romania (see Ivănescu (1980:30-46) for a summary of the debate). In the rest of this dissertation, Daco-Romanian will be referred to simply as Romanian. 'Romanian' here also refers to the form of Balkan Romance spoken in Moldova, where it may also be called *Moldovan* due to political circumstances in the separation of the Republic of Moldova from the Republic of Romania. The literary standard of Daco-Romanian as spoken in Moldova does not differ in any notable respect from Romanian as spoken in Romania. In contrast, Istro-Romanian, Megleno-Romanian, and Aromanian differ from standard Daco-Romanian in significant respects and may not be mutually intelligible for some speakers. These languages share a common origin in the Late Latin brought to the Balkans by Roman conquest and I refer to them collectively as Balkan Romance to distinguish them from Latin-descendent languages spoken outside the Balkans.

Figure 3.1: Map of the Balkan Romance languages

provides the basis for looking more deeply into its pattern of spread through space and time.

3.1 Locative determiner omission in Romanian

An important task in my analysis is to establish that LDO is present in other Balkan languages besides Albanian, in a form sufficiently similar that the phenomenon invites an explanation based on language contact. Romanian, deriving from Late Latin spoken in the Balkans, shows a pattern of determiner omission in prepositional phrases that strongly resembles what is found in Albanian. Romanian and Albanian share a similar syntax and case morphology in their noun phrases, which makes comparison between them easy. This section will provide evidence that the conditions for determiner omission in Romanian are like those of Albanian. Unmodified nouns within the domain of triggering prepositions express definiteness with the absence of a determiner. In contrast,

modified nouns require the presence of a definite determiner to express definiteness. All the prepositions triggering determiner omission in Romanian are primarily locative in meaning and select accusative case. These are the same semantic and grammatical conditions that trigger determiner omission in Albanian. On the basis of this evidence, it is appropriate to consider LDO a single phenomenon shared by Balkan Romance and Albanian.

Like Albanian, Balkan Romance normally indicates the definiteness of a nominal constituent by means of a definite determiner. The syntax of the noun phrase and the distribution of definite determiners in Balkan Romance mirrors that of Albanian. Both languages express definiteness through determiners post-posed to the first nominal or adjectival element in the noun phrase. As an illustration of this phenomenon in Balkan Romance as a whole, the syntactic distribution of determiners in Romanian can be seen in the examples below.

ROMANIAN

(47)   a. un prieten   bun
           a   friend.M.NOM good.M.SG
           'a good friend'

       b. prieten-ul   bun
           friend-M.NOM.DEF good.M.SG
           'the good friend'

       c. foarte bun-ul   meu   prieten
           very   good-M.NOM.DEF my.M.SG friend.M.NOM
           'my very good friend'

       d. *foarte-ul   meu   bun   prieten
           very-M.NOM.DEF my.M.SG good.M.SG friend.M.NOM
           'my very good friend'

In (47a), the indefinite determiner *un* occurs before the noun phrase as a whole. Outside of prepositional phrases, a bare noun occurring either with an indefinite determiner or alone signals an indefinite referent. In (47b), the definite determiner *-ul* occurs as a suffix to the first nominal element in the noun phrase. The determiner must attach to a noun or an adjective. Generally nouns precede adjectives, but the reverse order is also licit. Example (47c) shows that the definite determiner attaches as a suffix to an adjectival element, since the adjective occurs first in the noun phrase. Possessive pronouns follow the definite determiner, thus leading *meu* 'my' to occur after *-ul*, the definite article. The determiner cannot attach to an adverb, as demonstrated by (47d), since the adverb is not an inflecting element of the noun phrase. Thus, the surface syntax[2] of determiners in Romanian (and Balkan Romance broadly) abides by the same patterns that apply to determiners in Albanian, as analyzed in §2.1:
Indefinite DP
D + NP + (AP)
D + AP + (NP)

---

[2]Due to the complexity of the potential derivation, only the surface syntactic outcomes are illustrated here, without making any commitments to underlying representation. While the syntactic derivation post-posed determiners in Albanian and Balkan Romance is a subject of considerable interest, it is not relevant to the analysis presented here, which depends only on surface order of constituents.

Definite DP
NP + D + (AP)
AP + D + (NP)

As in Albanian, case in Romanian is expressed synthetically on the determiner, and case interacts with prepositional phrases to determine the contexts in which determiners may be omitted. The expression of case in Romanian, and Balkan Romance in general, is thus relevant to the understanding of determiner omission in these languages. Case in Romanian is more restricted than in Albanian: pronouns show a distinction between nominative, accusative, dative, and genitive, whereas nouns only distinguish between nominative/accusative and dative/genitive.[3] Table 3.1 gives the simple prepositions and their case selection. Balkan Romance prepositions are often compounded; the case selection follows the requirements of the final preposition in the sequence. The key distinction for the purposes of determiner omission is between prepositions that select dative/genitive and those that select accusative.

Table 3.1: Prepositions and case in Romanian

Accusative	Dative	Genitive
cu 'with'	*conform* 'according to'	*asupra* 'regarding, on'
către 'toward'	*datorită* 'due to'	*contra* 'against'
de 'of'	*grație* 'thanks to'	*dealungul* 'along'
despre 'about'	*mulțumită* 'thanks to'	*deasupra* 'above'
dinspre 'from'		
dintre 'among'		
dintru 'of, from'		
după 'behind, by'		
fără 'without'		
în 'in'		
la 'at, to'		
lângă 'beside'		
pe 'on, ACC'		
pentru 'for'		
peste 'over'		
prin 'by means of'		
printre 'among'		
până 'until, up to'		
spre 'to'		
sub 'below'		
în 'in'		
înspre 'into'		

---

[3]In the following examples, I will mark nouns according to the case that their dominating constituent selects for, even though nouns do not distinguish between nominative and accusative, or between genitive and dative, in their morphology.

Prepositions that select dative/genitive require a definite article for nouns with definite reference. This would be expected based on Romanian syntax in other argument positions. This also reflects the pattern found in Albanian, where prepositions selecting the ablative case require a definite article to express definite reference. In (48), this is shown by *deasupra* 'above,' which selects a noun in the genitive case that takes a definite article, *ocean-ului* 'of the ocean,' without any modifiers.

(48) Avion-ul   se află   deasupra ocean-ului.
      plane-M.NOM.DEF REFL located over   ocean-M.GEN.DEF
      'The plane was located over the ocean.'

The behavior of dative/genitive-selecting prepositions in Romanian contrasts with accusative-selecting prepositions. With the exception of *cu* 'with,' what is common to the accusative-selecting set of prepositions is the fact that their meaning makes reference to a location, either in a static or a dynamic relation. These locative, accusative-selecting prepositions impose different restrictions on determiner distribution depending on whether the noun is modified or unmodified.

ROMANIAN

(49) a. la o   universitate
          at F.ACC.INDEF university.F.ACC
          'at a university'

      b. la universitate
          at university.F.ACC
          'at the university'

      c. *la universitate-a
          at university-F.ACC.DEF
          (intended) 'at the university'

      d. la universitate-a   tehnică   din Cluj-Napoca
          at university-F.ACC.DEF technical.F.SG from Cluj-Napoca
          'at the Cluj-Napoca Technical University'

      e. *la universitate   technică   din Cluj-Napoca
          at university.F.ACC technical.F.SG from Cluj-Napoca
          (intended) 'at the Cluj-Napoca Technical University'

Unmodified nouns governed by accusative-selecting prepositions appear either with an indefinite article, as in (49a), or bare, without any article, as in (49b). An unmodified noun cannot appear with a definite article in the domain of accusative-selecting prepositions, as illustrated by (49c) (Ciobanu and Nedelcu 2005:615). The definiteness of unmodified nouns is expressed by the omission of a definite article; absence of an article indicates a definite interpretation. Thus, it is the fact that *universitate* 'university' in (49b) is in bare form, without any determiner, that indicates that a definite, specific university is meant (Nedelcu 2013:460). A definite determiner is not grammatical in this context, as in (49c). When a noun is modified, the use of a definite article to express definiteness becomes grammatical and necessary, as in (49d). Omission of the definite article does not correspond to a definite interpretation as in (49b) with *la universitate* 'at the unversity.' Instead, determiner omission with a modified noun is ungrammatical, as in (49e). The modification

condition on determiner omission in Romanian conforms exactly to the condition that applies in Albanian. Noun phrase modifiers in Romanian that trigger the use of the definite determiner include adjectives, genitive nouns, possessive pronouns, prepositional phrases, and relative clauses, the same set that control the appearance of determiner omission in Albanian.

Among accusative-selecting prepositions, the preposition *cu* 'with' is exceptional. *Cu* 'with' requires the noun it governs to appear with either an indefinite or a definite article—bare nouns are not permitted (Boretzky 1968:134). This is regardless of whether the relation described by the preposition is one of instrument, as in (50a-b) or association as in (50c-d). Thus, *cu* 'with' does not trigger LDO in Romanian.

ROMANIAN

(50) a. A   intrat   în coliziune cu   un   tren
          has entered in collision.F with M.ACC.INDEF train.M.ACC

          'It collided with a train.'

     b. Plec   cu   tren-ul
          depart with train-M.ACC.DEF

          'I will depart by train.'

     c. Am   un   interviu   cu   un   profesor
          I.have M.ACC.INDEF interview.M with M.ACC.INDEF professor.M

          'I have an interview with a professor'

     d. Am   vorbit cu   profesor-ul
          I.have spoken with professor-M.ACC.DEF

          'I spoke with the professor.'

Prepositions that select the genitive or dative case require both unmodified and modified nouns to express definiteness with an overt definite article (Ciobanu and Nedelcu 2005:617). These prepositions do not trigger determiner omission. In (51), *probleme-lor* 'of the problems' is in the genitive case and takes a definite article despite occurring unmodified.

ROMANIAN

(51) a. o   discuție   asupra   probleme-lor
          F.NOM.INDEF discussion.F.NOM regarding problems-F.GEN.DEF

          'a discussion of the problems'

     b. *o   discuție   asupra   probleme
          F.NOM.INDEF discussion.F.NOM regarding problems.F.GEN

          (intended) 'a discussion of the problems'

I refer to the phenomenon of determiner omission in the domain of accusative-selecting prepositions as *locative* determiner omission due to the meaning of the majority of the omission-triggering prepositions. However, not all the prepositions that trigger determiner omission in Balkan Romance are primarily locative in meaning. *Pentru* 'for' can have a locative meaning, but usually designates recipient; *pe* can mean 'on' but primarily marks animate direct objects of a verb; *de* is a preposition indicating possession and must be used in combination with *la* 'at' (> *de la* 'from') or *în* 'in' (> *din* 'from') to indicate origin. All of these prepositions select accusative and require

the omission of a definite article with unmodified nouns. All of these prepositions at one point had a locative meaning or secondarily express a locative meaning (Kurth 1904:2-3). This serves to maintain the association between locative semantics and the phenomenon of determiner omission.

The parallels between Romanian and Albanian are striking: the same set of grammatical and semantic conditions trigger the omission of determiners where their presence would be expected to indicate definite reference in both languages. The similarity justifies referring to this phenomenon as LDO in both languages. In addition, the parallels extend beyond just Romanian, into the broader set of Balkan Romance languages.

3.2   Locative determiner omission in Aromanian

The examples of determiner omission in Balkan Romance cited thus far have been drawn from Romanian, but the phenomenon is not limited to only this descendent of Balkan Late Latin. The other Balkan Romance languages also show determiner omission with similar sets of prepositions. Aromanian, a language which is closely related to Romanian but nevertheless distinct from it, does not allow definite articles with bare nouns governed by locative prepositions. It also shows parallels with Albanian in its semantic restrictions on determiner omission, wherein one preposition distinguishes between two different meanings on the basis of whether the meanings in question trigger determiner omission. The following examples demonstrate these parallels.

AROMANIAN

(52) a. ncalicã   pi cal
           he.mounted on horse.M.ACC
           'he mounted the horse'   (Liaku-Anovska 2007:74)

      b. s-lj-u   badz   cheale-a   pi cal-u   a tãu
           SUB-to.him-it you.put skin-F.ACC.DEF on horse-M.ACC.DEF PRT to.you
           'put leather over your horse'   (Liaku-Anovska 2007:73)

      c. *ncalicã   pi cal-u
           he.mounted on horse-M.ACC.DEF
           (intended) 'he mounted the horse'

These examples were confirmed with an Aromanian consultant. In (52a), the noun *cal* 'horse' appears indefinite, but has a definite meaning (corresponding to a noun with a definite article in the author's Macedonian translation: *javna koň-ot* 'he mounted the horse'). This contrasts with (52b), where the noun is modified by the possessive pronominal phrase *a tãu* 'your' and requires a definite article to indicate definite reference for the noun, thus *cal-u* 'the horse.' It is ungrammatical, however, to include a definite article with an unmodified noun governed by an accusative-selecting preposition is ungrammatical, as in (52c).

It would be expected that LDO in Aromanian would distinguish the preposition meaning 'with' from other accusative selecting prepositions, such that 'with' does not trigger determiner omission. Indeed, the Aromanian preposition *cu* 'with' shows exceptional behavior and allows unmodified nouns to take the definite article, as in (53).

AROMANIAN

(53) Ãl stãvrãsi atsel oarfãn-lu   om   cu cal-u
      him he.met that.M poor-M.SG.ACC.DEF man.M with horse-M.ACC.DEF

      'He met that poor man with the horse'   (Liaku-Anovska 2007:73)

This pattern also applies to other lexical, non-animate lexical items such as *mãn-a* 'the hand' in the following examples.

AROMANIAN

(54) nu cu mãn-a
      not with hand-F.ACC.DEF

      'not with your hand'   (Caciuperi 1985:103)

(55) stau pri scamnu
      I.sit on chair.M.ACC

      'I sit in the chair'   (Caciuperi 1985:104)

(56) stau pri scamnu-l   atsel mari-le
      I.sit in chair-M.ACC.DEF PRT big-M.DEF

      'I sit in the big chair'   (Caciuperi 1985:104)

In (54), *mãn-a* 'the hand' appears with a definite article, triggered by the governing preposition *cu* 'with,' which requires its subordinate noun to take the article in order to express definite reference. The behavior of *cu* 'with' contrasts with *pri* 'on,' which does not allow a definite article for unmodified nouns as in (55), but does allow it for modified nouns as in (56).

As a point worth noting for later analysis, Aromanian treats the preposition *cu* differently from Romanian. While Romanian requires the use of a definite article for uses of *cu* that denote a means of movement, as with *cu tren-ul* 'with the train' in (50) above, Aromanian allows the preposition *cu* to govern bare nouns, without a definite article, when a means of transportation is indicated. In this regard, Aromanian shows exactly the same behavior as Albanian in its treatment of the preposition 'with', as in (57 below.

(57) AROMANIAN
   a. Cu ți va s-calitorešt, **cu tren**,   ma **cu autobus**?
      with what will SBJN-you.travel **with train.M.ACC** or **with bus.M.ACC**

      'How will you travel, **by train** or **by bus**?'   (Ballamaci 2010:32)
      ALBANIAN
   b. Me çfarë do të udhëtosh, **me tren**   apo **me autobus**?
      with what will SBJN you.travel **with train.M.ACC** or **with bus.M.ACC**

      'How will you travel, **by train** or **by bus**?'   (Ballamaci 2010:32)

Distinctions between locative and non-locative uses of accusative-selecting prepositions extend further through Aromanian than they do through Romanian. As an additional example, while Romanian *pentru* 'for' requires LDO as an accusative-selecting preposition, whether it denotes a locative or non-locative meaning, the Aromanian equivalents *ti* and *trã* 'for' allow a definite article with an unmodified noun when the preposition has a non-locative meaning, as in (58)

AROMANIAN

(58) S-fac   emisiuni   radiodifusivi pi aromăneaşti şă **ti**
      REFL-they.make broadcasts.F.ACC radio.F.PL   in Aromanian and **for**

      aromâni-l'i   s-fac   şă Congresuri.
      Aromanians-M.ACC.DEF REFL-they.make and conventions

      'Radio shows are broadcast in Aromanian and conventions are held **for the Aromanians
      as well.**   (Ballamaci 2010:120)

Another preposition that selects accusative case, *la* 'to, in, at,' also distinguishes between a meaning that triggers determiner omission and one that does not. In an interesting parallel, these two meanings are expressed by two different prepositions in Albanian, one of which, *te* 'at,' requires nominative case and does not trigger determiner omission. Comparing the behavior of these two locative prepositions demonstrates how a lexical contrast in Albanian is mirrored by a structural contrast in Aromanian and highlights the importance of accusative case selection as one of the defining conditions of LDO.

Namely, the Aromanian preposition *la* 'to, in, at' normally triggers determiner omission. It can take an unmodified noun with a definite article. However, if the noun refers to a person and the reference is to a typical location occupied by that person, then the definite article is permitted with an unmodified noun (this use of *la* that corresponds with the meaning of the French preposition *chez* 'at, regarding,' or, within the Balkan linguistic area, to the meaning of the Macedonian preposition *kaj* 'at'). This is distinct from the simple locative use of the preposition, where a definite article is not permitted. Example (59) shows both uses in the same sentence, with *la murar-lu* 'to the miller's place' showing the definite article that is absent from *la moarã* 'at the mill' (Liaku-Anovska 2007:60).

AROMANIAN

(59) S-dusirã   la murar-lu,   la moarã
      REFL-they.went to miller-M.ACC.DEF in mill.F

      'They went to the miller's place, to the mill'   (Liaku-Anovska 2007:60)

This parallels Albanian in semantics, as in (60), where the preposition designating a person's typical place is *te* and takes nouns in the nominative case, with a definite article. Thus in both Aromanian and Albanian, this particular semantic domain covered by the prepositions *la* (Balkan Romance) and *te* (Albanian) shows exceptional behavior that distinguishes it from the general domain of locative prepositions selecting accusative case. Compare the presence of the definite article with *te* in *te mullixhi-u* 'to the miller's place' with the lack of the definite article with *në* in *në mulli* 'to the mill' which is a consultant's translation into Albanian of (59). Lack of determiner omission marks out this particular meaning as separate from other locative meanings.

ALBANIAN

(60) Vajtën   te mullixhi-u,   në mulli
      they.went to miller-M.NOM.DEF in mill.M.ACC

      'They went to the miller's place, to the mill.'

Aromanian thus shows a pattern of LDO that hews even more closely to the pattern observed in Albanian. These two members of Balkan Romance share with Albanian a pattern of accusative case selection and exceptional behavior for the preposition 'with' along with the pattern of unmodified

nouns showing definite reference despite the lack of a definite article. Aromanian, in addition, shares some of the semantic distinctions among prepositional meanings that determine whether LDO applies or does not apply to a given noun. As noted by Lyons (1999:51), this is a feature exclusive to Balkan Romance, as opposed to other branches of the Romance family. The lack of LDO in other Romance languages outside the Balkans makes it more likely that this convergence between Albanian and Balkan Romance is a result of language contact. But if so, this raises the question of whether one language possessed the structure before another, or whether they developed it through mutual influence without either one being the sole source.

The best way of answering this question is to include data from another source: Balkan Slavic languages. These languages show great similarities with Balkan Romance and Albanian in significant aspects of the grammatical systems. The parallels are not perfect, however. Only one of the Balkan Slavic languages, Macedonian, shows anything similar to LDO. In contrast, both Bulgarian and southeast Serbian Torlak dialects, though very closely related, lack LDO. Further data and comparisons between the Balkan Slavic languages help situate the distribution of LDO and show that it is a shared feature of Balkan languages, but not universally shared.

Chapter 4

Locative Determiner Omission in Balkan Slavic

Balkan Slavic, as a classification, is defined by the effect of Balkan language contact on the broader branch of South Slavic languages. West South Slavic, constituted by Bosnian, Serbian, and Croatian along with Slovenian, mostly lacks such common Balkan grammatical features as the postpositive definite article, replacement of synthetic case with analytical prepositional constructions, or loss of the infinitive, and thus is not classified as Balkan Slavic. East South Slavic languages, constituted by Macedonian and Bulgarian, were far more affected by processes of convergence that also incorporated Albanian and Balkan Romance. The separation between East and West South Slavic does not only depend on Balkan language contact, however. Many of the isoglosses that separate East and West South Slavic are relatively old and run north to south before fanning out over the territory of modern Macedonian, resulting in considerable dialect diversity within Macedonia and a slow transition into the southeastern Serbian (Alexander 2000a:8-9), including the Balkanized Serbian dialect of Torlak. Innovations of the Balkan linguistic area spread northwards and westward to include Torlak dialects spoken in southeastern Serbia and southern Kosovo (Friedman 2002/3). Thus, these dialects must be included for consideration in Balkan Slavic, and the notion of distinct boundaries between the standardized languages of Macedonian, Bulgarian, or Serbian is less useful for analysis when discussing the spread of Balkan grammatical features

   Unlike Balkan Romance languages, Balkan Slavic languages do not all show LDO. This chapter begins by addressing LDO in Macedonian in §4.1. In §4.1, I discuss the distribution of articles in Macedonian in order to establish a baseline for comparison between the standard use of the definite article and the exceptional behavior of the definite article in locative prepositional phrases. Macedonian noun phrase structure shares many properties with Albanian and Balkan Romance, including a post-posed definite article that attaches to the first declinable constituent of the noun phrase. Macedonian also shares the goal/location semantic merger in its prepositions. These are important preconditions for the phenomenon of LDO, as analyzed in chapter 2 and chapter 3. Macedonian, however, lacks the consistent use of an indefinite article, which complicates the interpretation of bare nouns. This contrasts with Albanian and Balkan Romance, which consistently use indefinite articles. Moreover, Macedonian does not show nominal case, which means that there is no distinction in the case selection of prepositions. Thus the accusative case selection condition of LDO in Balkan Romance and Albanian cannot apply here. Given the paucity of preconditions, and the differences in Macedonian noun phrase structure, we find that the standard variety consis-

tently attributes an indefinite reading to bare nouns, even in prepositional phrases. Unexpectedly, however, colloquial Macedonian regularly assigns definite reference to nouns in locative prepositional phrases, particularly when they are unmodified. In §4.1, I review the distribution of LDO in Macedonian. This phenomenon is especially common in dialects spoken by communities in close contact with Aromanian speakers. Thus, although LDO is not fully grammatically integrated into Macedonian, and does not occur in all speech forms and at all registers, some forms of Macedonian nonetheless show a tendency to repeat the same exceptional omission of determiners as in Albanian and Balkan Romance. The particular domains where this determiner omission occurs in Macedonian show parallels with the constellation of conditioning factors that trigger LDO in Albanian and Balkan Romance. In §4.3, I contrast such cases of determiner omission in Macedonian with their absence in Bulgarian. Thus it can be seen that Macedonian more greatly resembles Albanian or Aromanian in its treatment of definiteness in prepositional phrases, since Bulgarian lacks any grammatical pattern like LDO. This selective distribution of LDO in Balkan Slavic, in turn, is a key factor for the explanation of how LDO could have emerged as a contact-induced innovation among those Balkan languages that show it.

4.1 Locative determiner omission in Macedonian

The semantic distribution of articles in Macedonian

Standard Macedonian nominal syntax differs somewhat from that of Albanian and Balkan Romance. While the latter two language groups primarily position adjectives and possessive pronouns after the noun, these modifiers primarily occur before the noun in Macedonian. The rules for the distribution of the definite article, however, are the same. The article attaches to the first constituent of the noun phrase that expresses gender and number. Adjectives, nouns, or possessive pronouns fulfill these conditions and so may take the post-posed article, as in (61a-c), but adverbs may not, as seen by a comparison of (61d) with (61e).

MACEDONIAN

(61) a. idea-ta
          idea-F.SG.DEF
          'the idea'

      b. interesna-ta   idea
          interesting-F.SG.DEF idea.F.SG
          'the interesting idea'

      c. moja-ta   idea
          my-F.SG.DEF idea.F.SG
          'my idea'

      d. mnogu interesna-ta   idea
          very interesting-F.SG.DEF idea.F.SG
          'the very interesting idea'

      e. *mnogu-ta   interesna   idea
          very-F.SG.DEF interesting.F.SG idea.F.SG

(intended) 'the very interesting idea'

Macedonian allows both bare nouns and nouns with a pre-posed indefinite article that inflect for gender, as in (62a). Unlike Albanian (see §2.2), the Macedonian bare noun is allowed in predicate constructions without an indefinite article, as in (62b) where the indefinite article is disallowed. With possessive modifiers, the indefinite article is optional, as in (62).

MACEDONIAN

(62) a. (eden) dobar   student
          IND.M.SG good.M.SG student.M.SG
          'a good student'

      b. Toj beše (*eden) dobar   student
          he was IND.M.SG good.M.SG student.M.SG
          'He was a good student.'

      c. Toj beše (eden) moj   prijatel
          he was IND.M.SG my.M.SG friend.M.SG
          'He was my friend.'

Friedman (2003) observes, citing Weiss (1996), that Macedonian allows specific indefinite, non-specific indefinite, and kind-level readings for bare nouns. The indefinite article imposes a referentiality restriction such that the noun must have either a specific or a generic reading. Thus in the sentences in (63), where context requires a non-specific interpretation, the indefinite article is disallowed (Weiss 1996:436).

MACEDONIAN

(63) a. Ima   li kaj vas (*eden) lekar?
          is.there Q at you IND.M.SG doctor.M.SG
          'Is there a doctor at your place?'

      b. Vikni (*eden) lekar!
          call   IND.M.SG doctor.M.SG
          'Call a doctor!'

This results in the following semantic distribution of definites and indefinites in Macedonian shown in Table 4.1.

Table 4.1: The distribution of definites, indefinites, and bare nouns in Macedonian

	Specific	Generic	Lexically ungoverned	Predicational
Definite singular (*lekar-ot*)	✓	✓	✓	Yes
Indefinite singular (*eden lekar*)	✓	✓	✓	No
Definite plural (*lekari-te*)	✓	✓	✓	Yes
Bare singular (*lekar*)	✓	✓	✓	Yes
Bare plural (*lekari*)	✓	✓	✗	Yes

The distribution of determiner omission in Macedonian

While the evidence for LDO is very clear in Albanian and Balkan Romance, Macedonian shows a less grammatically integrated version of the phenomenon. Patterns of determiner omission that resemble LDO in these other languages are identifiable in Macedonian, but they depend on factors such as conversational register and dialect. An analysis of LDO in Macedonian needs to take these features into account, particularly in order to make a comparison with other forms of Balkan Slavic where LDO is absent.

The post-posed definite article in standard Macedonian expresses definiteness in both noun phrase arguments and in prepositional phrases, without regard to the modificational status of the noun phrase. Thus, at first glance, Macedonian appears to lack LDO as seen in Balkan Romance and Albanian. In (64), the noun phrase takes a definite article -*ta* whether the noun is unmodified (*zgrada-ta* 'the building') or modified (*crvena-ta zgrada* 'the red building'). This is despite the presence of the noun phrase within a locative prepositional phrase headed by *vo* 'in.' Modification of the noun here does not condition any alternation in the expression of definiteness.

MACEDONIAN

(64)   a.   Vo zgrada-ta
             in building-F.SG.DEF

             'in the building'

       b.   Vo crvena-ta   zgrada
             in red-F.SG.DEF building.F.SG

             'in the red building'

Nouns without definite articles in locative prepositional phrases are often interpreted as indefinite in the standard language. Compare example (65a), drawn from announcements collected during fieldwork, where the definite article indicates definite reference for *kniga-ta* 'the book,' with example (65b), where *kniga* 'book' is interpreted as indefinite.

MACEDONIAN

(65)   a.   Avtori-te   na promocija-ta   prezentiraa del od   aforizmi-te
             authors-PL.DEF at promotion-F.SG.DEF presented   part from aphorisms-PL.DEF
             sodržani   **vo kniga-ta**.   Kniga-ta   e posvetena   na počinati-ot
             gathered.PL in book-F.SG.DEF book-F.SG.DEF is dedicated.F.SG to deceased-M.SG.DEF
             ľubitel   na aforizmi Blaže Temov od   Veles.
             lover.M.SG of aphorisms Blaže Temov from Veles.M.SG

             'The authors at the promotion presented a portion of the aphorisms gathered in the book. The book is dedicated to a deceased aficionado for aphorisms, Blaže Temov, from Veles.'

       b.   Akademik-ot   Blaže Ristovski gi   sobra   istražuvaňa-ta   na Draganov
             academic-M.SG.DEF Blaže Ristovski them gathered researchings-PL.DEF of Draganov
             i   gi   izdade   **vo kniga**.   Kniga-ta   može da se   najde vo knižarnica-ta
             and them published in book.F.SG. Book-F.SG.DEF can SBJ REFL found in librar-F.SG.DEF
             Matica.
             Matica.F.SG.

'Professor Blaže Ristovski gathered Draganov's research and published it in a book. The book can be found at Matica Library.'

In (65a), a specific, known book is the topic of discussion. The noun *kniga-ta* 'the book' appears with the definite article throughout the section of discourse, both in the prepositional phrase *vo kniga-ta* 'in the book' and as the subject of the last sentence. This contrasts with (65b), where the referent of *kniga* 'book' first appears without a definite article, subsequently becoming a definite referent for which the noun *kniga-ta* 'the book' requires a definite article.

Given the considerable range of meanings that bare nouns can have in Macedonian, one would expect that definite reference would be consistently marked by the definite article. This is true in argument positions outside of prepositional phrases. However, there is variability in the appearance of definite articles in Macedonian prepositional phrases where the noun is intended to have definite meaning. The following examples, from standard register speech, exhibit the expected pattern.

MACEDONIAN

(66) a. Pronašle dvoglava zmija vo **dvor-ot**.
they.found two.headed.F.SG snake.F.SG in **courtyard-M.SG.DEF**

'They found a two-headed snake in **the courtyard**.'

b. Račen raketen frlač našol eden Čairčanec vo
handheld.M.SG rocket.M.SG launcher.M.SG he.found IND.M.SG Čair.resident.M.SG in
dvor na ulica "Džon Kenedi."
courtyard.M.SG on street.F.SG John Kennedy

'A Čair resident found a hand-held rocket launcher in **a courtyard** on John Kennedy St.'

Example (66) exhibits the expected distribution of definite and indefinite interpretation. In (66a), the courtyard is under the ownership of the individuals who are the subject of the verb *pronašle* 'they found.' There is only a single, known courtyard, and so the courtyard has definite reference, indicated by the post-posed definite article in *dvor-ot* 'the courtyard.' In contrast, there is no unique courtyard on John Kennedy street in (66b), and thus the reference of *dvor* 'courtyard' is indefinite and this interpretation is reinforced by the lack of a definite article in the prepositional phrase.

The next two examples, however, show—using the same lexical item—that this pattern of definiteness and specificity marking in locative prepositional phrases is not fixed.

MACEDONIAN

(67) a. Vo dvor mu padnal del od helikopter.
in courtyard.M.SG him.DAT it.fell piece.M.SG from helicopter.M.SG

'A piece of a helicopter fell into his courtyard.'

b. Kopal vo dvor i našol bogatstvo.
he.dug in courtyard and he.found riches.N.SG

'He dug in the courtyard and found riches.'

In (67a), the helicopter part has fallen into a particular courtyard possessed by known individual. While there is an external possessor represented by the dative pronominal clitic to the verb

mu 'to him,' the prepositional phrase *vo dvor* 'in (the) courtyard' contains no definite article, despite the definite reference of 'courtyard.' This is also the case in (67b): like in (66a) and (67a), the courtyard is a definite, individual courtyard uniquely owned by the person who has found the riches. Nonetheless, *vo dvor* contains no definite article. The same definite reference and the same surrounding structure is involved in (66a) and (67b), with the same lexical item, and yet they show different patterns of definiteness marking. The expression of definiteness in prepositional phrases is subject to variability that does not occur outside of the domain of prepositional phrases, which indicates that Macedonian shares with Balkan Romance and Albanian a peculiar treatment of definiteness in prepositional phrases.

Establishing whether LDO applies in Macedonian is a more complicated proposition than in Albanian or Balkan Romance. One essential difference is that Macedonian does not possess the same case distinctions that define the conditions for LDO in neighboring Balkan languages. In Macedonian, all prepositions take the accusative case, though case is normally only expressed in pronouns.[1] Since prepositions only select for a single case, nominal case plays no role in determining the distribution of definiteness in Macedonian prepositional phrases, a fact which already sets Macedonian apart from Albanian and Balkan Romance. In the latter two languages, LDO is a component of the standard, literary language as well as the spoken languages. In Macedonian, the standard, literary language does not require determiner omission with unmodified nouns. The lack of case specificity and the lack of a grammatical requirement for determiner omission appears at first glance to exclude Macedonian from significance in the study of LDO in Balkan languages. Nonetheless, Koneski (1967:509-513), in his summary of the grammar of Macedonian based on the works of early twentieth-century Macedonian authors, notes some trends towards the omission of the definite article in locative prepositional phrases. These trends establish a basis for comparison between those lexical exceptions that are embedded in Macedonian grammar and a more general pattern of determiner omission that occurs as an optional pattern of use in conversation.

There is a distinction between the use of the preposition *vo* and *v* 'in/to' that matches some of the features of LDO, though it also shows signs of a lexicalized distinction rather than a grammaticalized distinction. The preposition *vo* may be used with an unmodified noun to denote definite reference, as with *vo bunar-ot* 'in the well' in (68).

MACEDONIAN

(68) **Vo bunar-ot**   imalo   edna   vrata   otvorena
      In well-M.SG.DEF there.was IND.F.SG door.F.SG open.F.SG

      '**In the well** there was an open door.'   (Koneski 1967:510)

This example shows that the behavior of the noun with *vo* conforms to the expectation that Macedonian lacks LDO. The preposition *vo* can also take an unmodified, bare noun when the reference is to a place where an activity typically occurs. Here, the definite reference is not individuated; in (69a), where *reka* 'river' appears without a definite article, no particular river is meant—rather, the general location of any given river is emphasized. Likewise, in (69b), no particular *grad* 'city' is intended, only the general location of 'a city' as opposed to 'a village' or 'the countryside' (Koneski 1967:510). This use of the preposition *vo* and an unmodified noun puts

---

[1]In some dialects, accusative case may be expressed on personal names or with feminine nouns; however, this is not part of the standard language. Since case is not an aspect of Macedonian nominal morphology, either in the noun itself or in the definite article, case will not be marked in the following examples.

the emphasis on the typical activity in locations of that type rather than a specific instance of the location. This is to be expected from the lack of a determiner in these constructions.

MACEDONIAN

(69) a. Jas **vo reka**   sum se   kapela, a   vo more   ne
       I   **in river.F.SG** I.am REFL swum   but in sea.N.SG no
       'I have swum **in a river** before, but never in the sea.'

   b. **Vo grad** ne   sum živeal
      in city not I.am lived
      'I've never lived **in a city**.'   (Koneski 1967:510)

The locative preposition *v*, a variant of *vo*, shows different behavior and triggers different interpretations of the definiteness of nouns in its domain. *V* can only be used with an indefinite, bare noun, and cannot be used to denote a general location where an activity takes place, in contrast to *vo* as in (69) above. In these cases of *v* plus a bare noun, the noun is interpreted as definite, referring to a specific place (Koneski 1967:510). Thus *pole* 'field' in (70) refers to a specific, known field, not to 'fields' as a general location.

MACEDONIAN

(70) Tie   dve   piliňa   i   ti sam   si   gi videl i si   gi čul,
     those two.N.PL chickens-PL and you alone.M.SG you.are them seen and you.are them heard,
     sinko, **v pole**.
     son.VOC, **in field.N.SG**
     'You yourself saw and heard those two chickens **in the field**, son.'   (Koneski 1967:510)

This is reminiscent of the way in which bare, unmodified nouns are interpreted when governed by locative prepositions in Albanian and Balkan Romance. However, the alternation of *vo* and *v* is not governed by the same strict grammatical rules that apply to LDO in Albanian and Balkan Romance. Rather, it appears as a tendency of use, and one that was already falling into a pattern of fixed, idiomatic phrasal distribution as *vo* overtook more roles in Macedonian following the codification of the grammar of the literary language. Indeed, in the modern language, *v* appears almost solely as the preposition marking days of the week (where a meaningful alternation between *v sabota* 'on Saturday' and *vo sabota-ta* 'last Saturday' has been standardized and lexicalized), or in idiomatic expressions. Idiomatic prepositional phrases where the nominal constituent occurs without a definite article resemble adverbial phrases and occur with a range of prepositions. In some of these phrases, the indefinite noun can be replaced with a definite noun without changing the meaning of the phrase, as in (71a-b).

MACEDONIAN

(71) a. Toj ja fati za raka
        he her took for hand.F.SG
        'He took her by the hand.'

   b. Toj ja fati za raka-ta
      he her took for hand-F.SG,DEF

'He took her by the hand.'

In these cases, the insertion of a definite article does not cancel the interpretation of *za raka* as 'by the hand'; the expression of definiteness through an overt marker is irrelevant to the interpretation of the sentence (Koneski 1967:232-233). The fact that a definite article can be inserted into these phrases additionally distinguishes them from prepositional phrases with LDO in Albanian or Balkan Romance, where the use of definite articles on bare nouns within the domain of a closed set of prepositions is strictly ungrammatical. In Albanian or Balkan Romance, speakers outright reject a definite article in similar prepositional phrases with unmodified nouns. There is no appropriate semantic context in which the definite article is licensed for Albanian and Balkan Romance locative prepositional phrases that select the accusative case. In Macedonian, where case selection is not a factor, the definite article is acceptable in locative prepositional phrases, but may or may not make truth-conditional changes to the meaning of the sentence that extend beyond the mere addition of definiteness. As above, with *toj ja fati za raka*, it is sometimes possible to have variation without a change to the denotation of the sentence. There are other cases where changing the noun to a definite form changes the meaning of the phrase and cancels its idiomatic interpretation, as in (72a-b).

MACEDONIAN

(72) a. Mi   dojde veḱe   do guša
           me.DAT it.came already to throat.F.SG

           'I'm fed up!' (only idiomatic)

     b. Mi   dojde veḱe   do guša-ta
           me.DAT it.came already to throat-F.SG.DEF

           'It's already come up to my throat.' (only literal)   (Koneski 1967:232)

Here we see a distinction between *Mi dojde vekje do guša* 'I'm fed up' (with only an idiomatic interpretation) and *Mi dojde vekje do guša-ta* 'It's already come up to my throat' (with only a literal interpretation). This appears to be a case where the idiomatic interpretation requires a bare noun, but the standard rules of interpretation apply when the definite article is included. The distinction between these two phrases is consistent with an explanation where the definite article is required to express definiteness in prepositional phrases in standard, literary Macedonian.

The norms of the examples above apply to the written Macedonian language and constitute lexical exceptions to the normal distribution of the definite article. In colloquial Macedonian, however, unmodified nouns can be interpreted as definite within the domain of locative prepositions, in ways that cannot constitute lexical exceptions. As Koneski notes,

> "When we come close to personal names in the aforementioned situations, a general substantive that designates some place can become like names for locations, for which there is no need for an article as a marker of their definiteness. This occurs with nouns that designate places in the environs of a house, such that the residents individualize them, or more broadly in the environs of a settlement, again for the same reason, especially if the given object is the only one of that sort in the settlement" (Koneski 1967:227).

He gives such examples as:

MACEDONIAN

(73) Site se   pulele   kaj porta
      all   REFL gawked at   gate.F.SG
      'Everyone was standing at the gate gawking.'

(74) I   se   zatvorila porta so   silno   udruaňe
      and REFL closed   gate with strong.N.SG hitting.N.SG
      'And the gate closed with a great slam.'   (Koneski 1967:227)

The definite, specific referent of *porta* 'gate' occurs within a prepositional phrase in (73), but also occurs with the same word as a subject argument in (74). In either case the noun lacks a definite article. Koneski traces the origin of these unarticulated constructions to earlier stages of the Macedonian language in which the grammatical reorganization that introduced the post-positive definite article had not yet occurred. Nouns without the definite article in contexts where they would be expected to bear it are "remnants" of the state of the language when it more closely resembled Late Common Slavic, the ancestor of the South Slavic languages. In Late Common Slavic, the Balkan style of definite article had not yet been introduced. Koneski notes that articles are also lacking in general for nouns used in standardized, idiomatic greetings and blessings, whose ritualistic character would afford them some protection from the encroachment of definite articles. This presents the possibilities of two distinct pathways toward constructions in which an expected definite article is lacking within the prepositional phrase: (i) specific prepositions trigger the omission of the definite article in specific syntactic circumstances, and (ii) lexicalized remnants of an earlier stage of the language in which the determiner was generally optional were preserved in the modern language and bound to set phrases and nouns. If determiner omission in colloquial Macedonian prepositional phrases was restricted to these specific words used in restricted semantic domains, the second explanation would be a more plausible one.

4.2 Non-lexicalized LDO in Macedonian

Fixed phrases and localized landmarks, however, are not the only areas where determiners may be omitted in locative prepositional phrases in colloquial Macedonian. This suggests that determiner omission in Macedonian prepositional phrases cannot be accounted for as a lexicalization of bare nouns from an earlier stage of the language in which determiners were absent. Speakers prefer determiner omission in constructions for familiar, immediately identifiable, concrete locations that are known to both the speaker and listener, as in the following example:

MACEDONIAN

(75) a. Ostavi ja   **vo kola**.
          leave   her **in car.F.SG**
          'Leave it **in the car**.'

      b. ?Ostavi ja   **vo kola-ta**.
          leave   it **in car-F.SG.DEF**
          'Leave it in (that?) car (not this one?).'

c. Ostavi ja **vo crvena-ta   kola**.
leave   it **in red-F.SG.DEF car.F.SG**
'Leave it **in the red car**.'

While it is not strictly ungrammatical to have a definite article in the above construction, consultants find it to be strange, prompting an interpretation that there are multiple cars in the situation, and that the speaker is emphasizing a particular car in contrast to other contextually available cars. If a unique car is assumed to be known to both speakers, then (75a) is the only appropriate utterance in the situation. In contrast, in standard prose and in describing situations in which the speakers have no relation to the referent, the definite article is required to mark definiteness. It should be noted that the use of the word *kola* 'car' in this function is a modern innovation. Automobiles were not within the range of experience of speakers of Late Common Slavic, at the time when definite articles were not yet incorporated in Balkan Slavic grammar. Thus, it would be strange for this construction to be considered a remnant from earlier stages of the Macedonian language. Nor is *kola* among the nouns that can occur without a definite article when marking a definite referent in argument positions. This construction is specific to the locative prepositional phrase *vo kola*. Moreover, the introduction of a modifying attribute, as in (75c) requires the use of a definite article to express definite reference. These are the conditions for LDO observed in Albanian and Balkan Romance.

In another example, speakers consulted about their intuitions for the sentence 'put it in the fridge' show a marked preference for (76a) rather than (76b).

MACEDONIAN

(76)   a. Stavi go vo frižider
put   it in fridge.M.SG
'Put it in the fridge.'

b. ?Stavi go vo frižider-ot
put   it in fridge-M.SG.DEF
'Put it in (that?) fridge (not this one?).

c. Prodavačka-ta   vo posleden moment   se pomesti od   mesto-to   na
saleswoman-F.SG.DEF in last.M.SG moment.M.SG REFL move   from place-N.SG.DEF on
koe   padna frižider-ot
which.N.SG fell   fridge-M.SG.DEF
'At the last moment, the saleswoman moved out of the way of where the fridge fell.'

When asked for reactions to (76b), one consultant suggested that he found (76b) acceptable only in a context where there were multiple refrigerators in a single house and the speaker was pointing out a particular one. The application of the definite article in this case, while not ungrammatical, creates the expectation of a set to which a choice function is being applied. In order to obtain a definite reference in a normal context, the noun must appear bare. Trying to include a definite article results in a strange reading for the sentence. But when the noun appears in an argument position for the verb, as with (76c) where *frižider-ot* 'the fridge' is the subject of the verb *padna* 'it fell,' the definite article is required for definite reference.

This phenomenon in colloquial Macedonian is not limited to the locative preposition *vo*. When prepositions are compounded, the use of the bare noun to express definite reference becomes even

more frequent. Western peripheral dialects of Macedonian exhibit a pattern of prepositional doubling that is an exact reproduction of a similar pattern in Aromanian, which was widely spoken in the area (Markoviḱ 2012b:1). Here, locative prepositional phrases with nouns that are intended to have definite meaning govern a noun in bare form, without an article. This can then become the target for a second preposition using the first prepositional phrase as a localizer of a spatial relation—an additional feature that Macedonian shares with Aromanian. In a number of examples, the noun would take a definite article outside of the localizing prepositional context, but it appears bare, without a definite article, in both the Macedonian and its equivalent Aromanian translation.

MACEDONIAN

(77) Koga se vrativ od na pazar, bev mnogu umoren.
when REFL I.returned from at market.M.SG I.was very tired.M.SG

'When I returned from the market, I was very tired.'

AROMANIAN

(78) Kãn mi turai di la pazari, arem mult akrumat.
when me I.returned from at market.M.SG, I.was very tired.M.SG

'When I returned from the market, I was very tired.' (Markoviḱ 2013:3)

As can be seen in the comparison between Macedonian (77) and Aromanian (78), the structure is exactly parallel, including in the omission of the definite article. In such compound prepositional constructions in which one preposition expresses goal and another preposition acts as a localizer, Macedonian consultants prefer a bare noun even when a definite sense is intended. The structure produced by Macedonian speakers mirrors that of Aromanian speakers. The use of the definite article in these Macedonian sentences is dispreferred. This structural preference is particularly common with compound prepositions, but also applies to single locative prepositions.

MACEDONIAN

(79) Toj se skri zad vrata. Toj izleze od zad vrata.
he REFL hid behind door.F.SG he exited from behind door.F.SG

'He hid behind the door. He leapt out from behind the door.' (Markoviḱ 2011:196)

Macedonian consultants, when asked to describe situations involving a localizer of a spatial relation in combination with a non-locative preposition, also prefer a bare noun to express definite reference.

MACEDONIAN

(80) a. Sekoja godina kupuva petunii za na terasa.
every.F.SG year.F.SG he.buys petunias.F.SG for on terrace.F.SG

'Every year he buys petunias for the terrace.'

b. *Sekoja godina kupuva petunii za na terasa.
every.F.SG year.F.SG he.buys petunias.F.SG for on terrace.F.SG

'Every year he buys petunias for the terrace.'

In (80a), it is not any given terrace that is meant, but rather a specific terrace that is known to both speaker and listener. This would ordinarily prompt a definite article. In this context, however,

a definite article is judged inappropriate, as in (80b). Macedonian, then, exhibits extensive similarities with Aromanian in the ways definite articles are distributed in locative prepositional phrases. Although this structure originated in western peripheral dialects of Macedonian, it is now spreading to dialects outside of the main zone of contact with Aromanian. Speakers native to central dialects regularly employ compound localizing prepositions and LDO, in addition to other dialectal features once confined to the zone of contact with Aromanian (Markoviḱ 2012, 2013). Given the conjunction of features that Macedonian shares with Aromanian, it is likely that this feature is a result of contact between the two languages. Even though the standard Macedonian language does not show consistent, grammatical LDO, the colloquial language frequently opts for constructions that resemble LDO in Albanian and Balkan Romance. Contact with Aromanian provides a plausible route for the introduction of this feature into the structure of spoken Macedonian. It also accounts for the absence of LDO in other spoken forms of Balkan Slavic.

4.3   Comparison of locative prepositional phrases in Macedonian and Bulgarian

It is important to compare the behavior of locative prepositions in Macedonian with Bulgarian, where determiner omission is not found, despite the close relation between Macedonian and Bulgarian. In Bulgarian, a noun with definite reference must take a definite article, regardless of whether the noun occurs in a locative prepositional phrase or whether it is familiar. There is no link between the modification of the noun and the expression of definiteness in prepositional phrases, nor is the expression of definiteness for nouns in prepositional phrases distinctive when compared with other syntactic contexts (Boretzky 1968:132; Mladenova 2007:137). Article omission in Bulgarian has non-specific meanings (Nicolova 2008:89). There are exceptions, but these occur only in calcified phrases that remain from the period before Bulgarian had incorporated a definite article into its grammatical system. These are the sort of lexicalized constructions that exist in other languages with markers for definiteness, but without LDO. Indeed, these are the same kind of archaic preservations as previously analyzed for Macedonian in §4.1. They exist in opposition with more modern phrases with the same construction and meaning in which a definite article is required (Nicolova 2008:103).

BULGARIAN

(81) a. zaminavat na selo
          they.leave to village.N.SG
          'They leave for the village.'

      b. zaminavam v grad-a
          I.leave   in city-M.DEF.ACC
          'I leave for the city.'   (Nicolova 2008:103)

In (81a), the phrase *na selo* 'to (the) village' is fixed in a form lacking a definite article, but the referent is still a specific village. This does not extend to every object of this construction, however. Nicolova contrasts (81a) with (81b), where a definite article is required for definite reference in the prepositional phrase *v grad-a* 'to the city.' These kinds of lexicalized contrasts in the use of the definite article extend to other nouns of location.

(82)   a. otivam na loze
          I.go   to vineyard.N.SG
          'I go to the vineyard.'

      b. otivam na niva-ta
          I.go   to field-F.SG.DEF
          'I go to the field.'
          (Nicolova 2008:89)

Examples (82a) and (82b) provide evidence that even in very similar semantic contexts, the words which allow determiner omission in Bulgarian are lexically specified, such that *loze* 'vineyard' does not take a definite article, but *niva-ta* 'the field' does. This same contrast can be seen be seen between *učilište* 'school' and *universitet-a* 'the university' in (83a) and (83b).

(83)   a. vrǔštam se   od   učilište
          I.return REFL from school.N.SG
          'I return from school.'

      b. vrǔštam se   od   universitet-a
          I.return REFL from university-M.ACC.DEF
          'I return from the university.'

      (Nicolova 2008:89)

There is also a systematic pattern of differentiation between institutions at which an activity occurs and the buildings in which those institutions are housed. Nouns referring to institutions allow for the omission of the definite article, while nouns referring to concrete locations requires the definite article to specify the location. This is similar to the system of lexically determined determiner omission found in English.

(84)   a. otivam na pazar
          I.go   to market.M.SG
          'I go shopping.'

      b. otivam na pazar-a
          I.go   to market-M.ACC.DEF
          'I go to the market.'

                (Nicolova 2008:89)

Example (84a) indicates that the speaker is going to go shopping, not necessarily at the marketplace, but possibly in a store or in a mall. Nothing specific about the location designated by the referent of *pazar* 'market' is implied in (84a). In contrast, (84b) means specifically that the speaker is going to the marketplace, as opposed to a store (Nicolova 2008:103). This can be compared against the Macedonian (Marković 2012a).

(85) Od   na pazar   do doma dojdov so   trčaňe
      from at market.M.SG to home I.came with running.N.SG

      'I ran from the marketplace back home.'   (Markoviḱ 2012a:85)

In Macedonian, the use of *pazar* without an article is specifically localized to the marketplace, particularly in compound prepositional constructions. There is no possibility for an alternation between *pazar* 'market' and *pazar-ot* 'the market' to distinguish between the location and the activity, as in Bulgarian. Rather, Macedonian concurs with Aromanian and Albanian in using a bare noun to designate a specific, known location within a locative prepositional phrase that would otherwise take a definite article. It appears that the structure of the Macedonian language, under the influence of Aromanian, has allowed for a tendency of determiner omission to develop.

Macedonian shows some differences in its nominal syntax that distinguish it from Albanian and Balkan Romance. Adjectives primarily appear before nouns, and indefinite articles are not fully grammatically integrated into the system. This also means, however, that bare nouns are consistently interpreted as indefinite in the standard language. The expectation, therefore, would be that bare nouns would be interpreted as indefinite in locative prepositional phrases as well. Some exceptions to this rule appear to be the result of fossilized combinations retained from the period before Macedonian fully developed a post-posed definite article. But modern, colloquial Macedonian has developed a system of determiner omission for identifiable nouns with definite reference that does not depend on calcified lexical items. Although this system does not apply across all speech registers, it is common in spoken Macedonian, to the extent that determiner omission is preferred as an expression of definite reference in locative prepositional phrases. Indeed, in those western peripheral dialects that allow for compound localizing prepositions, determiner omission is required. These dialects developed compound prepositions, among other features, through long-term contact with Aromanian, for which LDO is a consistent grammatical feature. In contrast, Bulgarian does not show determiner omission in the same contexts where it is common in Macedonian. The determining factor appears to be contact with Aromanian and its influence on the structure of Macedonian dialects.

Thus from this data, it can be concluded that Macedonian shows LDO, although it is not fully grammatically integrated. The pattern of determiner omission in Macedonian strongly resembles that found in Albanian and Balkan Romance, languages with which Macedonian was known to have extensive contact. In contrast, Bulgarian lacks LDO, despite the close relation between Macedonian and Bulgarian as part of Balkan Slavic. In many local areas of contact, particularly between Macedonian and Aromanian, LDO looks like a prime example of a contact-induced innovation. But its distribution is only partial in the Balkan linguistic area. In addition, it is not fully integrated into Macedonian in the same way that it is a grammatical part of Albanian or Balkan Romance. In order to address these puzzles, I turn to a diachronic analysis of LDO, comparing its attestation in languages of the Balkans over time, and across dialects.

Chapter 5

The Diachronic View of Locative Determiner Omission

The task of this chapter is to establish that LDO first developed in Balkan Romance and then spread to other Balkan languages through periods of language contact. What makes LDO interesting is not only that it is an unusual structure for indicating definiteness, but that this feature appears in a cluster of Balkan languages, including Albanian. The same pattern of grammatically-conditioned determiner omission within prepositional phrases also applies in Aromanian (Romance), and to a certain extent, in Macedonian (Slavic). In addition, LDO is found in Romanian, which does not directly neighbor either Albanian or Macedonian (see Figure 5.1).

Considering that the pattern of LDO itself is uncommon, and absent from the broader branches of Slavic and Romance spoken outside the Balkans, the coincidence of determiner omission in this cluster of languages warrants closer examination and comparison. When a grammatical feature is shared among languages that already resemble each other in many respects, it is common to explain the convergence as a result of contact between the languages (Campbell 1985:26-29). But contact is always a matter of interaction between speakers, not languages in the abstract (Friedman 2006a). Any appeal to language contact must provide an account of the speaker interactions that would have resulted in the grammatical convergence (Masica 1976:173). Considering that some of the Balkan languages that show LDO are not spoken in contiguous areas with one another, a historical and dialectological approach to the development of LDO is required in order to establish that it was not an independent innovation in each of the languages. We must provide positive evidence of sustained contact between speakers in order to eliminate the possibility that the presence of LDO in these languages is coincidental or an inheritance from a common source (Muysken 2008:8).

In this chapter, I examine the diachrony of the development of LDO in each language of the Balkans. This is complicated by gaps in attestation. We have documentation for the predecessors to Balkan Romance and Balkan Slavic, in the form of Latin and Old Church Slavonic, respectively. The lack of consistent definite articles in Late Latin (Bauer 2007:117-118) or Late Common Slavic (Mladenova 2007:98), however, means that we can only establish that LDO was not present in the previous forms of these languages, since the identification of LDO depends on consistent application of a grammatical definite article in other syntactic contexts (see §2.3). For Albanian, we can only tentatively reconstruct earlier states of the language. Between unified Slavic, Romance, or Albanian and the modern descendent languages, documentation is unavailable for key periods of the grammatical development of Balkan languages. The earliest attestation of a language with

Figure 5.1: Map of the Balkan languages that show locative determiner omission

recognizably Balkan grammatical characteristics comes from Balkan Romance. Fortunately, this is also the Balkan language group in which LDO first emerged, as I will show.

In §5.1, I will first address the diachrony of LDO in Balkan Romance as a whole and demonstrate that LDO was an early innovation shared among all the descendants of Balkan Late Latin. In §5.1, I will show that it was already a feature present in the earliest documentation of Romanian. In §5.1, I introduce data from Istro-Romanian and Megleno-Romanian, two further branches of Balkan Romance, to demonstrate that LDO must have existed in common Balkan Romance prior to the divergence of its branches. This provides an early date for the integration of LDO into the Balkan Romance grammatical system and a basis for comparison with the diachrony of other Balkan languages.

Albanian, in its earliest attestations, does not show LDO. Rather, it shows a different structure involving a morphological locative case that depended on a complex interaction with the definite article system and prepositional phrases. It is possible that Albanian could have developed LDO

independently, but I argue from a diachronic analysis of the locative case and LDO in Albanian dialects that this is not what happened. Instead, LDO should be viewed as a contact-induced innovation in Albanian for which the locative case provided the structural preconditions. In §5.2, I summarize the grammatical features of proto-Albanian, to the extent that they can be reconstructed. I compare the early Albanian locative case to modern LDO and show that the triggering factors governing locative case and LDO resemble each other in a number of ways, but also differ in crucial respects that complicate the path from loss of the locative case to the grammatical integration of LDO. In §5.2, I discuss the ways in which the locative case could be elided in early modern Albanian, while still showing signs of its presence in the agreement morphology of other noun phrase constituents. In §5.2, I focus on the locative case in two dialects, Arbëresh and Arvanitika, that separated from the main body of Albanian speakers in the 14th and 15th centuries and that show the outcome of either the preservation or elision of the locative case. Importantly, these dialects do not exhibit LDO as is found in the main body of Albanian dialects. In §5.2, I discuss the different path taken by the main body of Tosk and Geg dialects toward the grammatical integration of LDO. In §5.2, I then compare the emergence of LDO in dialects of Albanian against LDO in Balkan Romance. In particular, I show that the development of LDO in Albanian could not have resulted from internally-motivated developments, even though the proto-Albanian locative case may have had an influence on the development of LDO in Balkan Romance during their early period of contact. Rather, LDO must have developed at a relatively late date in Albanian, in the 19th and 20th centuries, and its distribution corresponds to areas where Balkan Romance speakers assimilated into the Albanian community. The loss of the locative case, a process internal to Albanian, provides the necessary precondition for the adoption of LDO. Nonetheless, without ongoing contact with Balkan Romance and assimilation of Balkan Romance speakers, LDO does not become a fully grammatically integrated part of Albanian grammar.

In §5.3, I support my argument for Balkan Romance as the source of LDO with diachronic data from Macedonian. Like Albanian, Macedonian did not develop LDO until relatively late, and first in dialects that showed extensive contact with Balkan Romance. The influence of Aromanian is key to explaining the presence of determiner omission in Macedonian dialects and its absence in Bulgarian dialects. An analysis of the diachrony of definite determiner development shows that Macedonian could not have acquired LDO until the early 19th and 20th centuries, during the same period of Aromanian assimilation that exerted an influence on Albanian.

I thus argue that LDO developed in Balkan Slavic, Romance, and Albanian languages through contact between their speaker communities, rather than as an independent development in each language. This contact proceeded in multiple stages, as depicted in Figure 5.2

I argue that the phenomenon first appeared in Balkan Romance and then spread to Albanian and Macedonian in contact with Aromanian, with further processes of internal contact between dialects accounting for the peculiarities of the distribution of LDO in Albanian. The specific circumstances under which LDO developed in each language, and the ways in which the grammatical feature was integrated into the structure of each of these languages, provides us with greater detail about the kind of contact these speaker communities had with each other over time. It also serves as a test case for theories of language contact and shows how a linguistic area may be constituted from the clustering of local interactions between speaker communities. The best explanation for the similarities between Aromanian, Albanian, Macedonian, and Romanian as concerns their treatment of definiteness in prepositional phrases lies in contact-induced convergence among them, with Aromanian as the source of this specific structure.
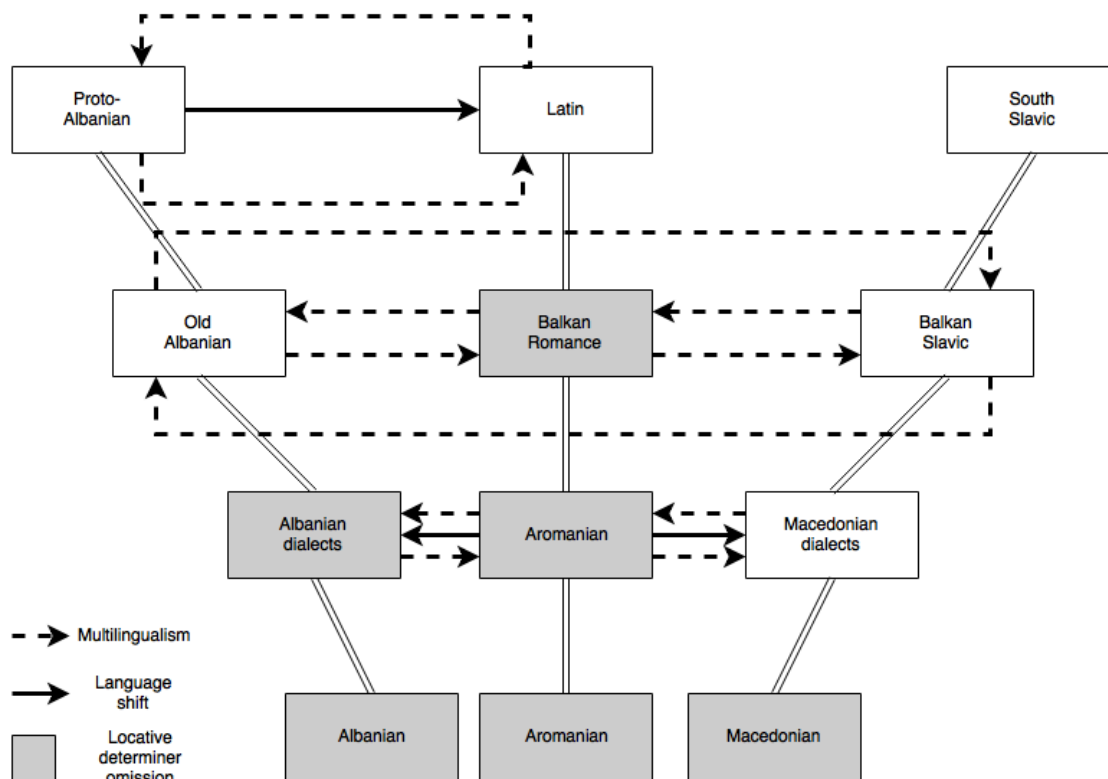
Figure 5.2: Diagram of the path of transmission for locative determiner omission

5.1 The diachrony of LDO in Balkan Romance

Unlike with Albanian, it is possible to compare Balkan Romance languages with an earlier, well-attested ancestor: Latin. Thus it is helpful to begin with an examination of the diachronic development of LDO in Balkan Romance. The Classical Latin language that preceded Balkan Romance did not distinguish definiteness with any consistent article (Löfstedt 1956, Rosén 1994:135, Bauer 2007:117-118). Demonstratives in Classical Latin could express article-like functions, but the development of the consistent, mandatory use of a definite article from Latin demonstratives (whether of the *ipse* or the *ille* sequence) was not complete even in Late Latin (Bauer 2007:119-120). The development of the definite article in Late Latin, later inherited in all modern Romance language, has been proposed as an outcome of sustained contact with Greek. Ancient Greek possessed an article well before cultural and linguistic contact with Latin (Leumann, Hoffman, and Szantyr 1965:191-192). However, the development of the definite article in Latin proceeded independently, at its own pace and following functional and syntactic pathways different from those of Greek (Bauer 2007:128-129, 136). Regardless of the origin of the definite article in Latin, it appears in all descendent languages, but with differing syntactic distributions, a matter of particular relevance for Balkan Romance.

The primary syntax for definite reference across Latin successor languages is definite article followed by noun phrase, which matches the primary order of demonstrative followed by noun phrase found in Latin (Bauer 2007:116). Balkan Romance, however, reverses this common order and post-poses the definite article derived from the demonstrative *ille* to the noun, or the first noun

phrase constituent. The consolidation of the Latin demonstrative + noun combination into the Balkan Romance noun + definite article combination occurred during a gap in documentation. The configuration of the post-positive article found in Balkan Romance is not at all typical for languages descendent from Late Latin (Lindstedt 2000:235); this swap of the order between demonstrative and noun is unique to Balkan Romance languages. In other Romance languages, the definite article may undergo assimilation to the preposition, but these phonological mergers produce consistent outcomes that can be derived from the syntactic placement of the definite article before the noun. When discussing LDO, where the article is not assimilated to a constituent, but rather omitted all together, it is more difficult to identify the beginnings of the pattern in Late Latin. It is not possible to compare LDO synchronically with any other Romance language, because the feature is entirely lacking in any other Romance language or dialect (Reinheimer Rîpeanu 2001:196-197).

Diachronically, the consistent usage of a definite article in other syntactic contexts is a precondition for identifying the unexpected 'omission' of the article in prepositional phrases. The demonstrative could certainly be found in locative prepositional phrases in Late Latin texts, at which time its demonstrative semantics were already becoming bleached and a definite meaning was developing (Bauer 2007:126-127). Thus, the omission of the definite article in Balkan Romance locative prepositional phrases cannot be traced to any tendency of the demonstrative in Late Latin, from which the Balkan Romance definite article was derived. As a point of comparison, a tendency toward the lack of definite article in prepositional phrases can be observed in Old French, even after definite articles had developed for nouns in other syntactic positions. The first prepositional phrase for which the article began to regularly appear with nouns was in positions governed by the preposition 'with,' which shows an intriguing parallel with the exceptional behavior of 'with' in Albanian and Balkan Romance (Reichenkron 1962:102). Nonetheless, this lag in the appearance of articles after prepositions in Old French does not match with the specific semantic restrictions on locative determiner omission to *locative* prepositions found in Balkan Romance. One possibility is that Balkan Romance failed to develop definite articles from Late Latin in the restricted domain of locative prepositional phrases, while developing them everywhere else. This, however, would not explain why definite articles do appear with modified nouns in locative prepositional phrases. An alternative proposal is that Balkan Romance developed definite articles across all domains, but began at a certain point to omit them in locative prepositional phrases. This is the more likely scenario, but because of gaps in the record of attestation, it is not possible to determine exactly when the post-posed definite article was fully developed in Balkan Romance. Nonetheless, by comparing the branches of descendent languages of Late Latin spoken in the Balkans with the earliest attestations, it is possible to establish that LDO must have been present in early Balkan Romance.

In §5.1, I establish the earliest point of attestation in which LDO appears in its fully grammatically integrated form and in §5.1 I discuss its appearance in the other branches of Istro-Romanian and Megleno-Romanian, with the goal of arguing that LDO must have been present in Balkan Romance at a time before its separation into distinct branches. This would place LDO in Balkan Romance at a time before the Slavic migrations in the late 6th century and during the period in which Balkan Romance was still in intimate contact with Albanian.

Consistent LDO in Early Romanian

Due to lack of attestation, major portions of the history of Balkan Romance are unavailable for examination. By the earliest attestation of Balkan Romance, definite articles are already used con-

sistently and according to the Balkan syntax of post-posing to the first nominal constituent. It is in the document the Letter of Neacşu of Câmpulung—a 1521 missal from a boyar in the Wallachian territories written in both Old Church Slavonic and early Romanian that provides the first attestation of Balkan Romance—that we see the outcome of developments from Late Latin in the Balkans (Rosetti 1968:468-469). In the letter, modified nouns with definite reference in prepositional phrases consistently appear with definite articles, as would be expected in a system where definiteness has been fully integrated into the grammatical system.

EARLY ROMANIAN (LETTER OF NEACŞU)

(86) se-au   prins neşte meşşter   den Ţarigrad   cum vor
      REFL-they.have realized some.PL masters.M.NOM from Constantinople.M.ACC how they.will
      treace aceale corăbii   **la loc-ul**   **cela strimt-ul**
      pass   those ships.F.NOM **at place-M.ACC.DEF PRT narrow-M.ACC.DEF**
      'Some experts from Istanbul realized how to make the ships move past that tight place.'
      (Avram 1962:149)

Example (86) shows a noun modified with an adjective (also bearing a definite article) *cela strimt-ul* 'narrow' in the prepositional phrase *la loc-ul cela strimt-ul* 'at the narrow place.' The definite article marking the noun phrase *loc-ul* 'the place' is repeated on the adjective *strimt-ul* 'narrow.' The expression of definiteness in this locative prepositional phrase is by means of a post-posed definite article.

EARLY ROMANIAN (LETTER OF NEACŞU)

(87) dau ştire domnietale **za   lucr-ul   torci-lor**
      I.give news to.you   **about work-M.ACC.DEF Turks-M.GEN.DEF**
      'I inform you of the deeds of the Turks.'   (Avram 1962:149)

Example (87) contains a noun with a genitive-case modifier *turci-lor* 'of the Turks' for the noun in the prepositional phrase *za lucr-ul* 'about the work.' The preposition *za* is Slavic in origin, no longer used in modern Romanian, and selects the accusative case. Once again, a definite article appears on the noun in combination with a modifier in the domain of a prepositional phrase.

EARLY ROMANIAN (LETTER OF NEACŞU)

(88) au   văzut **cu   ochi-i   loi**
      they.have seen **with eyes-M.ACC.DEF his**
      'They saw with their own eyes'   (Avram 1962:149)

Example (88) contains a noun with a possessive pronoun *lu* 'his' in *cu ochi-i lu* 'with his eyes.' The noun takes a definite article in addition to the possessive pronoun. Definiteness in prepositional phrases in this document is marked with a determiner in the presence of a diversity of modifiers. This is not because modifiers force the presence of a definite article. Not all modified nouns in the document show definite articles, as illustrated by (89).

EARLY ROMANIAN (LETTER OF NEACŞU)

(89) am   auzit **de boiari   ce   sănt megiiaş**
      I.have heard **from boyars.M.ACC that are neighboring**
      'I have heard from neighboring boyars'   (Avram 1962:149)

Here the relative clause *ce sănt megiiaş* 'that are neighboring' modifies *boiari* 'boyars,' an unspecified, bare noun designating indefinite referents. The absence of a definite article corresponds with an indefinite interpretation. The expression of definiteness in nouns with modifiers is thus dependent on the presence of a definite article. This is the case both in the domain of prepositions and across all grammatical contexts.

Unmodified nouns in the document contrast with modified nouns regarding the distribution of the definite article. They show a split pattern. In two cases, what appears to be an adverbial usage results in a noun without specific, definite reference occurring in the prepositional phrase without a definite article, as in (90) and (91).

EARLY ROMANIAN (LETTER OF NEACŞU)

(90) bagă   den tote oraşe-le   căte 50   de omin   să fie **în ajutor** î corăbii
      they.take from all   cities-F.ACC.DEF each fifty of men.M.ACC SUB be **in help**   in ships.M.ACC
      'They are taking fifty men from each town to help on the ships.'   (Avram 1962:149)

In (90), *î ajutor* 'to help,' lit. 'in help' serves as an adverb, where *ajutor* 'help' has no specific referent but rather denotes an activity, nominalized from a verbal root. This is the only appearance of the word in this document.

EARLY ROMANIAN (LETTER OF NEACŞU)

(91) se-au   dus **î sus**   pre Dunăre
      they.went up **in above** on Danube
      'They went up the Danube'   (Avram 1962:149)

A similar adverbialization of a prepositional phrase appears in (91), where the phrase *î sus* 'up' is a fixed adverbial phrase of movement, with *sus* 'up' functioning independently as an adverb rather than as a noun. These usages resemble what would be expected in English or in other languages that allow for some lexicalized phrases to avoid the use of definite determiners (as in, for instance, 'at home,' a lexicalized phrase that does not permit a determiner).

Some unmodified nouns in this document have definite reference, however. One example appears with definite reference in (90): *î corăbii* 'on the ships.' Earlier in the document, the phrase *aceale corăbii* 'those ships' appears, and this phrase is repeated following the line in (90). This indicates that *î corăbii* has an intended, anaphoric definite referent: a specific set of ships that has been made known to the listener. This context of reference would ordinarily prompt the use of a definite article. Nonetheless, *corăbii* appears bare, without a definite article. This is the result would be expected if LDO was already functioning in the Romanian of the time at which this document was written. While modified nouns as in (86-88) take a definite article to indicate definite referents, or no article to indicate indefinite referents as in (89), this unmodified noun in (90) marks definiteness with the absence of a definite article.

This text contains yet more examples of unmodified nouns occurring in bare form despite definite referents. Example (91) is significant for the appearance of the prepositional phrase *pre Dunăre* 'on the Danube.' The noun *Dunăre*, although a place name, refers to a river and normally occurs in definite form: *Dunăre-a*. This is the same with another place name in the text: Ţarigrad 'Constantinople' in (86) could also appear as *Ţarigrad-ul* '[the] Constantinople' outside the syntactic context of a prepositional phrase. But following the prepositions *pre* and *din*, respectively, the nouns *Dunăre* and *Ţarigrad* appear in indefinite form; this occurs twice with the phrase *pre Dunăre*.

This seems to indicate that determiner omission already applied as a rule in this early form of the Romanian language. However, there is not an attestation of *Dunăre-a* outside the context of a prepositional phrase in this document, so it is not possible to confirm that this is a conditioned alternation. Nonetheless, on the basis of the available evidence, we can place determiner omission at the earliest point of attestation of Balkan Romance, as a feature that already obeyed the conditions associated with the structure in modern Romanian and Aromanian.

A greater wealth of data is available from the early Romanian document *Letopiseţul Ţării Moldovei* (1642~47) (Neculce and Iordan 1975). LDO applies in these documents with conditioned exceptions. Here it is possible to compare a number of occurrences of nouns, including the same nouns in definite and indefinite form, as they fall under the governance of prepositions. With the preposition *în* 'in,' there are 276 instances of nouns with a definite article, all accompanied by a modifying adjective, possessive pronoun, genitive noun, prepositional phrase, or relative clause. There are no instances of definite nouns occurring without an accompanying modifier. This is to be expected, since with LDO, only modified nouns can occur with definite articles within the domain of locative prepositions, so this is as expected. The situation is more complicated for the preposition *la* 'at.' Here, of 547 instances of *la* occurring with a definite noun, 20 include nouns that are unmodified. For example, the noun *împărat* 'emperor' occurs eight times as *la împărat-ul* 'to the emperor,' showing a definite article without a modifier in *împărat-ul* 'the emperor.' The exceptions occur exclusively among animate nouns that act as titles for individuals: *la împărat-ul* 'to the emperor,' *la paş-a* 'to the Pasha,' and *la crai-ul* 'to the king.' These exceptional instances in early Romanian express beneficiaries or sources of transference, as in the following examples.

EARLY ROMANIAN (LETOPISEŢUL ŢĂRII MOLDOVEI)

(92) El îl va   trimite la crai-ul
      he it will send   to king-M.ACC.DEF
      'He will send it to the king.'

(93) Oprescu năval-a   de   la împărat-ul
      they.stop onslaught-F.ACC.DEF from at emperor-M.ACC.DEF
      'They defend the emperor from the onslaught.'

The animate referent of the nouns in (92) and (93) permit the use of the definite article without any modifying elements. The Romanian in this document shows a distinction between animate recipients and physical locations in its use of the preposition *la* 'at.' Were *la* to designate a static location or goal of movement,[1] it would be expected that the determiner could not appear here, since the nouns are unmodified. This is what is attested for the vast majority of instances of *la* with an unmodified noun. Such a situation of an animate/inanimate split was preserved in colloquial Romanian into the 20th century (Pop 1948:422), although it is now archaic (Ciobanu and Nedelcu 2005:615). It is the difference in the semantics, the non-locative usage of these prepositions to govern nouns with animate referents, that allows for these eight instances of unmodified nouns to take definite determiners.

This pattern extends to another preposition with both locative and non-locative semantics: *pre* 'on.' While nouns following *pre* normally must appear in indefinite form in the absence of modi-

---

[1] As in all Balkan languages, movement to and location at are collapsed together into a single functional expression in Balkan Romance.

fiers, a definite article is allowed when *pre* is used as the marker for an animate, accusative direct object:

(94) să apere pre   crai-ul
      to defend ANM.ACC king-M.ACC.DEF
      'to defend the king'

A definite article with an unmodified noun governed by the animate accusative marker *pre*, now *pe* in modern Romanian, is no longer possible except with kinship terms. Nonetheless, the use of a definite article for animate, unmodified nouns governed by *pre* was a consistent feature of older Romanian texts. Locative and non-locative semantics of *pre*/*pe* were distinguished by the distribution of the definite article (Reichenkron 1962). That is, in its earliest attestations, Romanian showed a consistent pattern of LDO that depended on the same conditions that I identified for Albanian: accusative case selection, locative semantics, and an unmodified noun. When these three conditions were satisfied, definite reference was indicated by the omission of a definite article. Thus, the attestation of LDO in early documents dates the structure at least to the 16th century in Romanian.

LDO in other Balkan Romance languages

In chapter 3, I already established that LDO exists in modern Aromanian and Romanian. In what follows, I show that LDO is also a feature of Istro-Romanian and Megleno-Romanian. This establishes that LDO is a feature of every variety of Balkan Romance, even those spoken at a great geographical remove from Romanian. This potentially dates the development of LDO even further in the history of Balkan Romance, to the point before the fragmenting of Balkan Romance in the 11th and 12th centuries. This very early date for the shared innovation of LDO in Balkan Romance provides a point of comparison for the later development of the feature in Albanian and Balkan Slavic.

The primary separation of Balkan Romance languages is into languages originating from communities north or south of the Danube River, all of which maintained contact with one another until sometime between the Slavic migrations in the late 6th century and the first mention of 'Vlahs' in Byzantine sources in the 11th century. (Friedman 2001b:30). Megleno-Romanians and Aromanians derived from a single population of Balkan Romance speakers south of the Danube (Capidan 1925:65; Todoran 1977:102-109). Megleno-Romanians migrated to the Meglen region of what is now Greece and began diverging from Aromanian some time in the 12th or 13th centuries, with heavy Slavicization indicating that there had been sustained contact with Balkan Slavic speakers. The map in Figure 5.3 shows the outcome of these migrations.

Megleno-Romanian, spoken to the south of the Danube and carried to its present location by migration, also shows LDO with unmodified nouns. The semantics of its prepositions show significant influence from Balkan Slavic (Capidan 1925:182), but LDO remains a feature of the system of definite determiner distribution in Megleno-Romanian, despite the fact that LDO is not a feature of Bulgarian (see §4.3). This further supports the theory that LDO is an inherited feature of Megleno-Romanian that predates the split of Balkan Romance into its constituent languages.

Figure 5.3: Map of the Balkan Romance languages

(95)   Si   dusiră   pri pulată
       REFL they.bring to palace.F.ACC
       'They went to the palace.'   (Capidan 1925:182)

In (95) the preposition *pri* 'on, to' shares the semantic ambiguity as found for the Bulgarian preposition *na* 'on, to.' This is likely to be an influence of Bulgarian, as this ambiguity for *pri* (cognate with *pre* 'on' and *pe* 'on' in other branches of Balkan Romance) is not shared across Balkan Romance. Nonetheless, the definite reference of the unmodified noun *pulată* 'palace' is expressed by a bare noun without definite determiner, whereas in Bulgarian, definite reference would require a definite article. LDO in Megleno-Romanian is best explained an inherited feature from Balkan Romance.

Istro-Romanian, the fourth distinct branch of Balkan Romance, is derived from common Balkan Romance as was spoken north of the Danube, and diverged from Daco-Romanian through migration

to the Istrian peninsula sometime from the 11th to the 12th centuries (Todoran 1977:106-107, 175; Zegrean 2012:175). Istro-Romanian displays the fewest 'Balkan' features of Balkan Romance due to its long separation from the core zone of language contact in the southern Balkans (Friedman 2001b:27).

The attestation of LDO in Istro-Romanian, then, is the most significant piece of data for dating the appearance of LDO in Balkan Romance, since LDO could not have appeared in Istro-Romanian due to language contact in the Balkans at any time after the 11th∼12th centuries. Example (96) exhibits the expected behavior for a locative prepositional phrase governing an unmodified noun with definite reference: *su pamint* 'under the earth' shows no definite article despite the definite reference of *pamint* 'earth,' which contrasts with *cercec-u* 'the cicada,' which bears the definite article *-u*.

ISTRO-ROMANIAN

(96) Cercec-u   se bodit su   pamint
      cicada-M.NOM.DEF REFL he.hid under earth.M.ACC
      'The cicada hid under the earth.'   (Kurth 1904:88)

For comparison, the first Aromanian texts can be dated from the late 18th century and consisted of religious texts produced in the city of Moscopole in Albania, from a community of speakers centered on southern Albania and the Pindus mountain range in Greece (Caragiu-Matioțeanu 1968:4). Aromanian speakers are concentrated primarily in western Greece, southern Albania, and in the area of Bitola in the Republic of Macedonia, which resulted from a secondary migration back northward after the destruction of Moscopole (Golab 1959:425-426, Markoviḱ 2007:51).

All branches of Balkan Romance, including Megleno-Romanian, Aromanian, and Istro-Romanian, show determiner omission for unmodified nouns in locative prepositional phrases. The earliest attestation of Romanian already post-dates the period of divergence between (Daco-)Romanian and the other Balkan Romance languages (Zegrean 2012:175), and so the appearance of LDO in these other branches of Balkan Romance is significant. Istro-Romanian in particular did not participate in the latter stages of Balkan language contact between the 12th century and the 20th century, as it was territorially separated from the Balkan heartlands (Zegrean 2012:175, Wahlström 2015:91). Although it cannot be conclusively established due to the lack of sufficiently early attestations, it is reasonable to conclude that LDO was a feature of Balkan Romance that predates its separation into the languages of Daco-Romanian, Megleno-Romanian, Aromanian, and Istro-Romanian (Kurth 1904:8). LDO was inherited in these languages from a common origin in Balkan Romance and carried with them to the current, non-contiguous territories of their speaker populations.

LDO extends across the languages of Balkan Romance and obeys the general conditions expected for Balkan LDO: accusative-selecting prepositions with locative meaning trigger the omission of determiners for unmodified nouns with definite reference. This phenomenon must have already existed in Balkan Romance before the divergence of its constituent branches into Romanian, Megleno-Romanian, Aromanian, and Istro-Romanian, and thus can be dated to a period before the 11th century when the first migrations of Balkan Romance speakers separated these speakers territorially from one another.

Figure 5.4: Map of Albanian dialects

5.2 The diachrony of LDO in Albanian

The development of LDO in Albanian followed a more complicated route than that of Balkan Romance, with its uniform attestation. In Albanian, one must pay attention to dialect geography in tracking the spread of LDO as a grammatical feature of modern Albanian. Albanian dialects can be divided into two rough groups: a northerly group of dialects called Geg and a southerly group of dialects called Tosk. Between these dialect zones, there is a transitional set of dialects. In addition, two dialects broke off from Tosk and no longer participated with the main body of Albanian speakers after the 15th century. These dialects are known as Arbëresh (spoken in Italy) and Arvanitika (spoken in Greece), as depicted in Figure 5.4.

Of particular importance is the time span over which these changes happened and the dialect distribution of the shift from the original Albanian locative case to the modern LDO construction. In §5.2, I discuss the Albanian locative case, which shared the syntactic domains that govern LDO in the modern language (analyzed in §2.3). In §5.2 I show nevertheless that this case lacked important features of LDO such as the modification condition, the semantic restriction to solely locative prepositions, and the omission of the definite article. Over time, Albanian texts begin to show a loss of the locative case, but until the 19th century, this involved the elision of the locative case ending without concomitant loss of locative case agreement morphology. It was not yet a shift to LDO, as I argue in §5.2. In §5.2 I show that the dialects of Arbëresh and Arvanitika lack LDO and instead retain remnants of the locative case, thereby illustrating the independent outcome of this process. In §5.2 I contrastively argue that the main body of Albanian dialects underwent a significant change leading to the replacement of the locative case with LDO.

In §5.2, having established the path of the locative case and its divergence into LDO, I compare Albanian with Balkan Romance. Determiner omission only becomes a consistent feature of Albanian in the 19th and 20th centuries, with its earliest attestation in southern Tosk dialects. For northern Geg dialects, attestation is much more limited. Tosk dialects spoken in extensive contact with Aromanian and undergoing assimilation of Aromanian speakers develop fully grammatically integrated LDO. For northern Geg dialects, the loss of the locative case appears to have proceeded independently. A region of central Albania, in the transitional zone between Geg and Tosk dialects, continues to show sporadic attestation of the locative case until the mid-20th century. This seems like the distribution of a remnant, rather than an innovation. Yet LDO is clearly an innovation in Albanian. I account for the distribution of LDO in Albanian by attributing it to contact and assimilation of Balkan Romance speakers in the northern and southern range of Albanian. I argue that this contact was more intense and assimilation happened earlier in the southern Albanian dialects, whereas assimilation of Aromanian speakers did not occur until late and the absorption of LDO into the grammatical system of Albanian last occurred in isolated transitional dialects of Albanian.

Early Albanian

While it is clear that Albanian is an Indo-European language, occupying its own branch of the family that diverged relatively early from Proto-Indo-European (Byron 1976:31), the prehistory of Albanian remains murky. Knowledge of Albanian beyond what can be reconstructed from sound-change relationships with other Indo-European languages can be derived only from the point of its contact with Latin. It is common to posit Albanian as the direct descendent of one of the languages of the Balkans that predate the period of Roman colonization. The paleo-Balkan languages, as they are known, were diverse and very poorly documented (Woordard 2004:9-15). Most records of their existence come from ambiguous tribal designations in Roman documents, which may not have consistently matched up with languages spoken in the region. Some information about the languages spoken in the Balkans prior to Roman colonization can be derived from place names filtered through the currently attested Balkan languages. For example, Hamp (1982:79) concludes from a diachronic analysis of the Romanian place name *Dobreta* that an autochthonous language of the area possessed a definite article, and connects this with the form of the definite article found in Albanian, with the implication that a link of inheritance can be drawn between the language that supplied this place name and modern Albanian. But in general, toponymy is of limited utility in determining grammatical structure or lexical correspondences, both of which would be necessary

to determine a continuous lineage between modern Albanian and a Balkan language autochthonous to the region prior to Roman colonization (Friedman 2001b:29-30).

Despite these difficulties, scholars versed in the literature have made suggestions about possible paleo-Balkan lineages for Albanian, which would clarify the timeline and geographical range of the earliest contact between Albanian and Late Latin (later to become Balkan Romance). Illyrian or Thracian are forwarded as the primary candidates (Çabej 1971:42), with Illyrian having some scholarly consensus (Thunmann 1774:240, Kopitar 1829:85, Katičić 1976:184-188, Polomé 1982:888)—but there is a significant lack of verified inscriptions (Çabej 1971:41, Woodard 2004:11, Mann 1977: 1) and it is unclear whether 'Illyrian' as a term used in Roman records even referred to a single common language from which modern Albanian could descend (Hamp 1994). There is, however, evidence that Albanian was spoken over a broader territory in the Balkans than the contemporary range of territory occupied by its speaker community today (Çabej 1971:41, Demiraj 2004:98, 104). A number of important toponyms in Macedonia, southern Serbia, and Kosovo show reflexes of Albanian phonological developments; e.g. Astibos > Albanian Shtip, Slavic Štip (in eastern Macedonia), Naissus > Albanian Nish, Slavic Niš (in southern Serbia) (Pulaha 1984:11). The toponym *Dobreta*, because of its greater distance as mentioned above places, the potential range of the predecessor to Albanian up to the banks of the Danube. This was also the range of significant Latin influence (north of the Jireček Line, which is recognized as the customary division of a northern zone of Latin language influence and from a southern zone of Greek language influence, [Jireček 1911, Friedman 2001b:29]) and corresponded to the area inhabited by pastoral speakers of Balkan Romance well into the Middle Ages. This fits well with a theory forwarded in particular by Hamp (1994) that Albanian is the result of an autochthonous Balkan language that has undergone partial Romanization, while Balkan Romance represents a full language shift, whereby the early form of this Balkan language (its "proto-Albanian" linguistic predecessor) was fully absorbed into Late Latin.

Albanians as an identifiable ethnic group are first mentioned by the Byzantine historian Attakite, who identifies them as soldiers participating in a rebellion dating from 1043/1078 (Pipa 1989:200, Byron 1976:200). This establishes their presence as a distinct community, but does not reveal anything specific about their language. The first attested fragment of written Albanian dates from 1462, in a Latin-language letter containing a single-sentence translation of a baptismal formula into Geg Albanian (Mann 1977:1). In this text, features of modern Albanian such as the particle of concord and the post-posed definite article already appear. The earliest complete text in Albanian is the Missal of Gjon Buzuku, written in 1555, which contains various sections of the Bible in translation into Geg Albanian (Elsie 2005). A number of other written texts in Albanian appear from the 16th and 17th centuries, but documentation from the 17th to the 19th century remains very sparse. Early writing in Albanian appears primarily in Geg, which was spoken in mountainous areas of northern Albania that were difficult for the Ottomans to pacify. These areas of the Albanian-speaking range were thus subject to less repression of the Albanian language than southerly areas where Tosk Albanian was spoken (Konica 1957:112). Nonetheless, some examples of Tosk also appear among the early documents. Early Albanian authors include Matrënga (Tosk, Arbëresh - 1592), Budi (Tosk - 1618), Bardhi (Geg - 1635), Bogdani (Geg - 1685), and Kazazi (Geg - 1740). But only during the period of the Albanian literary awakening (the *Rilindas*), traditionally dated to the meeting of the League of Prizren in 1878, does a significant body of work in Albanian appear (Byron 1973:39). It is thus difficult to draw extensive conclusions about the range of grammatical structures in Albanian dialects before the late 19th century (Xhuvani 1957:151-152). Selections,

while available, are sporadic and specific to local dialects.

Nonetheless, it is possible to draw conclusions about aspects of the nominal system of Albanian prior to documentation. This is important for the establishment of LDO in Albanian, since the presence of particular case and definiteness markers interacted with the development of the later structural configuration of determiner omission. The expression of definiteness can be established for Albanian well before documentation. It is fully developed in the earliest attestation of Albanian (Orel 2000:247) and some authors argue that the period of the Albanian article's emergence can be confirmed before attestation through toponymic evidence (Hamp 1982:79). Topalli (2009a:378) makes a specific proposal, dating the development of the Albanian definite article to the period of contact with Latin, between the 4th and 6th centuries. This would accord with proposals that place the development of the demonstrative into a fully fledged definite article in Late Latin to the 6th century (Löfstedt 1956:373). Topalli argues that, while the article is old (dating from the time of the Roman occupation of the Balkans), it is not ancient (not present in the paleo-Balkan predecessor to Albanian). Others assert that it is not possible to make such definite conclusions about the period in which the article appeared in Albanian (Boretzky 1968:132, Hock and Joseph 2009:405, Wahlström 2015:147). Considering that Balkan Romance developed a similar post-positive article, which fails to align with the pre-posed definite article found in other Romance languages, and that Albanian and Late Latin speakers underwent extensive contact, it is plausible that proto-Albanian already possessed a post-positive definite article by the 6th century, with the invasion of the Slavs under the Avars, and that this was the source of the unusual configuration of post-posed definite article derived from the Late Latin demonstrative *ille* found in Balkan Romance. The lack of documentation prior to the 16th century, however, makes this impossible to determine with certainty.

The post-posed definite article was fully developed and integrated into the early system of Albanian case declension (Topalli 2009a:344), which involved considerable loss of inherited Indo-European nominal case and the reintroduction of new, agglutinated, phrasal case markers (Demi-raj 1993:70). The earliest attestations of Albanian had the same nominative, accusative, dative-genitive, ablative case system as observed in the modern language. But the configuration of definiteness marking with the case paradigms, and the marking of the cases themselves, differed from the modern language. In addition, there was a further locative case, expressed only in definite nouns, and used exclusively after prepositions (Topalli 2011:83).

Table 5.1 juxtaposes the early Albanian case system against its modern form using as an example the definite declension of the masculine noun *mal* 'mountain' (derived from Byron 1973:111-113).[2]

The locative case is now absent from the modern Albanian language. Its loss coincides with the development of LDO. However, the decay and elimination of the locative case did not automatically produce the grammatical, semantic, and morphological configurations that define LDO. Determiner omission of the modern pattern required a more complicated path of development from the locative case to modern LDO.

---

[2]There were some alternations not shown in this table, for example the genitive plural could also take the suffix *-et* rather than *-vet*. These are omitted here. It should also be noted that in Modern Albanian, dative/genitive plural does not show distinctions of definiteness.

Table 5.1: Definite declension of noun *mal* 'mountain' in Early and Modern Albanian

	Early Albanian		Modern Albanian	
	Sing.	Pl.	Sing.	Pl.
Nominative	mal-i	malë-t	mal-i	malë-t
Accusative	mal-në	malë-t	mal-in	malë-t
Dative/Genitive	mal-it	malë-vet	mal-it	malë-ve
Ablative	mal-it	malë-shit	mal-it	malë-ve
Locative	mal-t	malë-t	—	—

Comparison of the distribution of the locative case with LDO

The locative case is the most noteworthy feature of Early Albanian that has been lost in Modern Albanian. The locative case was strongly associated with the prepositions that trigger LDO in the modern language, but it did not obey the same syntactic or semantic constraints as LDO. The locative case only occurred in the definite and only within the domain of prepositions that select the accusative case (Elezi 2010:161-162). This set included all the prepositions that trigger LDO in modern Albanian, with the addition of the prepositions *pa* 'without' and *me* 'with,' which only trigger LDO when used indicate a means (Ajeti 1985:114, Karagjozi 2005:73, Demiraj 1988:375-376).

It is worth noting, here, that the locative case alternated with the accusative case, much as LDO in modern Albanian alternates with the accusative case in modified nouns. But the alternation of locative with accusative in early Albanian was not patterned. In particular, it did not depend on the modificational status of the noun: the locative definite article in addition to the accusative definite article could occur either with or without modifiers. In addition, the locative case could appear within the domain of any preposition that selected the accusative case. There was not a specific, semantic restriction that ruled out the use of the locative case with the prepositions *pa* 'without' and *me* 'with.' Since these prepositions selected the accusative, they could also select the locative. In this respect, the locative case in Albanian was an alternative to the definite accusative case, rather than existing in a grammatical pattern of alternation with it in the way that modern determiner omission alternates with the definite accusative. For indefinite nouns, only the accusative case was used, resulting from a restriction of the locative case only to definite nouns. As a result, simply losing the locative case could not have yielded the modern pattern of LDO.

When the locative acted as an alternative to the accusative case with modified nouns, the locative case required a distinct particle of concord, *të* (Demiraj 1988:378). Unlike Modern Albanian LDO, which occurs only with unmodified nouns, the locative could occur as a definite inflectional suffix either for unmodified nouns, or for nouns with modifiers such as adjectives or possessive pronouns, as in these examples:

(97) a. ndë mal-t
   on   mountain-M.LOC.DEF

   'on the mountain'   (Riza 2002:58)

b. mbë mal-t   të Liban-it
    on   mountain-M.LOC.DEF PRT Liban-M.GEN.DEF
    'on Mount Liban'   (Riza 2002:58)

In (97a), the noun *mal* 'mountain' is governed by a preposition *ndë* 'in, on' that denotes locative meaning and takes the accusative case in pronouns. Here, it selects the locative case, which is marked on *mal* with the suffix *-t*. The locative case appears on both unmodified nouns, as in (97a), and modified nouns, as in (97b). The modification status of the noun is not relevant to the distribution of the locative in Early Albanian, which shows a difference from the conditions that trigger LDO in modern Albanian.

The consensus among Albanian scholars is that the locative in Early Albanian was an alternative form of the accusative case that appeared only with a closed set of prepositions that otherwise selected the accusative (Demiraj 1988:378, Mulaku 2012:264, Elezi 2010:161-162). In this way, it could also be considered a prepositional accusative case. It is a newer case development in comparison to the older accusative (Demiraj 1993:7, 13). The locative never appeared with prepositions that select cases other than the accusative, nor did it appear in isolation without a governing preposition. There was no locative case form for pronouns. Prepositions that selected the locative in nouns selected the accusative in pronouns (Topalli 2011:84). Were the locative an inherited case from Proto-Indo-European, it would be expected to appear in pronouns as well. Despite its more recent development, the locative was widely distributed across Albanian dialects and documents (Demiraj 1988:377). It appeared in the works of Buzuku and Bardhi, authors who wrote in Geg (Byron 1976:197, Demiraj 2006:195-210) as well as those of Matrënga, an author writing in Tosk (Riza 2002:58). This dates the development of the locative case to a period after the development of definite articles in Albanian, which attached to case endings inherited from Proto-Indo-European. It also places the development of the locative case at the time when Albanian was still in intensive contact with Late Latin.

Why Albanian would employ a "prepositional accusative" case is unclear. It is an innovation in the Albanian case system, not directly inherited from Proto-Indo-European nominal morphology (Topalli 2009a:333), and it only applies to definite nouns. Pronouns never showed a distinct locative form and always appeared in the accusative case with the prepositions that selected the locative (Topalli 2011:83). There was nothing in morphology or phonology that would prevent the original accusative case from appearing on nouns as well. Even in the works of authors who employed it, the locative often alternates with the accusative case for both modified nouns and unmodified nouns. Riza (2002:59) observes that alternations between locative and accusative with the same prepositional governors can occur even within the works of the same authors:

(98)   a. me Jozef-në
           with Joseph-M.ACC.DEF
           'with Joseph'   (Buzuku, 1555)

       b. me të shënë Mëri-t
           with PRT holy Mary-F.LOC.DEF
           'with Mary'   (Buzuku, 1555)

While in (98a), the noun *Jozef* appears with the accusative definite suffix *-në*, another personal name, *Mëri* appears with the locative definite suffix *-t* in (98b). These examples occur with the same preposition, in the same text, by the same author. While the alternation of definite locative with definite accusative for the same prepositions in this text reinforces the point that the locative case was like an alternative, prepositional accusative case, the alternation found here is not the same pattern of alternation found in modern Albanian with LDO. However, Topalli (2009b:323-324) asserts that already at this time, a tendency was emerging: when nouns appear with *me* 'with' and the prepositional phrase designates a means, definiteness is expressed by the locative case. When *me* 'with' designates accompaniment rather than means, definiteness is expressed by the accusative case. These are the same conditions that determine whether LDO occurs with the preposition *me*. Topalli argues that this is also the case for the preposition *për* 'for,' in which the differentiation of meaning between locative and non-locative semantics corresponds to a differentiation in the expression of definiteness by case and article, or lack thereof.

(99) a. Ecëni ëndë Betlem-t   e   kërkoni mirë për djalë-t.
          go   to   Bethlehem-M.LOC.DEF and search well for boy-M.LOC.DEF
          'Go to Bethlehem and look carefully for the boy.'   (Matthew 2:8, Buzuku, 1555)

      b. Ëndë Betlem-t   të   Jude-së,   për-se ashtü anshtë shkruom për
          in   Bethlehem-M.LOC.DEF M.LOC.DEF Judea-F.GEN.DEF for-that so   is   written for
          profetë-në.
          prophet-F.ACC.DEF
          'In Bethlehem of Judea, for so it has been written by the prophet.'
          (Matthew 2:5, Buzuku, 1555)

In (99a), the preposition *për* designates a goal in *kërkoni mirë për djalë-t* 'look carefully for the boy.' The object of the preposition in this instance, *djalë-t* 'the boy (LOC)' takes the locative case. But only a few lines before, the preposition *për* designates the agent of a passive voice verb in *ashtü anshtë shkruom për profetë-në* 'so it has been written by the prophet.' In this instance, the object of *për* takes the accusative, definite case. The exceptional behavior of *me* and *për* with regard to the locative case parallels the exceptional behavior of these prepositions with regards to LDO in modern Albanian. Nonetheless, the alternation was sporadic, with the locative better seen as an alternative to the accusative case within the domain of prepositions that selected the accusative. The locative case lacked the modification condition that is a key trigger for modern LDO.

Elision of locative case morphology over time

The transition from locative case, where the modification status of the noun is not a factor in distribution, to LDO, where modification is a key grammatical condition, was not immediate. Intermediate stages involved the loss of the locative case ending, but the retention of the locative case particle of concord *të* for adjectives and genitival modifiers. Thus internal developments in Albanian attestations prior to the 20th century do not show LDO. Rather, they show morphological reduction of the locative case. Another step was necessary to transition from locative case to the pattern of LDO found in modern Albanian and shared with Balkan Romance.

Already authors in early Albanian tend to elide the locative ending in the presence of the particle of concord, *të*, resulting in a construction where the noun appears to take accusative indefinite form, but in fact has definite meaning (Topalli 2009a:322).

(100) në mal　　　　të　　　　Tomor-it
　　　　on mountain-ø ACC.IND/(LOC?) Tomor-DAT.M.DEF
　　　　'on Mount Tomor'

Rather than the expected structure *në mal-in e Tomor-it*, with a definite accusative article to correspond with the definite reference, or *në mal-t të Tomor-it* with a definite locative article and locative particle of concord, *mal* in the above example is found in the indefinite accusative morphologically. This shows a syntactic reconfiguration *në malt të Tomor-it > në mal-ø të Tomor-it* in which the geminate sequence of the locative case ending *-t + të* of the pre-posed locative adjective article was reinterpreted as an accusative, indefinite *të*, while the elided locative ending of the noun was interpreted as a bare noun in the accusative case. The particle of concord *të* could represent either indefinite accusative or definite locative, blurring the distinction between these two case/definiteness paradigms. In particular, this construction would result in a morphology that resembled indefinite accusative, but represented a definite meaning, triggered in the domain of locative prepositions. This is closer to the modern construction of LDO, but remains different enough that further grammatical developments were necessary to reach the point of LDO. This stage is better interpreted as the phonologically conditioned elision of a morphological locative ending in the presence of the locative particle of concord, rather than incipient LDO.

Figure 5.5 schematizes the transition involved in this elision. In its earliest state, Albanian possessed a set of prepositions that could select either the locative (loc) or the accusative (acc) case for definite nouns. When these nouns occurred with a modifier such as an adjective (A), number (num) and gender (gen) agreement would be expressed on the noun (N), its suffixed definite article (=D), and the adjective. The particle of concord, when present, would express agreement with either definite accusative or definite locative case. With elision of the definite locative case, a new state of affairs was available. While definite accusative case remained expressed in all agreement positions as expected, the locative case could now be phonologically elided. Thus, while definiteness was no longer overtly expressed with a locative suffix *-t*, the presence of the locative particle of concord *të* would still indicate the presence of a null locative case ending, and thus that the noun should be interpreted as definite, as indicated by the brackets. Solid lines indicate direct morphological expression, while dotted lines indicate interpretation without a morphological reflex. Arrows indicate the outcome of grammatical change.

Such elided locatives appear in some later Albanian documents that predate the modern period, such as in Kristoforidhi's Bible translation of 1872. Another example of this shift comes from translations of the Bible from different periods of Albanian. The New Testament translation of Konstantin Kristoforidhi, produced in 1872 before the full retreat of the locative case (Kristoforidi 1872), provides an example of the above construction in which a null accusative appears to be accompanied by a preserved locative adjectival article, as in (101).

(101) Eδe Iesu-i　　　　mbassi l'ēu ndе Beθlehem　tе　　　　Jude-sе
　　　　and Jesus-NOM.M.DEF held　birth in　Bethlehem.M ACC.IND/(LOC?) Judea-DAT.F.DEF
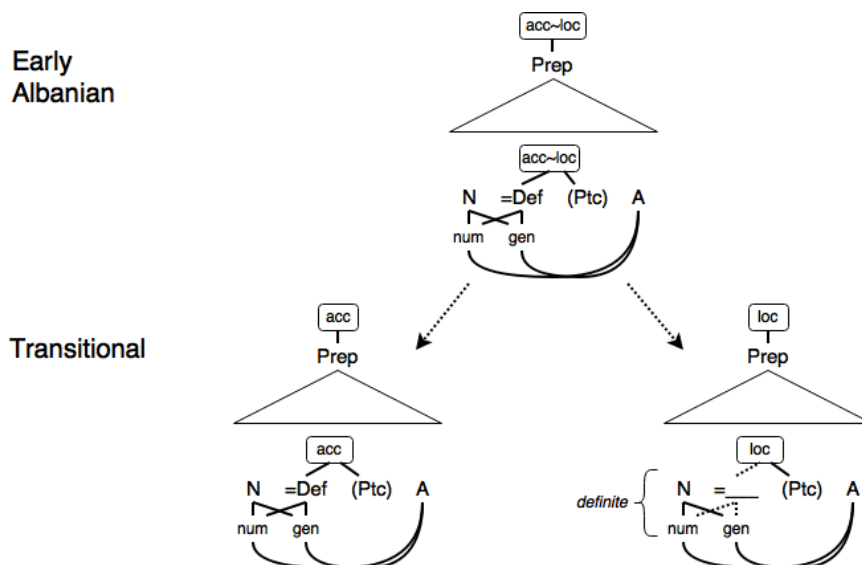　　　　'and Jesus was born in Bethlehem of Judea'

Figure 5.5: The transitional state of the Albanian case agreement paradigm for modified nouns with the elided locative case

Compare this with a modern translation of the Bible into standard Albanian in (102), in which the modification of *Bethlehem* by the possessive noun *Jude-së* requires an accusative definite article, as expected by the modern pattern of determiner omission. The text in (101) gives evidence for a transitional stage between the loss of the archaic locative and the full reconfiguration of locative prepositions into accusative-selecting elements that trigger determiner omission only for bare nouns.

MODERN ALBANIAN

(102) Pasi Jezus-i   lindi   në Bethlehem-in   e   Jude-së
       after Jesus-NOM.M.DEF was.born in Bethlehem-ACC.M.DEF ACC.M.DEF Judea-DAT.F.DEF

       'After Jesus was born in Bethlehem of Judea'

The locative was not entirely absent from Kristoforidhi's translation. Other parts of the document show a residual locative case, which demonstrates that examples such as (101) were a matter of the elision of a category that was still present, although sporadically, in the Tosk dialect to which Kristoforidhi was translating (Mann 1977:102).

(103) a. nd-atë   udhë-t
           on-that.ACC road-F.LOC.DEF

           'on that road'

       b. mbë tokë-t
           on   earth-F.ACC.DEF

           'on the ground'

At this stage of documentation, even in the late 19th century, works written in Albanian retain a locative case. But it could be elided in the presence of the locative particle of concord *të*, which

resulted in a construction that looked like indefinite accusative, but retained a definite meaning. The locative case ending still appeared with unmodified nouns. This is not yet the same construction as LDO, where definite accusative appears with modified nouns and indefinite accusative appear with unmodified nouns. Instead, what we have here is nearly the reverse situation. Even late in the development of the Albanian case and definiteness paradigm, the locative case still predominated. The transition to LDO must have happened later, in the late 19th century to early 20th century.

Geographically non-contiguous Albanian dialects and the absence of LDO

This idea that developments of the locative case within Albanian did not directly yield LDO is supported by evidence from two dialects spoken outside the main Albanian linguistic area, which offer important contrasts with contiguous dialects of Geg and Tosk. Because of their geographic isolation, these non-contiguous dialects were not subject to the same influences of Balkan multilingualism that affected Geg and Tosk.

One of these diaspora dialects, Arbëresh, is spoken by communities of Albanians in Italy. It derives from the southern Tosk dialects spoken by refugees fleeing the Balkans during the 13th and 14th centuries, following the death of the national hero Skanderbeg and the Ottoman reprisals for his rebellion (Altimari 1994, Fine 1987:602). The other diaspora dialect, Arvanitika, also derives from southern Tosk; it is spoken by a community of Albanians in Greece, in the area of Athens (Hamp 1971:665-666). After their speakers migrated from Tosk-speaking areas of what is now modern-day Albanian to southern Greece in the 14th and 15th centuries (Alibali 2013:138), these dialects were no longer in contact with the main body of the Albanian speaking community. In the case of Arbëresh, the main language of contact, if any, was Italian. Arvanitika continued contact primarily with Greek. Both dialects ceased, after migration, to occupy the particular multilingual milieu of contact between Albanian, Balkan Romance, and Balkan Slavic that affected the Albanian-speaking community along the Adriatic and into Macedonia.

That the locative had a wide dialectal base at the earliest era of documentation is also attested by the fact that it remains in use in Arbëresh, the diaspora dialect spoken in Italy, which split from the general body of Albanian speakers during the Middle Ages and failed to participate in several subsequent changes that were generalized across most or all Albanian dialects spoken in the Balkans. At the earliest period of Albanian documentation, and in dialects removed from the influence of general Albanian relatively early, the locative shows itself as a pervasive and integrated part of the case system.

For instance, Arbëresh, spoken in Italy, shows a straight-forward preservation of the locative case, as in the following example:

ARBËRESH ALBANIAN

(104) ndë mal-t
       on mountain-M.LOC.DEF

       'on the mountain'   (Çamaj 1984:32)

This is equivalent to a direct attestation from the state of early Albanian and gives further evidence for the presence of the locative throughout both Geg and Tosk dialects before the late 19th century.

The diaspora dialect of Arvanitika, shows an intriguing pattern in which the locative case particle of concord remains for an adjectival modifier of a noun within a prepositional phrase that has

definite reference, in which the noun also takes the definite accusative case marker. This results in a 'mixed state,' with a definite article from one newer case paradigm and an particle of concord from another, older paradigm, as in these examples drawn from (Sasse 1998:40).

(105) ARVANITIKA

   a. U jam kathijití   i   ghlosolojí-së   ndë panepistimio-n   të
      I am professor.M.NOM M.NOM linguistics-F.GEN.DEF at   university-ACC.DEF LOC
      Kollonji-së   ndë Jermanji   (Sasse 1998:40)
      Cologne-F.GEN.DEF in   Germany-F.ACC

      'I am a professor of linguistics at the University of Cologne in Germany.'   (Sasse 1998:40)

      STANDARD ALBANIAN

   b. Unë jam profesor   i   gjuhësi-së   në Universitet-in   e
      I   am professor.M.NOM M.NOM linguistics-F.GEN.DEF at university-M.ACC.DEF M.ACC.DEF
      Këln-it   në Gjermani
      Cologne-M.GEN.DEF in Germany-F.ACC

      'I am a professor of linguistics at the University of Cologne in Germany.'   (Sasse 1998:40)

In (105a) we see a modified noun phrase governed by a locative preposition *ndë panepistimio-n të Kollonji-së* 'at the University of Cologne,' in which the noun *panepistimio* (borrowed from Greek) takes the masculine accusative definite article *-n*, as would be expected in Standard Albanian. This would correspond to the configuration for LDO, since the presence of a modifier triggers the appearance of the definite, accusative case article. The expected particle of concord, however, would be definite, accusative *e*, whereas in (105a) we see the locative particle of concord *të*. This is explicitly compared against (105b) in Sasse (1998), where *në Universitet-in e Këln-it* 'at the University of Cologne' shows both the masculine accusative definite article and the accusative definite particle of concord *e*. Otherwise, Arvanitika matches Standard Albanian in its case usage, including in the omission of the accusative definite article with *ndë Jermanji* 'in Germany,' exactly the phenomenon of LDO that we are considering. Speakers consistently drop the definite article (either locative or accusative) for unmodified nouns. This is also exhibited in the following examples, one of which is drawn from a fluent speaker and one of which is drawn from a semi-fluent speaker (Tsitsipis 1998:47).[3]

(106) ARVANITIKA (FLUENT)

   a. Një núse   e   re   ka árdhur në katúnd
      a   bride.F.NOM F.NOM young.F.SG has come to village.M.ACC

      'A young bride has come to the village.'   (Tsitsipis 1998:47)

      ARVANITIKA (NON-FLUENT)

   b. Nji re   núsa ka árdhur në katúnd
      a   young.F.SG bride.F has come to village.M.ACC

      'A young bride has come to the village.'   (Tsitsipis 1998:47)

---

[3] Arvanitika, now in a state of endangerment, is no longer consistently spoken throughout its traditional community.

Both speakers use the indefinite accusative form *katund* 'village' within the domain of the locative preposition *në* 'to, in,' a usage which is consistent across texts for unmodified nouns. Whether speakers show a fully Albanian grammatical system as in (106a), or a grammatical system affected by interference with Greek as in (106b), definiteness within a locative prepositional phrase is still expressed by the absence of a definite article if the noun is unmodified. Unfortunately, Arvanitika is a highly endangered language, with relatively poor documentation and few remaining speakers (Hamp 1989:199). It is also subject to collapse and consolidation in its case system, beyond that observed in standard Albanian. Non-fluent speakers of the language have lost consistent use of articles of concord, regularly confusing the appropriate articles as opposed to the correct article use seen in the speech of elderly fluent speakers (Tsitsipis 1998:51-53). That is, the use of a particular article in any given prepositional context, or the presence or absence of case forms, does not indicate a stable re-evaluation of the case system Moreover, even fluent elderly speakers appear to have lost the distinction between the definite accusative and indefinite accusative articles of concord, using the indefinite accusative *të* across the whole case paradigm (which would render the particle of concord system for the Arvanitika accusative indistinguishable from the locative). In the example below, the particle of concord is *të* despite the presence of an accusative definite article, which is also the case in the first Arvanitika example (105) above (Tsitsipis 1998:46).

(107) Atë   klíc-in   të dér-ës   e móra u
      this.ACC key-M.ACC.DEF ACC door-F.GEN.DEF it I.took I
      'I took this key to the door.'   (Tsitsipis 1998:46)

Given that Arvanitika speakers consolidate the definite accusative *e* and indefinite accusative *të* into a single form *të*, the scope for variation in Arvanitika locative prepositional phrases is even greater. This differs from the variation shown in documents throughout the period of development from earliest recorded Albanian to modern Albanian, where the use of definite locative, bare accusative, or definite accusative within locative prepositional phrases was not consistent, but the agreement paradigm for any one case usage was comparable to its agreement paradigm outside of locative prepositional phrases. Arvanitika exhibits the variation found in earlier Albanian documents, but a variation that is unstable across the entire case system. There is, in any case, lack of consistent LDO in Arvanitika.

Thus, the development of consistent and grammatically conditioned LDO in contiguous dialects of Albanian must have post-dated the departure of diaspora dialects such as Arbëresh and Arvanitika from the main body of Albanian speakers. As exhibited by these examples, the locative case was never in full consistent usage across all early Albanian authors. Rather, it was an alternative construction to the accusative, which appeared within the domain of a restricted set of prepositional phrases. Since it was present in Arbëreshë (Çamaj 1984:32), and it was found in both Geg- and Tosk-speaking writers, it must also have existed in the earlier period of unity in the Albanian speaking community (Topalli 2011:85). That is, the disappearance of the locative from Albanian and the transition to a modern pattern of LDO post-dates the early history of the Albanian language. And since these main dialects of Albanian still showed variable attestations of the locative into the late 19th century, the development of LDO must have been relatively recent, coinciding with the end of the Ottoman Empire and the establishment of Albanian nationhood in the early 20th century.

The grammatical integration of LDO in Albanian

LDO of the modern variety—in which the accusative case is universalized across the entire paradigm of nouns governed by accusative prepositions, and the alternation is only in the appearance of the accusative definite article—begins to appear in the late 19th century, in the works of some writers of the Rilindas (movement for national renewal) beginning in 1878 (Byron 1976:39). While early Rilindas authors such as Sami Frashëri make consistent use of the locative case, in what appears to be a conscious appeal towards older forms of Albanian, others begin to avoid the locative case entirely and instead opt for expressing modified nouns with a definite accusative case article and unmodified nouns with no article at all within the domain of accusative prepositions (Riza 2002:61). This movement towards a reduced case inventory was not limited to any one dialect. Authors from both Tosk and Geg speaking communities showed a preference for dropping the locative case (Byron 1976:41-42, 114) as part of a general pattern of case loss and reconfiguration. In addition to the loss of the locative, both dialects also experienced the loss of the definite ablative plural and the definite dative/genitive plural. The former was syncretized with the dative/genitive plural, while the latter was merged with the indefinite dative/genitive plural to yield a situation of no definiteness distinction in the dative/genitive plural. Definiteness remains distinctive in the singular for prepositions that select dative/genitive, and one cannot see any synchronic tendency towards its omission synchronically from the period of the Rilindas to that of the modern language. The phenomenon of the development of LDO did not affect prepositions selecting dative, genitive, or ablative case.

Evidence for the loss of the locative and the introduction of systematic LDO for accusative prepositions is increasingly more common from the first citations in the texts of the Rilindas toward the introduction of the standardized literary language in the mid-20th century (Boretzky 1968:130), but the loss of the locative appears in both Geg and Tosk dialect documents (Boretzky 1968:150). These modern Albanian documents show some important tendencies in regards to the semantic and syntactic conditions that seem to trigger a shift away from locative case and toward LDO. In particular, prepositions which once selected either the locative case or the accusative case begin to select only the definite accusative. These prepositions are *me* 'with' and *për* 'for,' which are primarily used to denote relations that are not locative.

First, there is a diachronic tendency across texts before the mid-20th century for the preposition *me* to take a definite accusative article with both modified and unmodified animate nouns. As seen in §5.2, early Albanian allowed the locative to appear with *me*, but by the period of the Rilindas, the preference was for a uniform accusative article across nouns designating animate referents. Thus, the use of locative case as in *me djalë-t* 'with the boy' or determiner omission as in *me djalë* 'with the boy,' which still appear in the oldest Albanian texts, become dispreferred, and definite accusative case, as in *me djalë-n* 'with the boy,' becomes the norm (Boretzky 1968:136-138), see Table 5.2 for a depiction of this development.

Table 5.2: Change in the behavior of *me* 'with' and case/definiteness selection in Albanian

Early Albanian	Modern Albanian
me djalë-t (loc. def.)	
or	*me djalë-n* (acc. def.)
me djalë (acc. ind.)	

Moreover, other prepositions that can denote either concrete spatial relations or abstract, associative relations, such as *mbi* 'on, about' *për* 'for, about, toward,' or *nënë* 'under,' also become associated with a preference for a uniform usage of the accusative definite article with abstract, non-locative semantics. Particularly when the nouns they govern refer to a person, use of the locative case or omission of the determiner becomes less and less common. Thus *për djalë* 'for the boy' with determiner omission becomes less preferred than *për djal-in* 'for the boy' with the accusative definite article (Boretzky 1968:138-139, 142). In circumstances where these prepositions govern animate nouns, they are particularly likely to denote abstract relations and not concrete meanings of location or direction of movement: *për djal-in* is much more likely to mean 'for the boy' or 'about the boy' than 'toward the boy.' It is the abstract semantics in the use of these prepositions that determines the transition away from locative and towards a uniform definite accusative article, regardless of the modification status of the noun.

The behavior of accusative-selecting prepositions such as *me* 'with' or *për* 'for, about, toward' can be contrasted with that of prepositions such as *në* 'in, on, to' which only designate concrete locational or directional relations. By the early 20th century, concrete locative prepositions tend to lose the accusative definite article with unmodified nouns (Boretzky 1968:147). This is equivalent to a loss of the locative case with unmodified nouns, first sporadically and then consistently, represented in Table 5.3.

Table 5.3: Change in the behavior of *në* 'in/on' and case/definiteness selection in Albanian, unmodified nouns

Early Albanian	Modern Albanian
në mal-t (loc. def.)	
or	*në mal* (acc. ind.)
në mal (acc. ind.)	

In addition, those noun phrases modified by an element that does not take a particle of concord like a prepositional phrase or a subordinate clause, begin to take the accusative definite article instead. This involves a loss of the locative definite article and its replacement with the accusative, but no change to the locative particle of concord, which is not present in these constructions.

Table 5.4: Change in the behavior of *në* 'in/on' and case/definiteness selection in Albanian, modifiers without particle of concord

Early Albanian	Modern Albanian
në mal-t afër fshatit (loc. def.)	
or	*në mal-in afër fshatit* (acc. ind.)
në mal afër fshatit (acc. ind.)	

Finally, the category of adjectives, which only sometimes show a particle of concord, begins to exhibit a definite accusative agreement pattern rather than a definite locative pattern. Elided locatives alternate with the use of the definite accusative across the system of modified nouns. Those nouns taking genitive modifiers, which must always occur with a particle of concord, begin

to appear with the accusative definite article along with the definite accusative particle of concord, and thus the LDO configuration begins to hold across the noun system. This is schematized in Figure 5.6, where the option of elision of the locative in the presence of modifiers collapses into a single system in which prepositions that once governed either the locative or the accusative now trigger only the definite accusative for modified nouns with definite reference.

Table 5.5: Change in the behavior of *në* 'in/on' and case/definiteness selection in Albanian, modifiers with particle of concord

| Early Albanian	Transitional	Modern Albanian
në mal-t të Tomorit or në mal të Tomorit	në mal të Tomorit or në mal-in e Tomorit	në mal-in e Tomorit
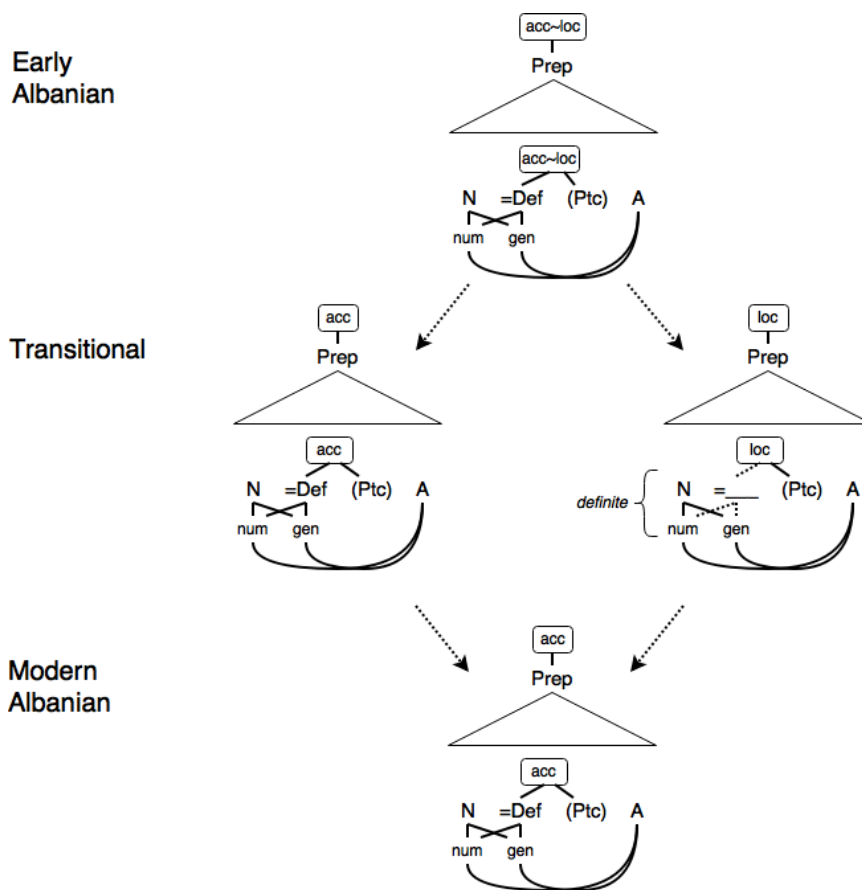
Figure 5.6: The outcome of the Albanian case agreement paradigm for modified nouns with full loss of the locative case

As a result of this change, prepositions governing modified nouns that once selected either the locative or accusative case now selected just the accusative case. Bare nouns, without modifiers,

were subject to a different development. Two transitions were key for the development of LDO and the loss of the locative case. This is schematized in Figure 5.7. First, accusative-selecting prepositions denoting non-locative meanings (like *me* 'with' or *për* 'for') became a distinct domain from accusative-selecting prepositions denoting locative meanings (like *në* 'in, on'). While the non-locative prepositions began to take the accusative definite article for both modified and bare nouns, the locative prepositions took the accusative definite article only for modified nouns, while unmodified nouns appeared bare, without a determiner. Whereas earlier in the Albanian system, the locative could appear with either set, now determiner omission only occurred with the locative set of prepositions.

Second, the locative particle of concord was lost from the regular system of Albanian. Whereas previously a locative ending in -*t* could be elided before a locative particle of concord *të*, now the definite accusative case ending -*n* with its definite accusative particle of concord *e* was substituted where the locative case had once occurred, conditioned by semantic restrictions on the governing preposition and syntactic restrictions on the noun phrase.
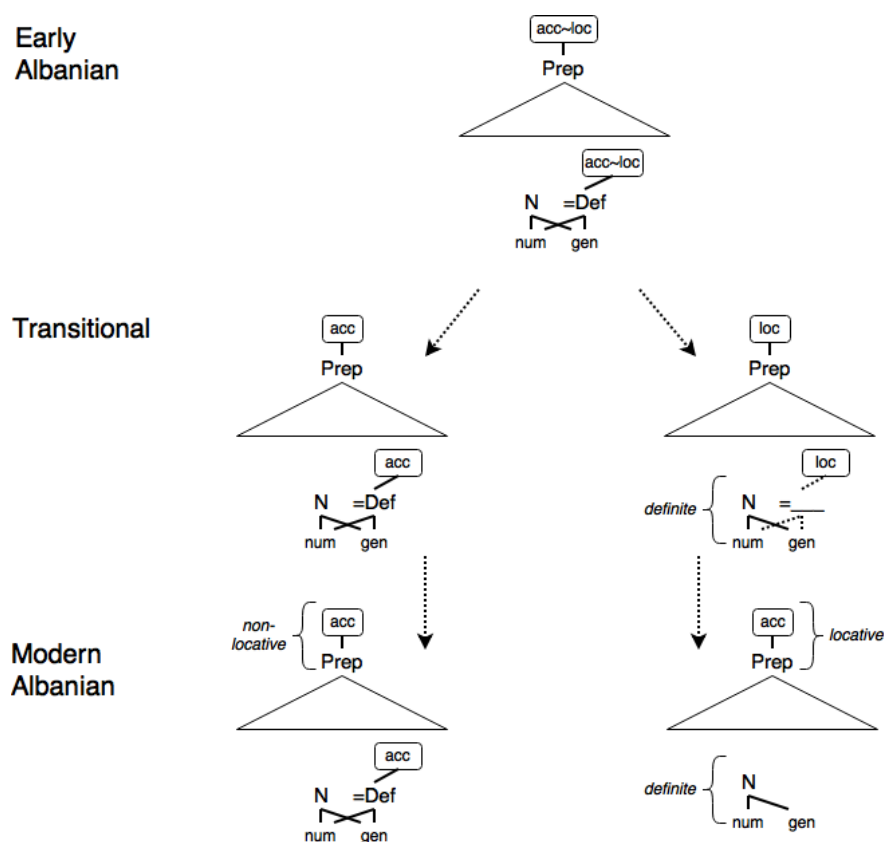
Figure 5.7: The outcome of the Albanian case agreement paradigm for unmodified nouns with locative determiner omission

The full loss of the locative and systematization of LDO is widespread by the time of the introduction of the standard Albanian language, following 1944. The standard language is based primarily on the southern Tosk dialect (Byron 1973:73-75). In the standard language, the locative was eliminated entirely and replaced by an alternation of bare, accusative case marking for express-

ing definiteness in unmodified nouns and the accusative definite article for expressing definiteness in modified nouns. The locative case persisted sporadically in scattered spoken dialects located in a belt transitioning from northern Tosk to southern Geg, in the center of the country of Albania (Byron 1976:48), bounded by the Mat and the Vjosa rivers (Demiraj 1988:377, Gjinari and Shkurtaj 2003:216, Cacaj 1999:76)—for more discussion, see Transitional dialects in (5.2). Since the late 20th century, the locative has disappeared entirely from the contiguous territory of Albanian speakers, and LDO now predominates throughout.
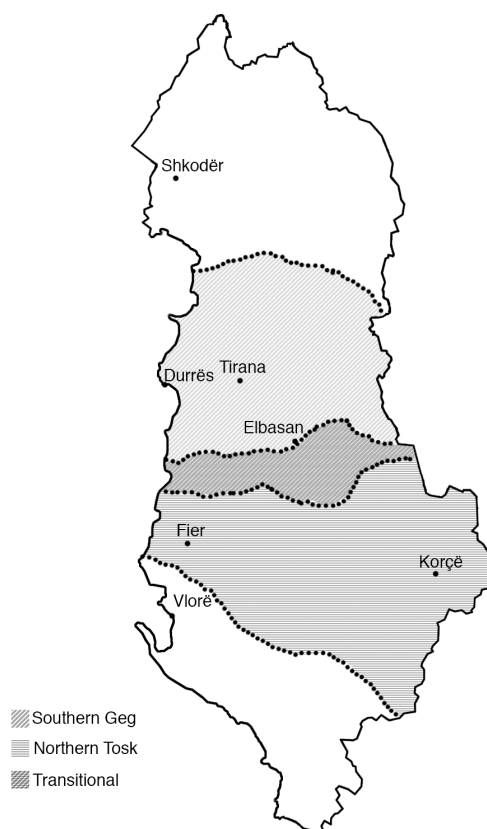
Figure 5.8: Map of the zone of sporadic locative case attestation in northern Tosk and southern Geg dialects in the mid-20th century

In southern Tosk and northern Geg speaking regions, the locative disappeared entirely from spoken dialects (Gjinari 1965: 245, 254, 1966:54), with the result that the innovation of locative loss appears at the northern and southern ends of contiguous Albanian speaking territory, while the dialects in the middle preserve pockets of conservatism regarding the locative case somewhat longer. This is an unusual pattern for an innovation, and yet locative loss is without question an innovation in Albanian grammar, as demonstrated in §5.2-§5.2. Its replacement with LDO, however, required a further step. Systematic documentation of northern Geg dialects only truly begins in the 20th century, which makes it difficult to tell exactly what the state of LDO development in

these dialects was until this point. Tosk dialects, better attested, make a clear transition to LDO between approximately 1875 and 1950.

Although the loss of the locative could be interpreted as part of the general pattern of case reduction and reanalysis that applied across the Albanian case system in the modern period, one would have expected, based on earlier patterns in Albanian grammar, that both modified and unmodified nouns would express definiteness with an indefinite accusative reanalyzed from the old locative case. This is the pattern that is found in non-contiguous Albanian dialects such as Arvanitika in Greece and Arbëresh in Italy, as analyzed in §5.2. Two transitions, one regarding semantic conditions, and the other regarding syntactic conditions, were necessary to yield LDO in the contiguous Albanian dialects, and these transitions did not occur in Albanian dialects that separated from the main body of the Albanian speaking community before the late 19th century.

What occurred in the loss of the locative case was not just a reinterpretation of the case selection paradigm for locative prepositions, but also the introduction of a new grammatical alternation for the expression of definiteness. Whereas the locative case was a marker of definiteness for both modified and unmodified nouns under the old Albanian case paradigm, in the new paradigm, a construction that looks like the indefinite accusative case expresses definite reference only for unmodified nouns, while the accusative definite article plays this role for modified nouns. The use of the indefinite accusative for unmodified nouns can be easily derived as a reinterpretation of the old locative case, in which the case ending *-t* is dropped, yielding a form identical to the unmodified particle of concord. Were this reinterpretation to extend across the entire case paradigm, however, we would expect to see the dropping of the old locative case marker extended to modified nouns, with the locative case particle of concord *të* reinterpreted as the indefinite accusative particle of concord (also *të*). This would result in an elimination of definiteness distinctions for nouns within the domain of locative prepositions. Such a state of affairs pertains in the diaspora dialect of Arvanitika, where the particle of concord remains *të* within the domain of locative prepositions, whether or not an accusative definite case marker is present. Moreover, Albanian has eliminated definiteness distinctions in other restricted case paradigms, such as the dative and genitive plural. Indeed, it is precisely a desinence in *–t* that would be expected to be lost entirely based on other changes to the Albanian system. Both the dative/genitive definite article and the locative definite article involved a suffixed *–t* followed by a particle of concord with the form *të*. An internally motivated loss of the locative case would be expected to yield accusative selecting prepositions that failed to morphologically distinguish definiteness for either unmodified or modified nouns, as a result of the loss of the locative desinence.

In the main Albanian dialects, however, elision of the locative ending *-t* and reinterpretation of definite locative as indefinite accusative applies only to unmodified nouns. Modified nouns express definiteness with the accusative case and the corresponding accusative particle of concord *e*. In addition, the accusative definite article begins to be used across all nouns governed by accusative selecting prepositions with non-locative semantics. This pattern of that which is here called LDO can be clearly documented as predominant in southern Tosk dialects over the course of their development into the basis of the standard language from the late 19th century to the mid-20th century. LDO was also found to be already predominant in northern Geg dialects at the time of their documentation by the mid-20th century. Only sporadically, in isolated transitional dialects between Geg and Tosk, was the locative case still found by the mid-20th century. In current Albanian speech, LDO predominates throughout all dialects in the contiguous speech area of Albanian.

Albanian retained a distinctive pattern of case distribution that applies only to nouns within

the domain of locative prepositions, but transitioned from the use of a special case marker *-t* to the pattern of LDO, i.e. an alternation between the bare accusative and the definite accusative. In maintaining a definiteness distinction for nouns in locative prepositional phrases, Albanian transitioned from a state in which locative prepositions triggered special case treatment for the nouns they governed (replacing the default accusative case with a distinct locative case for nouns in constructions with locative semantics alone), and arrived at a state exactly parallel to that of Balkan Romance. This process happened during the late 19th and early 20th century, a period when Aromanian assimilation into surrounding speaker populations of Albanians, Greeks, and Macedonians was beginning to pick up pace. In order to understand LDO in its broader Balkan context, therefore, it is necessary to trace the path of the development of LDO in Balkan Romance and comparing it against this development in Albanian.

Comparison of LDO in Balkan Romance and Albanian

The similarities between Balkan Romance languages and Albanian in their behavior with regards to determiner omission and prepositions are striking. These similarities are most extensive when comparing Aromanian with Albanian. Table 5.6 shows the extent of these similarities.

Table 5.6: Determiner omission in Aromanian and Albanian compared

Aromanian determiner omission	Albanian determiner omission
• restricted to accusative-selecting prepositions	• restricted to accusative-selecting prepositions
• *cu* 'with' and *ti* 'for/about' show exceptional behavior	• *me* 'with' and *për* 'for/about' show exceptional behavior
• modification of the noun by any constituent suppresses omission	• modification of the noun by any constituent suppresses omission
• all triggering prepositions have locative meaning	• all triggering prepositions have locative meaning

The similarities between LDO in Balkan Romance and in modern Albanian would seem to find a straightforward explanation in contact-induced convergence, given that Albanian and Balkan Romance have a long history of contact, dating from before the historical record of attestation (cf. §5.1). LDO in Balkan Romance is an embedded grammatical pattern with striking similarities in syntactic distribution and conditions to Albanian.

One of the important indicators of a deep, long-term relationship of contact between Albanian and Balkan Romance is their lexical similarities. Thus, one might try to explain the similarity between these two languages by positing lexical borrowing, carrying along with it the grammatical conditions for determiner omission from whichever language was the source. It is important to note that the similarities between Romanian and Albanian in terms of determiner omission are structural, rather than lexical. While the case-selection and modification conditions are the same, and the two languages also match with respect to the exceptional behavior of accusative-selecting prepositions with non-locative meaning, there is no direct lexical correspondence between Romanian and Albanian in regards to the prepositions that cause determiner omission themselves. As can be seen in Table 5.7, prepositions with similar meaning have different case selection requirements and different derivations in Romanian and Albanian. The vast majority of Romanian prepositions are directly derived from Late Latin. Lexical calquing from Romanian to Albanian, or vice versa, does not provide an immediate explanation for the shared grammatical structure of LDO. Rather, it is the combination of structural triggers—accusative case-selection, locative meaning, and an unmodified nominal constituent—that condition determiner omission in both language groups.

Table 5.7: Comparison of Romanian and Albanian prepositions (with etymologies)

Romanian	Albanian
• *în* (< Lat. *in*) + ᴀᴄᴄ 'in, on, to,' *la* (< Lat. *illac ad*) + ᴀᴄᴄ 'in, at, to'	• *në* (< PAlb *en-da* < IE *en* 'in' + IE *do* 'to') + ᴀᴄᴄ 'in, on, to'
• *cu* (< Lat. *cum*) + ᴀᴄᴄ 'with'	• *me* (< PAlb. *me(t)*) + ᴀᴄᴄ 'with'
• *lângă* (< Lat. *longu ad*) + ᴀᴄᴄ 'beside'	• *buzë* (< PAlb. *budjā* 'lip') + ᴅᴀᴛ 'beside'
• *de* (< Lat. *de*) + ᴀᴄᴄ 'of, from,' *din* (< Lat. *de in*) + ᴀᴄᴄ 'from,' *de la* (< Lat. *de illac ad*) + ᴀᴄᴄ 'from'	• *nga* (< PAlb. *en-ka* 'in' + 'which is') + ɴᴏᴍ 'from,' *prej* (< IE *prō* 'forward, before') + ᴀʙʟ 'from'

What Albanian and Balkan Romance share is not a common set of prepositions, but rather a structural pattern in which accusative-selecting prepositions with primarily locative semantics trigger determiner omission depending on the modification status of the noun phrase. The set of accusative-selecting prepositions in Romanian that trigger determiner omission are not all primarily locative in meaning. But they are mostly locative, and if they no longer have a locative function now, they used to have one in Late Latin. Moreover, Aromanian shows a more extensive relationship between locative semantics and LDO. This implies not only a structural, but a semantic parallel between determiner omission in Balkan Romance and in Albanian. For example, both Balkan Romance and Albanian show exceptional behavior in the preposition 'with,' in that it is accusative-selecting, but fails to trigger determiner omission (and also lacks locative semantics). This common treatment of 'with' occurs despite the differing etymological origins of the corresponding lexical items in either language.

The lack of LDO with the accusative-selecting preposition translated as 'with' in both Albanian and Balkan Romance deserves particular attention. As in Albanian, the Balkan Romance preposition meaning 'with,' *cu*, behaves differently than the other prepositions selecting accusative case. As would be expected based on the use of the article in other syntactic positions, *cu* selects nouns with a definite article to express definite reference. In Romanian, there is a contrast with Albanian, however, in that there is no distinction between the selectional properties of 'with' expressing accompaniment and 'with' expressing means. Whereas the Balkan Romance preposition *cu* requires all nouns it governs to take either a definite or indefinite article, in Albanian, it is only the use of *me* 'with' to express accompaniment that requires the presence of an article. In Aromanian, the relationship is more parallel; *cu* as an expression of means triggers determiner omission, as in Albanian. Thus, while Balkan Romance and Albanian share the broader behavior of separating out 'with' from other accusative-selecting prepositions in regard to LDO, Aromanian is closer to Albanian in the specific behavior.

Romanian, in particular, also differs from Albanian in that accusative-selecting prepositions with non-locative semantics still trigger LDO. This is another way in which Aromanian shows more similarities with Albanian. In Romanian, although *pe* 'on' can express locative semantics, it nevertheless triggers LDO in its role as a preposition designating an animate, accusative direct object (cf. §3.1). Likewise, uses of *la* 'to' designating a beneficiary rather than a destination or a location still trigger LDO (cf. §3.2). This was not, however, the behavior of early Balkan Romance. Earlier states of the language matched what is now found in Albanian, in that they distinguished locative from non-locative uses of the same accusative-selecting prepositions, with only locative uses triggering LDO. Likewise, Aromanian maintains the distinction between locative and non-locative uses of *la* into the current period. That is, at an earlier time in its history, Balkan Romance exhibited the same set of rules for LDO that now hold in most dialects of Albanian and in standard Albanian, as well as Aromanian.

To review, Balkan Romance and Early Albanian as documented before the turn of the 19th century shared the following properties in relation to definiteness marking within the domain of prepositions:

- Both languages normally express definiteness with a definite article, which declines for case

- Both languages had a distinct set of prepositions that select the accusative case

- Pronouns receive uniform marking for accusative case with these prepositions

- Accusative-selecting prepositions with locative meaning trigger distinct marking of definiteness on nouns, a pattern which does not appear outside the domain of prepositions

- From among the set of prepositions that trigger distinct definiteness marking, prepositions used in a non-locative sense fail to trigger distinctive marking and definiteness is marked with the standard article appropriate to the accusative case

- In particular, 'with' shows exceptional behavior, different from other accusative-selecting prepositions in regard to definiteness marking

These similarities can be situated among a number of shared features between Albanian and Balkan Romance that developed in these languages before historical attestation (MacRobert 1982: 147). The predecessor to Balkan Romance, Latin, did not possess definite determiners. Late Latin texts that show the early stages of definite determiner development have no distinctive pattern of differentiation between prepositional phrases and other contexts. By necessity, then, the distinctive omission of definite determiners in locative prepositional phrases must have been an innovation in Balkan Romance, a feature that followed the introduction of a consistent pattern of indexing definite reference with a post-posed definite article. It is likely that the definite article emerged in both Albanian and Balkan Romance during a period of sustained contact (Topalli 2009a:376). Both Albanian and Balkan Romance speakers emerge in the historical record at about the same time, before direct attestation of either language (Pipa 1989:200). Both languages make use of a particle of concord that shows agreement between the head of the noun phrase and certain classes of modifiers such as possessives and adjectives; at the time of the earliest attestations of Romanian, the particle of concord was used in a distribution that is more similar to that of the p particle of concordarticle of concord in Albanian (Mann 1977:103-104, Orel 2000:247). Albanian and Balkan Romance also share a common set of words that are unattested in other languages of the Balkans, which implies earlier bilateral contact before the period of Balkan convergence that affected Balkan Slavic in the Middle Ages (Demiraj 2004:98). The extensive Romanization of Albanian vocabulary requires a very early period of intense cultural exchange and bilingualism in Latin; Balkan Romance may have been the outcome of language shift of ancient Albanian speakers to Late Latin (Hamp 1989:47, Hamp 1994:67, Joseph 1999:225), and there was a period during which Latin speakers along the coasts of the Adriatic shifted to Albanian as Albanian tribes migrated down from the mountains and toward the coasts (Demiraj 2004:112). A special configuration of definiteness marking for accusative-selecting prepositions fits within a broad pattern of convergence between these two languages beginning from the ancient period of their contact. In particular, it seems to have been the result of language shift of an Albanian speaking population into Late Latin, resulting in grammatical changes to Late Latin that brought it closer in structure to Albanian, with Balkan Romance as an outcome and special treatment of definite reference within locative, accusative-selecting prepositional phrases as an integrated feature shared by both languages.

Here, I have been careful to distinguish what was shared between Albanian and Balkan Romance at the earliest period of their contact from LDO, referring rather to a special treatment of definiteness marking in accusative-selecting prepositional phrases. In Balkan Romance, this pattern of exceptional definiteness marking was, in fact, LDO. But in Albanian, it was a locative case, not inherited from proto-Indo-European but rather innovated to express definiteness on nouns in the specific context of locative prepositional phrases. The Albanian locative case occurred in no

other domain than accusative-selecting prepositional phrases; likewise, Balkan Romance determiner omission appeared only with this class of prepositions. The Albanian locative case was distributed throughout its dialects, bridging the Tosk/Geg divide, a fact which suggests that it must have been an innovation in the case system that preceded the divergence of these dialects during the 8th to 10th centuries (Rusakov 2013). Likewise, LDO is attested throughout Balkan Romance, in all languages descended from Balkan Late Latin including Istro-Romanian, Aromanian, and Megleno-Romanian. It must also have been an early innovation in Balkan Romance, but one that followed the introduction of definiteness into the Balkan Romance grammatical system. The key timeframe for convergence between Albanian and Balkan Romance would have been their period of sustained, intensive contact between proto-Albanian speakers and Late Latin speakers from the beginning of the first millennium until the 7th and 8th centuries (Rusakov 2013). Without documentation, it cannot be established conclusively that a shared exceptional treatment of definiteness in locative prepositional phrases was the result of this contact. Nor can it be established whether one language was the source of this pattern of differential definiteness marking. Nonetheless, the convergence between these languages in their treatment of the junction between definiteness and locative prepositional phrases can be interpreted as part of a broad pattern of convergence in grammatical structure between Albanian and Balkan Romance that was the result of early language contact, preceding the Slavic migrations into the Balkans.

5.3 The diachrony of determiners in Balkan Slavic and the influence of Aromanian on Macedonian

The divergence between Macedonian and Bulgarian in the treatment of locative prepositional phrases is important when considered in the context of their different histories of language contact. While Macedonian underwent extensive contact with Aromanian due to the mingling of their speaker communities, Bulgarian did not engage in the same degree of contact. This is reflected in the presence of a number of distinctively 'un-Slavic' features, such as the have-perfect and prepositional doubling in Macedonian, or analytical marking of the accusative case with the preposition *na* 'on' (equivalent to the Aromanian accusative-marking preposition *pi* 'on') (Markoviḱ 2007:91). These extensive similarities with Aromanian are absent in Bulgarian, whose grammatical structures otherwise closely correlate to those of Macedonian (Gołąb 1959, Solta 1980, Gołąb 1984, Friedman 2008) as East South Slavic languages. Macedonian distinguishes itself from Bulgarian by the greater extent to which it allows or requires determiner omission in locative prepositional phrases (cf. §4.3). This, like other grammatical factors that distinguish Macedonian from Bulgarian, must be viewed to have its origin in contact with Aromanian. Determiner omission in Macedonian, while not grammatically integrated to the same extent as in Balkan Romance or Albanian, does appear as a feature shared with Aromanian among several other grammatical features these languages have in common.

The historical extent of this pattern of determiner omission in Macedonian is difficult to determine, but it is certainly not an early pattern in the language in the way that LDO can be found in early Balkan Romance. The first attestations of the Slavic expansion into the Balkans date from the 6th century. Slavic literacy, however, did not emerge until approximately three centuries later, in the late 9th century. The oldest directly preserved Slavic texts may have been written in the tenth

century, although none are dated (Lunt 2001:4). The record of Macedonian attestation is marked by a 'dark period' extending from the 11th to the 16th centuries in which no documentation is available (Wahlström 2015:12). Nonetheless, it can be derived that a Paleo-Balkan substrate cannot account for the distribution of definiteness in Macedonian prepositional phrases. The first written Slavic documents date from the 9th century, and surface as copies in the 10th century. These documents show an extensive case system very close to what has been reconstructed for Late Common Slavic; on the basis of this documentation, the language of the South Slavs from this period lacked definite articles (Wahlström 2015:13). A distinction of definiteness specific to prepositional phrases cannot exist without a grammaticalized definite article. Since the Slavic migrations into the Balkans began in the late 6th century (Fine 1983:29, 31), any substrate effect on definiteness or case should have already manifested in the Old Church Slavonic writings found three centuries later. While there were uses of the demonstrative that corresponded to definite reference, these were not consistent and did not amount to a grammaticalized definite article, nor did they show the distinctive post-positive syntactic distribution, in texts from the 9th to 11th centuries (Kurz 1938, Mladenova 2007:98).

By the time of the 17th century, when Balkan Slavic reappears in writing corresponding closely to the spoken language of the time (Dëmina 1985:322, Wahlström 2015:32), the post-posed definite article appears. The article first appeared with plurals, then inanimate nouns, and finally feminine followed by masculine animates. For syntactic position, the article was present earlier in object and oblique position than in subject positions (Mladenova 2007:134-135). This article could still express case distinctions in the nominative-genitive for animate, masculine nouns in non-subject position (Demina 1985:315). This was a byproduct of the grammaticalization of the marking of definiteness from a demonstrative that also expressed nominative-genitive case for animate, masculine nouns (Wahlström 2015:109). It could also be used with unmodified nouns in locative prepositional phrases (Mladenova 2007:110), which is contrary to the pattern of LDO.

BULGARIAN (17TH CENTURY)

(108) i   pomysli   da ide na b[o]žii   grobĭ   i   da sedi tamo izokolu **po**
       and he.thought to go to divine-M.SG grave.M.SG and to sit   there around in
       pustynja-ta
       desert-F.SG.DEF
       'And he thought to go to the Holy Sepulchre and to stay around there in the desert.' (Demina 1985:156)

A locative case, as opposed to LDO, did exist in the OCS documents of the 9th century. Unlike the Albanian locative case, the Slavic locative case played no role in the development of LDO in Macedonian. The Slavic locative was used exclusively with locative prepositions and indicated a location rather than a goal. To indicate a goal, the combination of a locative preposition with the accusative was used. By the time of the 17th century Damascene documents, however, the indication of goal and location had collapsed as in other Balkan languages, and an accusative or oblique case was used with locative prepositions instead (Wahlström 2015:62), and this merger may already have been underway in the 13th century (Steinke 1968:105-106). The preposition *sŭ* 'with' was exceptional in this regard, taking the instrumental case instead and already showing a collapse between instrumental and accompaniment uses in the OCS documents (Duridanov 1956:113). This preposition with oblique case already begins to occur in Slavic documents in the 12th century (Sobolev

1991:31-32). The definite article arose unevenly in Balkan Slavic (covering both Macedonian and Bulgarian), with some dialects post-posing the article to both nouns and adjectives, while others introduced the definite article with nouns but retained a definite inflection for adjectives inherited from Late Common Slavic (Mladenova 2007:345). The Damascene documents show a mixture of the new, post-posed definite article with the old definite inflection in adjectives, indicating that the grammaticalized definite article was already penetrating through most Slavic dialects at this time. Both a source in Greek (with innovation radiating outward from the Rhodope dialects) and a source in Moesia (with innovation proceeding from bilingualism with Balkan Romance) have been suggested (Mladenova 2007:350-352). The Balkan Romance feature would better fit the structural typology of definiteness in Balkan Slavic. In any case, locative determiner omission was neither a feature of the earliest documents of modern Balkan Slavic, nor is it a feature of modern Bulgarian dialects. The distinctive treatment of definiteness in locative prepositional phrases is a feature that appears specifically in Macedonian speech forms as later documented in the 19th and 20th centuries.

In schematization, Macedonian underwent the transformation detailed in Figure 5.9. In Early Slavic, prepositions selected for a number of nominal cases, including dative, accusative, genitive, and instrumental. Case, as well as gender (gen) and number (num) were expressed on both the noun (N) and any declinable modifiers to the noun, such as adjectives (A). Definiteness did not play a role in nominal morphology or semantics. Macedonian, in contrast, morphosyntactically codes definite reference with a definite article (Def) post-posed to the first declinable constituent in the noun phrase. Since nouns no longer express morphological case, prepositions do not select for nominal case. When present, LDO occurs only with prepositions that express locative semantics, in which case the bare, unmodified noun expresses definite reference.
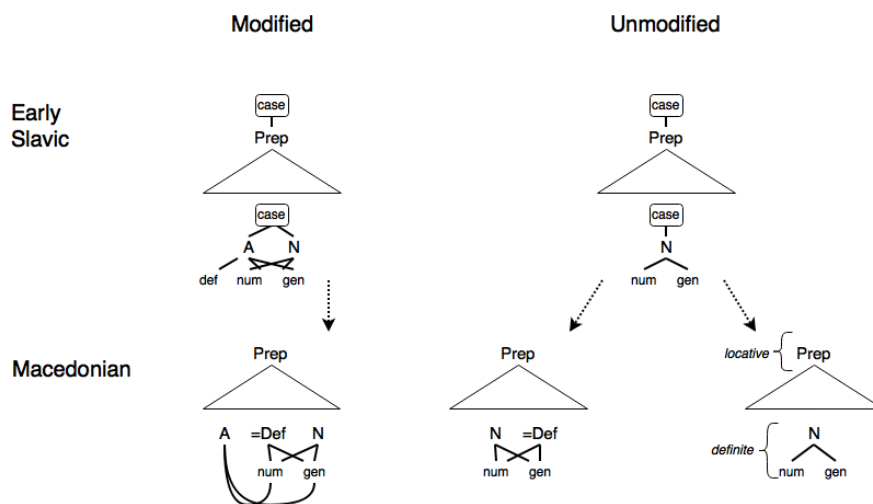

Figure 5.9: The outcome of the shift toward LDO in Macedonian

The most striking similarity in LDO is found between Albanian and Balkan Romance. These languages share the entire set of defining LDO characteristics, including locative semantic triggers, a defined set of triggering prepositions, and alternations between LDO and the accusative-case definite article for modified nouns. Albanian and Balkan Romance also share other features such as a head-initial noun phrases, particles of concord for adjectival and possessive modifiers, and mor-

phological distinctions among nominative, accusative, and dative/genitive case; these are shared between Albanian and Balkan Romance in addition to a set of lexical items that do not have parallels in other Balkan languages or outside of the Balkans. This naturally has led some authors to conclude that LDO is an inherited feature(Michov 1908, Schaller 1975, Joseph 1999), introduced by a predecessor language that served as a substrate for both Balkan Romance and Albanian. LDO, then, could be listed among the inherited factors that bind together the Balkan Romance and Albanian speaker populations prior to the historical record. Macedonian, whose noun phrases have a different structure and which was introduced to the Balkans at a later time, could easily have developed its own unstable expression of definiteness in prepositional phrases due to its relatively recent development of definite articles and its loss of the case system inherited from Late Common Slavic, as documented in Old Church Slavonic. In order to rule out the possibility of LDO as an inherited structure for Balkan Romance and Albanian, and to rule it out as a remnant of Common Slavic in Macedonian, I have examined the evidence of structure of these languages before the historical records, to the extent it can be reconstructed.

In the record, Latin is available the earliest of any of the predecessors to the Balkan languages, followed by Old Church Slavonic. Thereafter, Balkan Romance first appears with extensive documentation in the historical record, followed by Balkan Slavic and finally Albanian. We can show from reconstruction and the historical stages of these languages as they become available in the records that LDO was not directly inherited as a feature in all the languages that now possess it. Rather, it first occurred in Balkan Romance, and must have developed prior to the separation of Balkan Romance into multiple branches in the 11th century. This would have been at the end of a period of intense, intimate contact between early Albanian and Balkan Romance, a period that is not documented but which we can trace in the numerous structural and lexical correspondences between these two Balkan language branches. While Balkan Romance possessed LDO from its earliest documentation, Albanian instead showed a locative case that alternated with the accusative case. The locative case appeared in prepositional constructions that were very similar to those that triggered LDO in Balkan Romance, but in Albanian this environment produced a morphological alternation with the definite accusative article, and a syntactic modification condition did not apply. The gradual loss of the locative case did not necessarily lead to the development of LDO, as peripheral dialects unconnected with the main body of Albanian attest. Rather, it was a period of further contact between Albanian and Aromanian in the 19th and 20th century that led to the spread of LDO, at least from southern Tosk dialects into more northern dialects and into the standard language. Northern Geg dialects, outside the range of documentation prior to the 20th century, appear to have already lost the locative case. The source of this loss, however, must have been the process of elision of the locative case ending -t in the presence of modifiers accompanied by the locative particle of concord të. This tendency has been noted throughout Albanian dialects in early attestations, and it is an independent tendency connected to other Albanian processes like the loss of the definite article in -t from the dative, genitive, and ablative plural. LDO when found in Geg, then, came about not through the influence of Balkan Romance, but due to the spread of this feature from southern Tosk. We can see the pattern of development for LDO in Tosk Albanian, where assimilation of Aromanian speakers corresponds with the transition from a remnant locative case to grammatically integrated LDO. We can also see it in Macedonian, where contact with Aromanian was the key factor that led to the partial development of LDO in southwestern dialects of Macedonian, as opposed to the absence of LDO in Bulgarian, where contact with Aromanian was not a significant factor.

   This diachronic account of the development of LDO in Albanian, Balkan Romance, and Balkan Slavic not only points to language contact as a major factor, but points to repeated periods of contact between languages in different states of grammatical convergence and, moreover, contact among dialects of individual languages. The languages of these groups are part of the well-known Balkan Sprachbund, one of the most extensively studied linguistic areas of the world. But the development of this feature was not a matter of simple borrowing. Nor is it enough to say that LDO is another feature that may be designated a 'Balkanism.' Multiple instances of migration, inter-communal contact, and assimilation were necessary for the final distribution of LDO observed in the modern languages. In this respect, Balkanisms cannot be simply lumped together as the common outcome of simple convergence. From a structural perspective, a number of grammatical patterns, such as post-positive markers for definiteness and the goal/location merger, needed to become common prior to the development of LDO. This turns our attention to the processes that define the Balkan linguistic area, whose rich literature and whose contributions to the study of contact linguistics inform the fully developed account of how LDO came to be an areal feature of some, but not all Balkan languages.

Chapter 6

Language Contact and the Balkan Linguistic Area

In this chapter, I will situate the phenomenon of LDO shared among Albanian, Balkan Romance, and Balkan Slavic languages in the literature on contact in the Balkan linguistic area. In §6.1, I begin by addressing early work on Balkan languages and discuss the insight of how features could come to be shared among languages through diffusion rather than through common inheritance from a genetic ancestor. Balkan languages are distinctive both for the features they share in common and for the ways in which they diverge from their genetic relatives outside the Balkans. In §6.2, I discuss how this recognition helped define the idea of the Sprachbund, or linguistic area. In §6.3, I review the literature on the linguistic area, a theoretical construct that faces challenges of definition and applicability. In §6.3, I discuss how the occurrence of numerous commonalities may give the impression of a clear linguistic area, but in §6.3, I review the arguments for why this definition of a linguistic area from circumstance fails to exclude accidental resemblance due to internally motivated innovation or shared inheritance. In §6.3, I discuss the corresponding importance of developing a historical perspective in the definition of a linguistic area. In §6.4, I turn to the social-ideological factors that allow for a linguistic area to develop and discuss the mechanisms and motivations for language change due to contact among speakers. This allows me to address the specific social and structural factories that would have facilitated the spread of LDO among languages of the Albanian, Balkan Romance, and Balkan Slavic branches and account for the distribution of this feature geographically, diachronically, and dialectologically among the 'Balkanisms' that together define the Balkan linguistic area.

6.1   Early work on Balkan languages: Kopitar, Schleicher, Sandfeld

Balkan languages drew the attention of linguists in the 19th century, during a period when work on the reconstruction of the Indo-European language family was proceeding rapidly. In particular, Balkan languages stood out for their failure to fit into the developing understanding of the interrelations among the languages in Europe. At a time when empirical methods of investigation were identifying the ways in which commonalities among languages could be connected to shared ances-

tors reconstructed through regular sound change, the similarities among Balkan languages posed a challenge to this genetic, "tree" model of linguistic differentiation.

Kopitar (1829) first brought the commonalities of languages in the Balkans to the attention of scholarly literature. Kopitar began with the premise that the differences between 'Walachian' (Balkan Romance) and the other Romance languages of Western Europe were particularly notable. It was simple to show that Balkan Romance was descended from the same Latin predecessor that yielded the other languages of the Romance branch of Indo-European. But grammatically, Balkan Romance stood apart in a number of key features of grammar. Moreover, these grammatical features seemed to show that Balkan Romance bore a closer relation with Albanian and 'Bulgarian' (Balkan Slavic)[1] than with its genetic relatives outside the Balkans as determined by sound and lexical correspondence. Kopitar accounted for this puzzle with the argument that, whereas the Romance languages outside of the Balkans were subject to uniform Germanic influence, Romanian and other branches of Balkan Romance were the result of Late Latin lexical material imposed on 'Thracian' grammar. Thracian featured among the ancient languages of the Balkans that had been named, but not directly attested in Greek and Roman works. Kopitar connected Albanian, another language spoken in the Balkans, with Thracian, and proposed that Albanian maintained both lexical and grammatical material from Thracian as genetic ancestor (notably, this would have placed the proto-Albanian speaker population somewhere in what is now modern Bulgaria). In effect, Kopitar 'solved' for the undocumented language of Thracian by matching it to the only Balkan language for which he could not trace a pre-historic ancestor. He also noticed that Balkan Slavic had replaced its inherited nominal case inflections with analytical constructions. He compared this with the nominal case systems of the neighboring Albanian and Balkan Romance speech languages and concluded that these latter two must have had an influence on Balkan Slavic. This triangle of mutual influence was summed up in his famous formulation of a zone of "three lexically different but grammatically identical languages that form a barrier between Greek and Slavic, from the lowest Danubian valleys across the whole of the Haemus Mons [Balkan Mountains], from sea to sea" (1829:106).[2] The outcome of this comparison was the idea that Balkan Romance, Albanian, and Balkan Slavic needed to be examined as a unit, independent of the study of the branches of Indo-European to which they belonged. The result of Kopitar's observation was the idea that it could more informative to look at relations among languages in terms of geography than in terms of reconstructed genetic ancestors. Thus, the identification of commonalities among Albanian, Balkan Romance, and Balkan Slavic introduced the idea of areal studies to linguistics.

Kopitar's analysis depended on a comparison of the languages of the Balkans with each other in addition and complementary to a comparison between each language and other members of its branch. It was important that a definition of the Balkan zone of 'grammatical identicality' made reference not only to geography—the situation of speakers in a specific geographical range—but also to grammatical traits analyzed with reference to languages spoken outside that geographical range, in cross-comparison. In practice, this meant that Kopitar compared Balkan Romance against other Romance languages, and Balkan Slavic against other Slavic languages. Albanian at the time

---

[1]It was common in the literature until the mid-20th century (and sometimes even thereafter as a result of a political considerations, see (Friedman 2002/3:164-167)) to group the East South Slavic dialects that now constitute Macedonian and Bulgarian under a single label of 'Bulgarian.'

[2]In the original German, "drey lexikalisch verschiedenen, aber grammatisch identischen Sprachen, die vom untersten Donauthale an längs des ganzen Hämusgebirges von Meer zu Meer zwischen den Greichen und Slawen die Grenzscheide."

appeared to have no close relative outside the range of the Balkans, and so instead Kopitar compared it against the historical record (limited though it was). Simultaneously, he compared these languages with each other in order to derive a distinct group whose similarities crossed the family lines separating Slavic, Romance, and Albanian within the Indo-European family. This coordination of comparison across geography, diachronic ancestors, and synchronic relatives revealed that the features shared in common among the Balkan set of Slavic, Romance, and Albanian languages were not traceable to a common ancestor, but that they could be traced to a common geographical range.

Though geography was an important component of his definition of the Balkan language area, geography alone was not sufficient to place a language in the group Kopitar had discovered. To this point, Kopitar provided examples of the same text in Bulgarian, a Slavic language of the Balkan type, and Serbian, a closely related Slavic language. Even though Bulgarian and Serbian could be easily shown to belong to the same South Slavic grouping within the Slavic branch of Indo-European through lexical and sound correspondences, their grammatical characteristics were notably divergent. Thus the geographical basis for comparison was delimited by the comparison with other languages in the same area. After all, Serbian is also spoken in the Balkans, but it is widely agreed to stand apart from the Balkan linguistic area (with the exception of southern Serbian Torlak dialects, which appear to share the older phonological and morphologic developments common to West South Slavic, but the later morphosyntactic developments common to Balkan Slavic, see Alexander 1983, 1984). Kopitar thus introduced the conceptual distinction between languages of the Balkans—those languages currently or once spoken on the Balkan peninsula—and *Balkan languages*—languages of the Balkans whose geographical ranges are in contact or overlap and which share structural similarities that are not a result of genetic heritage. Both geography and a particular kind of interrelation with other contiguous languages defined this group and tied them together. Thus, both geography and cross-comparison have been essential to defining such language groups in all subsequent work (Hamp 1977; Jakobson 1958/1971).

The nature of the features that tie languages of the Balkans together has been a subject of analysis in its own right. The grammatical commonalities among the Balkan languages not only mark them for their similarities, but also distinguish them from genealogically related, non-Balkan languages, particularly in morphosyntax. This was argued, with an ideological overtone, in the work of Austrian censor August Schleicher (1850:143), who, like Kopitar, compared Romanian to Romance languages and Bulgarian to Slavic languages. He also connected Albanian to Greek, another language spoken in the Balkans, and described the same set of three languages analyzed by Kopitar as the 'most corrupt' of their families. Grammatical deviation from the languages outside of the Balkan linguistic area was framed as a kind of degradation, in that the grammatical similarities of the Balkan languages stood in contrast to the 'purer' structure of languages outside the Balkans. From this perspective, it is not shared innovations that brought the Balkan languages together, but rather a collective failure to maintain inherited linguistic structures. Although without the moral overtone, this ideological frame has persisted into later work on the Balkan languages and that has likewise been contested (Topolińska 2007). Even when avoiding a discourse of corruption and degradation, a persistent question relevant to both Balkan languages and the study of linguistic areas in general has been whether there is a connection between sharing structure across genealogical lines and simplification or reduction in features.

The case of LDO in Albanian demonstrates how it could be difficult to determine whether the process that yielded the Balkan linguistic area involved simplification of grammatical distinctions.

On the one hand, early Albanian possessed a distinct locative case, which was lost in the later language as demonstrated in §5.2-§5.2. The reduction of a morphological case system from five cases to four seems like a straightforward matter of loss of complexity. But as demonstrated in §5.2, the introduction of LDO involved the addition of complexity to Albanian, in that it required a new, grammatically encoded alternation between a zero-marking for definiteness in one syntactic context and an accusative-case definite article in another context, without parallel in the rest of the Albanian grammatical system. Moreover, while case loss preceded the introduction of LDO into Albanian, the development of grammatically conditioned determiner omission actually maintained a distinction of definiteness in accusative-selecting prepositional phrases that was lost in other grammatical contexts, such as the dative/genitive plural. Loss of one structure and innovation of another resulted both in the conservation of a grammatical distinction internal to Albanian and a convergence toward similar surface structure in Balkan Romance in structure. This demonstrates how a failure to carefully compare languages in the Balkans to one another may bias an observer to perceive the Balkan linguistic area as defined solely by 'what's missing' when these languages are compared to their ancestors or to related languages outside the Balkans. This approach misses the ways in which they may have developed their shared features, not through loss, but through innovation of new complexity and irregularity.

The classic study of the Balkan languages for their own sake, rather than as remarkable variants of already established branches of Indo-European, began with the work of Sandfeld (1930), first published in Danish in 1926. In his book, Sandfeld was particularly concerned with the social relations that could have resulted in Balkan language similarities. Thus his study moved beyond merely observing an association between geographical range and grammatical similarity that could not otherwise be accounted for by genetic inheritance, and took on the task of explaining the function of geographic contiguity in producing convergent language development. Sandfeld went about this by establishing shared vocabulary among the Balkan languages, tracing loanwords from Turkish, Greek, Latin, Slavic, and Albanian (1930:16-99). These loanwords, significantly, were not the outcomes of reconstructed roots shared among ancestral languages, as in historical reconstruction. Rather, they appeared to have spread through the Balkan languages due to their geographical contiguity and an extended period of shared culture among their speaker populations. Religious words, words for artifacts, and vocabulary referring to institutions formed much of the inventory of borrowings. Many of these words had an obvious source in Greek or Turkish, culturally dominant languages of this region for a period of many centuries. Here, then a set of similarities could be explained by a shared cultural domain defined by dominant language(s), and in addition, Greek and Turkish were introduced for comparison into a body that had previously been restricted to Albanian, Balkan Slavic, and Balkan Romance.

In addition borrowings that seemed to have resulted from the cultural dominance of Greek and Turkish during the Byzantine and Ottoman eras of influence over the Balkans, Sandfeld noted words of uncertain origin that seemed to indicate a time-depth of mutual influence among some Balkan languages that predated either Greek or Turkish. These lexical commonalities, rather than deriving clearly from a synchronic Balkan language, seemed to be a shared inheritance from some unidentified predecessor. The list of these words he provided showed a pattern of pertaining to shepherding and pastoral semantic domains. This semantic cue offered another means of explaining similarities, by making connections between cultural and occupational properties of the speaker populations and language distribution. These pastoral words, in particular, are often shared between Balkan Romance and Albanian, languages that were both associated with pastoral communities that retreated

into mountainous terrain following the fall of the Roman Empire. Slavic, in contrast, came to the Balkans during the invasions of the 6th and 7th centuries, linguistically inundating the peninsula all the way down to southernmost Greece and spoken extensively along the river valleys, where invasion was easy. (Friedman 2007:206-207). Thus in language of agricultural practice, Slavic roots dominated. After this point, Greek was spoken primarily as a coastal, urban language, with prestige in urban communication, literacy, and religion (for Christians). Turkish, introduced in the 14th century by the Ottoman Empire, also served as an urban prestige language (for Muslims). Although not analyzed in Sandfeld's work, two additional languages can be included among Balkan languages with particular social roles. Balkan Romani—a series of sometimes widely diverging dialects belonging to the Indo-Aryan branch of Indo-European—arrived during the tenth to twelfth centuries (Friedman 2007:207) and shows some structural Balkan features (Matras 1994). Judezmo—a language descended from Castilian Spanish spoken by the Balkan Jewish community that migrated to the Ottoman Empire in 1492 following the Reconquista in Spain (Friedman 2007:207). Both were spoken as home languages, with the Roma and Jews generally speaking Turkish as well as other local languages, but without members of other ethnic/religious groups speaking Romani or Judezmo. In this way, patterns of lexical borrowing and terms from specific domains could be connected to cultural properties of the speaker populations, providing a possible explanation for why terms from particular languages seemed to dominate across all languages in the area.

Borrowing vocabulary is a well-known process for language contact. What stands out in the Balkans is rather the apparent borrowing of grammatical structure, which cannot be traced so easily to cultural dominance or the need to describe new cultural domains. Beyond vocabulary, Sandfeld provided a key insight for accounting for similarities among the Balkan languages by analyzing correspondences "outside the lexicon," which he categorized by proposed source and listed in terms of pair-wise relationships between languages: Greek, Albanian, Aromanian, Balkan Slavic, or Turkish (1930:100-162). Outside the lexicon, here, refers to features of grammatical structure. Sandfeld (1930:161) drew a distinction between sources of vocabulary borrowing and sources of structural influence. Thus, he saw Turkish as a source of lexical and phraseological borrowings, but he argued against it as a source of grammatical convergences, citing its more recent introduction among the community of speakers in the Balkans and its highly divergent grammatical structure compared to the other Indo-European languages that make up the proposed Balkan linguistic area. In making this distinction, Sandfeld was taking into account both time-depth of contact and structural receptivity to borrowing and interference; that is, he made a point of including these among the factors affecting the convergence of Balkan languages. Both these factors of time-depth and structural receptivity have since been elaborated as important factors determining grammatical convergence between languages in other areas. Having established individual pair-wise correspondences, Sandfeld (1930:163-216) turned to convergences that apply across many of the Balkan languages as a whole, excluding Turkish. Although lists of convergent Balkan features had been produced before, notably by Miklosich (1862:6-8), Sandfeld's list became the classic starting point for analyzing Balkan linguistic convergence. His list included the following features:

1. the post-posed definite article;

2. the loss of the infinitive and its replacement with an analytical construction involving a conjugated verb and a subordinating particle;

3. future tense formed with a particle derived from the verb 'want;'

4. identical forms for 'where' and 'whither,' correspondingly 'here' and 'hither;'

5. nominal complement rendered as an accusative object;

6. beginning an affirmative clause with 'and' following a negative clause;

7. paratactic in preference to hypotactic constructions;

8. separate clitic and full pronoun forms, with doubling of object pronouns and the use of object clitic pronoun agreeing with an object noun in the same clause (object reduplication);

9. age rendered with 'have X years' rather than 'be X years';

10. 'how' used to mean 'approximately;' and

11. miscellaneous phraseology.

Two of these famous Balkan structural features, or 'Balkanisms,' can be identified as precursors to LDO. Without a post-positive article (1), the distinctive omission of the article that defines LDO is not observable, and those Balkan languages which show pre-positive articles that would be in direct phonological contact with the preposition (Greek and Romani) do not show LDO. In addition, a collapse between 'where' and 'whither' in the grammatical system (4) is a necessary requirement for the semantic trigger of LDO. Only those prepositions that indicate a collapsed goal/location distinction require LDO in the noun phrases they govern. While LDO was not proposed by Sandfeld as a Balkan structural feature and has not normally been included among subsequent enumerations of Balkanisms, his contribution already identified grammatical elements that would be relevant to the identification of LDO as a shared Balkan feature.

Sandfeld (1930: 212) attributed these linguistic commonalities to the influence of Greek, which he saw to be the prestige language of the Balkans during the Byzantine period and extending into the Ottoman period. This would render the explanation for Balkan grammatical features the same as the explanation for patterns of lexical borrowing: a single language, culturally dominant, imposes a structure that other languages borrow and adopt into their own grammatical constructions. This connects lexical borrowing with grammatical convergence. It also shows an approach to language convergence in which commonalities are listed and traced to a single source, similar to the way in which lexical items can be listed and traced to a source. This listing approach has been a fruitful means of providing concrete definitions for membership in a group of languages related to one another by contact, but it is also a potential source of conflict, as later literature on the Balkan languages illustrated. Greek, for instance, cannot be taken as the source of the Balkan post-positive article, for the simple reason that Greek definite articles are pre-posed to the noun phrase and appear in documentation during a period that predates contact with Latin and Common Slavic (Leumann, Hoffman, and Szantyr 1965:191-192). Taking a cultural dominant language such as Greek to be the logical source of pressure toward Balkan grammatical structure runs aground against the problem that Greek lacks several of the features shared by the more core group of Albanian, Balkan Romance, and Balkan Slavic identified by Kopitar.

In the early works of Kopitar, Schleicher, Miklosich, and Sandfeld, we see patterns of analysis emerging that continued to be relevant in later studies of Balkan languages and areas of languages related by contact in general. A geographically bounded set of languages is observed to stand out

through inter- and intracomparison with synchronic relatives and diachronic ancestors. Defining the Balkan linguistic area required comparing geographically contiguous languages not only with their relatives outside the area, but also with each other, in order to establish a set of similarities. Analyzing a contiguous set of languages allows for the possibility of cultural and social influences to have affected the speaker populations within the area. Similarities between Balkan Slavic, Balkan Romance, Albanian, Greek, and Turkish[3] were situated in a setting of shared culture and exchange, drawing connections between culture and language change. Structures that distinguished the languages from their relatives were just as important as structures in which they resembled one another. In some ways, these cases of convergence toward one another and divergence away from one another were viewed only as cases of grammatical simplification. Lexical borrowing seemed to be connected to the possession of structural similarities, but not all languages engaging in lexical borrowing seemed to be affected by convergence in the same way. Drawing up a list of commonalities helps to clarify membership and provides criteria for testing whether other languages beyond the initial set under analysis should be included in such a group, which has come to be called a *linguistic area*.

Since Sandfeld, a number of overviews of the Balkan languages have been produced, cf. Schaller (1975), Haarmann (1978), Solta (1980), Cyxun (1981), Banfi (1985), Feuillet (1986), Desnickaja (1990), Demiraj (1994), Reiter (1994), Steinke and Vraciu (1999), Hinrichs (1999), Asenova (2002), and Friedman (2008a; 2011), inter alia. Sandfeld's study became the starting point for an extensive program of study into how the Balkan languages came to resemble each other so greatly in grammatical structure. His discoveries also initiated research into the general phenomenon of areally distributed grammatical commonalities that cross family lines.

6.2 Balkan languages and defining linguistic areas

The Balkan group of languages has provided a baseline not only for the study of processes specific to this geographical region and its social structures, but also for the study of areal linguistics in general. Once the phenomenon of languages influencing each other in structure in ways that could not be traced to inheritance from a single ancestor was noted, it could be applied more generally to the study of human language similarity and variability. Observing the commonalities among the languages of the Balkans, Trubetzkoy (1923) proposed the idea of a 'linguistic union,' also known as a Sprachbund. This term, used in Trubetzkoy (1928), corresponds to the general term 'linguistic area' employed across areal linguistics (Emeneau 1965; Velten 1943). Trubetzkoy defined the Balkan linguistic area as a cultural-historical group whose outcome was linguistic convergence. Thus a linguistic area, in his terms, was constituted not only by geography, but also by a shared culture and a specific historical process. Trubetzkoy (1931:233-234) saw a linguistic area as comparable to dialect geography: isoglosses, denoting a particular feature, could extend beyond not only a single dialect in a dialect continuum, but beyond a single language. A Sprachbund, then could be interpreted as a kind of 'language continuum.' He argued that this kind of isogloss diffusion could occur not only among distantly related languages, as in the Balkans, but also across language families, which would provide an explanation for how Balkan languages so greatly resembled one another despite their diverse genetic origins. In setting up his proposal, he made a point of contrasting the systematic sound correspondences between morphological units in languages of a language

---

[3]With the languages of Balkan Romani and Judezmo later added to the list.

family with the relative paucity of shared basic vocabulary in a linguistic area. Grammatical correspondence in structure, not form, was the calling card announcing a linguistic area. The spread of features through a linguistic area would then proceed by the same mechanisms as the spread of an innovation through a dialect continuum. Rather than thinking in terms of a "tree" model of language differentiation, this approach promoted the idea of language commonalities spreading like a "wave" from a geographical source of innovation.

Exactly what elements of language structure could signal the presence of a linguistic area across a geographical range quickly became a point of discussion. Jakobson (1938:353), in taking up and promoting Trubetzkoy's theoretical concept, followed Trubetzkoy's definition closely by emphasizing that the languages of a linguistic area possess 'remarkable resemblances' among one another in syntax or morphology, while language families share common grammatical morphemes and words that indicate genetic heritage from a common ancestor. Jakobson went further, however, in proposing zones of shared phonological features, such polytonicity in the circum-Baltic languages or palatalization in Eurasian languages (Jakobson 1931/1971). Whereas Trubetzkoy's definition of a linguistic area was derived from observations of the Balkan languages, and assumed the Balkan linguistic area as a model for identifying other such areas in the world, Jakobson worked with far broader affinities. These proposals anticipate problems that have occupied areal linguistics ever since. If a linguistic area is to be distinguished from a language family, then how are linguistic areas composed of languages from the same language family to be distinguished from mere inherited commonalities among related languages? With language families, regular sound change is the widely accepted mechanism for explaining lexical and phonological correspondences. In addition, it is known that lexical correspondences can result from speaker communities sharing a cultural domain. But lexical correspondences do not necessarily result in structural correspondences, as can be witnessed by the extensive lexical borrowing from Turkish in the Balkans, but relatively few structural borrowings. If the contrast is to be drawn between linguistic inheritance and a linguistic area, then what mechanism produces the structural convergences of linguistic areas? For historical reconstruction, the systematicity of correspondences is key for distinguishing coincidental similarities between words in unrelated languages from inherited lexical items indicating a familial relationship. For linguistic areas, how is accidental coincidence of structure to be distinguished from the outcomes of contact between languages?

By expanding the possible scope of a linguistic area, Jakobson extended the theoretical notion, but also brought into question the necessary criteria for determining a linguistic area and identifying the languages that belong to it. In this way, it might be possible to provide a causal explanation for patterns unaccounted for through historical reconstruction, such as the polytonic phonology of languages from multiple families in the area of the Baltic Sea. But there is also a danger in allowing too broad a definition for a linguistic area and rendering the concept unworkable. Ruling out coincidental shared innovations from structural convergence is relevant even for widely agreed-upon linguistic areas like that which defines the Balkan languages. Some of these definitions of shared features in a linguistic area require more precise delineation of common innovation from inheritance. The establishment of a linguistic area invites the question: where do you draw its boundaries?

6.3 The linguistic area as a concept and a tool

In this section, I will discuss the development of the idea of the linguistic area, and the conceptual problems with viewing languages as part of an area of convergence. In §6.3 I address the view of linguistic areas as the recognition of mere similarities or as the artifact of of a process, which is called 'circumstantialist.' This view relies on the listing of features, which brings complications in the definition of features to be included or excluded from the definition of a linguistic area. The complications themselves may be illuminating, in that they show that linguistic areas are best seen of as the layering of structural commonalities, not all of which must be shared among all the languages in the area. It is also necessary to make sure accidental convergence is ruled out as an explanation, given that languages within an area may be related to one another, as I discuss in §6.3. Given that some features may be inherited commonalities, it is important to take a historical perspective and look at the development of individual features—as I argue in §6.3 and throughout this section, LDO is a particularly vivid illustration of the advantage of this approach. LDO is best viewed as the outcome of multiple clusterings of local contact between speech forms over time and space, and its path of development shows how this view of the Balkan linguistic area as iterated clusters of contact allows for areas of greater or lesser convergence between individual languages, even some that may be geographically separated from one another.

The Balkan linguistic area: Arguing from circumstance

Campbell (1985:26-29) provides an extensive survey of the literature on defining a linguistic area. In it, he notes two ways in which the term linguistic area may be used: as a recognition of similarities among neighboring languages, not necessarily related, or as an artifact of a process of feature diffusion among languages for which there is historical evidence. If the starting point is to merely note similarities, which may be all that is possible for under-documented languages or languages without a historical record, then the designation of a linguistic area is assigned through "guilt by association." With a very broad definition such as that provided by Jakobson, the presence of shared features across languages spoken contiguous to one another is sufficient for defining a linguistic area; Campbell labels such an approach 'circumstantialist.'

Circumstantialist ways of defining a linguistic area by a 'listing approach' of factors, such as found in Seliščev (1925) for the Balkan linguistic area, can be an excellent way of identifying avenues for further investigation. This approach has been applied to numerous other potential language contact situations resulting in linguistic ares, for example the Indian Sub-continent linguistic area (Emeneau 1956), the Ethiopian linguistic area (Ferguson 1970), and the Meso-American linguistic area (Campbell et al. 1986). Thus Sherzer (1976:132), in defining a linguistic area, emphasizes that member languages may differ, but they must share several diagnostic traits, spread by a plausible mechanism of contact between speakers (generally geographic contiguity). What constitutes a diagnostic trait can be complex to determine. For the initial traits that Sandfeld used to define the 'Balkan type,' it helped that there were related languages outside of the Balkans that lacked the features. Thus, sharing of given features among the Balkan languages, along with the lack of these features in relatives outside the Balkans, was sufficient to designate a feature as diagnostic. In this approach, a linguistic area is sufficiently defined *both* by commonalities among a multiplicity of features and by the social circumstances of speaker contact necessary to facilitate convergence. But even in the case of the Balkan linguistic area, often considered one of the

best studied and most well-defined subjects of study in areal linguistics, there is disagreement on the features that count as 'Balkanisms,' the diagnostic traits that mark a language as Balkan. If a linguistic area is to be defined through feature counting, then how do you decide which features count?

Taking into consideration the sort of convergent features that Sandfeld identified, Joseph (1983:247) identifies multiple ways of determining whether a linguistic feature is a Balkanism:

1. Any similarity between or among Balkan languages

2. Any similarity between or among Balkan languages that is an innovation vis-à-vis their common ancestor language (in this case, Proto-Indo-European)

3. Any similarity between or among Balkan languages that is due to language contact

4. Any similarity between or among Balkan languages that is unique to the Balkans

In the case of LDO, chapters 2-5 established that is is shared among Balkan languages (1) and an innovation in comparison to their common ancestor language (2). Circumstance and diachrony also strongly suggest that LDO must have arisen due to language contact (3), and the particularly configuration of features that define LDO appear to be unusual, and perhaps unique to the Balkans (4). In this way, LDO could be easily viewed as a feature that defines the Balkan linguistic area. Yet its uneven distribution would make it a relatively poor gauge of the linguistic area as a whole. Without further analysis, it is not clear how a language like Istro-Romanian, very far removed from the core Balkan range proposed by Kopitar, should have come to developed LDO while Bulgarian, situated between Romanian and Macedonian and sharing many of the core Balkan structural features with both, should fail to show LDO. Moreover, the existence of LDO depends on other shared features, making LDO easy to overlook in favor of more immediately identifiable features like a goal/location semantic merger or post-posed definite article. Nonetheless, I argue that LDO came about under the same processes that resulted in other 'core' Balkanisms. The identification of LDO and the understanding of its development is complicated, but such complications apply to any of the Balkan structural features when one looks closely.

From the very first, counting and listing features, even for well-studied languages like those of the Balkans, can be a fraught procedure. One problem is defining what it means for a feature to be 'shared.' If features are counted according to whether they are present in all the proposed members of the linguistic area, the inclusion of any new language inevitably dilutes the count. Lindstedt (2000:232), for example, focuses on twelve frequently noted Balkanisms and produces an index of their presence in individual languages that have been proposed as members of the Balkan linguistic area, reproduced here as Table 6.1.

Table 6.1: Lindstedt's (2000) grammatical Balkanisms

	Modern Greek	Albanian	Balkan Slavic	Balkan Romance	Balkan Romani
argument marking					
enclitic articles	(+)	+	+	+	−
object reduplication	+	+	+	+	+
prepositions instead of cases	(+)	(+)	+	(+)	(+)
dative/possessive merge	+	+	+	+	(+)
goal/location merge	+	+	+	+	(+)
relativum generale	+	+	+	+	+
verb system					
Aux (+Comp) + finite verb	+	(+)	+	(+)	+
'will' future	+	(+)	+	+	+
past future as conditional	+	(+)	+	(+)	(+)
'have' perfect	+	+	(+)	(+)	−
evidentiality	−	+	+	(+)	(+)
other					
analytic comparison	(+)	+	+	+	+

At first glance, the table presents a compelling picture of commonalities. But sometimes one must strongly qualify the circumstance to determine whether a feature is 'shared.' Features designated with the (+) sign are arguably attested in the listed language or set of languages, but this means different things depending on the language. Albanian, for example, does not replace nominal case entirely with prepositions, unlike most Macedonian dialects, which have entirely replaced the system of noun cases inherited from Common Slavic with prepositional constructions. Though the case system of Albanian has undergone numerous changes, it retains a robust system of case inflection reflected in nouns, definite articles, and modifiers. It has shown an expanded use of prepositions in place of nominal case over time (Mansaku 1994:80), but without anything approaching the near total loss of nominal case in Balkan Slavic. Thus the reduction of cases is particularly important with Balkan Slavic languages, where the retention of complex morphological case in their relatives is easy to note. But it is a less defining feature for Albanian or Balkan Romani, which both feature extensive nominal case morphology. And in the case of Balkan Romance, the label may even be misleading, in that Romanian retains *more* nominal case morphology from Latin than other Romance languages outside of the Balkans. To designate the 'loss of case' in Romanian as one of its notable Balkan features in the same way as this feature applies in Balkan Slavic would imply the existence of a set of other Romance languages showing more nominal case. However, the retention of case is what distinguishes Romanian from its genetic relatives, not its loss. Indeed, the Romanian case system resembles that of Albanian in significant aspects. Thus, the implication of a 'shared' feature is rather different depending on the context of the individual language and its history of development.

The 'will' future is another example of a shared Balkan feature that must be qualified. In the case of the 'will' future, in which the future is marked by a particle derived from the verb 'to want,' Albanian does in fact possess this feature in the standard dialect. In many dialects of Albanian,

however, the future is formed with the verb 'to have.' Moreover, this difference in the expression of future tense cannot be neatly divided along Albanian-internal dialectal lines. While the 'will' future is generally associated with southern Tosk dialects, which are often thought to be more 'Balkanized' than northern Geg dialects, there are in fact Tosk dialects that use the 'have' future, and Geg dialects that use the 'will' future. This considerably complicates the determination of whether the type of verbal expression for future is a mark of the Balkan linguistic area in the structure of Albanian overall, or whether only individual Albanian dialects have participated in the process that yielded the 'will' future. (Friedman 2005:36-37). These detailed considerations of individual structural Balkanisms show that it can be more fruitful to treat Balkanisms as individual units of analysis, as I have done in this dissertation, and observe how they spread among some speaker groups but not others in a linguistic area, rather than focusing on what the broadest analytical category of 'languages' share or do not share. Indeed, this approach toward breaking down linguistic area features into individual units of analysis and abandoning the emphasis on group cohesion has been recently advocated as a program of a study for areal linguistics by Enfield in his definition of the Mainland Southeast Asia linguistic area (2005:198).

There are thus two ways in which a feature-cataloging approach such as that exemplified by Table 6.1 simplifies the true situation of the Balkan languages. One way is that it does not specify whether a feature is an innovation in a given language or whether it is a conservatism retained from an ancestor language, as with nominal case in the Balkans. The table also depicts languages as if they are monoliths for which a binary measurement of feature presence or absence is possible. However, some languages show individual differentiation according to dialect, with certain dialects sharing more Balkan features than others. Other languages show a single feature uniformly across all dialects. This is an important difference, which may be connected to the diachronic structure of the language. Together, these objections to the feature-cataloging approach show how inheritance and diffusion can be difficult to separate in the study of linguistic areas, but how important it is to keep both possibilities for linguistic differentiation in mind when studying commonalties of a given linguistic area.

When it comes to the specific feature under study here, complexities of both dialectal and diachronic distribution are important to LDO. As regards diachrony, in Balkan Slavic, this feature is clearly a recent innovation. In Balkan Romance, LDO is a conservatism retained from before the documented record. In Albanian, a mixed situation prevails—the precursors to LDO long existed in the language in the form of locative case. The semantic conditions that trigger LDO were already associated with the locative case. But the loss of the locative and the introduction of determiner omission in its place is an innovation in Albanian. In dialectal distribution, LDO is a remarkably uniform feature of Balkan Romance, but has a very local dialect distribution in Macedonian. Once more, Albanian stands out from these two for the unevenness of its dialect distribution, with northern and southern dialects of Albanian showing loss of locative case earlier than central dialects, but full development of LDO happening first in southern dialects, while peripheral and non-contiguous dialects lack it entirely. This sort of complexity applies not only to LDO, but to most features of the broader Balkan linguistic area.

While Table 6.1 demonstrates extensive commonalities between individual languages in the Balkans, it shows few universal properties shared by all the proposed member languages. When holding the set of diagnostic features constant, only the presence of a generalized relative clause is unambiguously shared by all. The use of a generalized relative is remarkably *common* across the world's languages, and would hardly be sufficient as the basis for the proposal of structural

convergence between languages. When seen in aggregate, the extent of commonality in features is compelling, but when individual features are examined, the picture of Balkan linguistic area membership becomes complex. The picture that emerges depends on whether one distinguishes individual related languages (eg, distinguishing Romanian from Aromanian in Balkan Romance or Bulgarian from Macedonian in Balkan Slavic) or dialects of the same language (ie, the Geg and Tosk dialects of Albanian, or the several dialects of Balkan Romani). And yet, shared grammatical features are considered key for defining the Balkans as a linguistic area; the identity of the features themselves is vitally important to understanding the situation of Balkan languages as distinct from their genetic relatives or their 'non-Balkan' neighboring languages.

It may be noted that the set of features in Table 6.1 is not the same feature set as originally proposed by Sandfeld. Tomić (2006:24) gives an extensive overview of proposals for Balkan features, and takes a different tack. Rather than holding the set of features constant and noting which languages possess them, as in Lindstedt's overview, she presents data on which features major researchers working with Balkan languages have proposed to be Balkanisms, yielding a somewhat different set, reproduced here as Table 6.2.

Table 6.2: Prominent Balkan Sprachbund features discussed by individual authors

	Sandfeld (1930)	Schaller (1975)	Solta (1980)	Gołąb (1984)	Lindstedt (2000)	Campbell et al. (1986)	Thomason (2001)
postpositive articles	+	+	+	+	+	+	+
prepositional cases	+				+		
dat/gen merger	+	+	+		+	+	+
location/direction merger	+				+		+
object reduplication	+	+			+	+	+
two direct objects	+						
dative clitics in NP		+		+			
vocative case markers			+				+
loss of infinitives (and use of subjunctives)	+	+	+	+	+	+	+
'will' future	+	+	+	+	+	+	+
"will" future in the past + subjunctive				+			
'have' perfect				+	+	+	+
'have' past perfect					+		
future perfect (used as conditional)				+	+		
general relativizer					+		
periphrastic comparison				+			
'and' instead of 'but'	+						

There is only agreement on three features across all authors: the loss of the infinitive, the post-posed definite article, and a future tense using an auxiliary derived from 'want.' Combining data from Table 6.1 and 6.2, it can be seen that Greek and Romani are often considered to be Balkan languages, but neither language possesses post-posed definite articles, and instead show only pre-posed articles. The features agreed upon universally, then, are only reduced to two: loss of the infinitive and the 'want' future auxiliary (Alexander 2000b:11). If one insists on a multiplicity of shared features, this barely qualifies the Balkan linguistic union as a coherent entity. For example, forming a future tense from the auxiliary 'want' is quite common crosslinguistically, and so this seems to be a shaky foundation upon which to build a hypothesis that the grammatical resemblances of the Balkan languages are a result of feature diffusion, rather than coincidental and independent developments due to language-internal pressures. Nonetheless, Weinreich (1953:378-379), Katz (1975:16), and Masica (1976:172) argue that as little as a single synchronic isogloss between multiple languages is sufficient for designating a linguistic area, noting that there is no logical distinction between languages that share only a single trait and languages that share multiple traits. Campbell (1985:29) takes the difference between linguistic areas determined by a single diagnostic trait and those determined by a cluster of isoglosses as one of degree, with some linguistic areas being stronger and better established than others. Taking into account this concept of intensity, the Balkan linguistic area then seems to be defined by a set of only a few core features, but numerous more peripheral features that are shared by some, but not all of the languages in the area.

One way to resolve this is to consider Balkan languages as part of a spectrum, with some languages treated as more core to the linguistic area and others considered more peripheral—this is the approach that Lindstedt takes in response to the partial attestation of his defining features across the languages. Thus, a Balkan Slavic language like Macedonian, where all features are attested, would be treated as more core to the Balkan linguistic area than Balkan Romani, for which Balkanisms are more sparsely attested. But proposals of an absolute hierarchy of "more or less Balkan" languages based on counting the number of features have failed to garner consensus (Harris and Campbell 1995:120-50; van der Auwera 1998). Lindstedt, in his approach, posits Albanian, Macedonian, Bulgarian, and Aromanian as the most Balkan languages. Other authors, such as Topolińska (2007:66), have narrowed this down to highlight Macedonian as the 'most Balkan' of the Balkan languages. But Campbell et al. (1986:561) has been argued that Romanian is the 'most Balkan' language from a probabilistic approach based on only enumerating features. While listing features has been an important aspect of defining the Balkan linguistic area, and has proven useful to the definition of other linguistic areas (see Aikhenvald and Dixon 2006; Enfield 2008; Matras and Sakel 2007), it is not sufficient to account for a linguistic area, or to differentiate between different kinds of linguistic areas.

To this point, Joseph has proposed that one can speak of the 'broad' Balkans and the 'local' Balkans (Joseph 2009:125): some features appear to be common across all Balkan languages, while others are confined to a particular set of dialects or a particular subset of languages in contact with one another. On their own, the features that define the "broad Balkans" seem to delimit, but not define the Balkan language area. No one standard Balkan language has all the Balkan features that can be found in local dialects (Topolinska 1991:102). If one limits the Balkan linguistic area to the languages that most resemble one another and which constituted the earliest proposal of a Balkan linguistic league—Albanian, Balkan Romance, and Balkan Slavic—then the overlapping isoglosses are very dense. If one includes other languages of the Balkans, such as Romani, Judezmo, and

Turkish, the total shared features decrease. Thus the factors that produced the distinct pattern of the Balkan linguistic area must be examined in particular, between individual languages, before backing away toward the big picture of the linguistic area as a whole. This is reasonable when considering the mechanical necessities of any structural convergence: contact-induced innovations can occur between two languages alone, and so 'linguistic convergence' can apply to multiple languages, or only just two. (Jakobson 1931/1971:145; Thomason and Kaufman 1988:95). Nothing in principle separates this process from processes among three or more languages. In addition, some authors propose that contact between dialects is the same and is governed by the same processes as contact between languages (Alexander 1983; Alexander 1984-1985; Trudgill 1986).

The consideration of how many features are shared between languages, rather than asking merely whether features are shared or not, relates to Schaller's (1997:203) distinction between extensive and intensive linguistic areas, a proposal also originally suggested by Trubetzkoy (Campbell 2006:3; Tosco 2008). While an extensive linguistic area shares common features, particularly phonological, an intensive linguistic area is the result of specific processes of language contact and results in multiple correspondences, often grammatical. One set of grammatical convergences can lay the groundwork for another, and given sufficient time, and overlapping of feature innovations create bear out as an intensive linguistic area. As mentioned above, Jakobson wished to extend the idea of the linguistic area very widely, to include matters such as phonological structure. For this, the idea of an extensive linguistic area defined by more individual, with limited 'layering,' can be useful. Eurasia, for its features of vowel harmony, might be considered an extensive linguistic area. Though vowel harmony is widespread across this geographical zone, it is also one of the few universally or widely shared features of the area.

The difficulty of pinning the linguistic area down as a concept that can be consistently defined has led some authors to reject the idea of a 'linguistic area' in general as too difficult to employ as a theoretical tool (cf. Dahl 2001, Muysken 2000, Stolz 2002, Tosco 2008). Campbell (2006:459), despite extensive work in contact linguistics, considers the term 'linguistic area' to be too fuzzy, seeing it as just a case of localized diffusion extending across a wider area. For other authors, however, it is more useful to think of localized diffusion as something that can occur in degrees of intensity, with only dense clusters of localized diffusion constituting a true linguistic area, even if it has more diffuse peripheries. The Balkan languages show many more correspondences than the Eurasian languages, and so offers a viable example of an intensive linguistic area. Moreover, the idea of localized intensity of diffusion can play out at an even smaller scale: Balkan phonological convergences, for example, are local, not extensive across the Balkan linguistic area (Friedman 2006a:660), and Friedman (1994) notes that the 'Balkanization' of Albanian morphology depends on local contact and diverges across dialects. This might be rather surprising, considering that phonological features are sometimes the sole means of defining extensive linguistic areas such as the Eurasian zone of vowel harmony. Friedman (2008a:143-145) proposes that all contact must be thought of as 'local' when considering the Balkan linguistic area, even when approaching features that are widely shared across a broad geographical range of speaker populations in the Balkan. These commonalities, under a localist interpretation, are merely the result of iterated local diffusion, reinforced by structural convergences that only extend to a limited portion of the linguistic area, as has been pointed out for phonological convergences (none of which extend across all members or all dialects of the Balkan linguistic area). Rather than taking linguistic areas to be uniform, a linguistic area can instead be seen as a 'cluster of clusters' (Hamp 1989:44), and localized diffusion can be seen as a spectrum of differential bindings (Hamp 1989:47; Joseph 2010:628-29). Thus the

strength of a linguistic area, per Campbell's proposal, depends on the density of localized diffusions, and these depend on the time-depth and the social dynamics of contact between speakers and speech varieties.

Ruling out accidental convergence in a linguistic area

If only partial sharing of features is admitted as a diagnostic for a linguistic area, then another problem emerges: deciding whether shared features between related languages, however, distant, should be counted toward the determination. Especially if the feature is also attested in related languages outside the proposed linguistic area, it can be difficult to distinguish between convergence by social mechanisms of contact and inheritance from an earlier ancestral language. The *habeo*, or 'have' perfect, for example, refers to the formation of a perfect verb tense with the verb 'to have.' This is a noteworthy feature of Macedonian, a member of Balkan Slavic, though it is not a fully developed part of Bulgarian. But the presence of a 'have' perfect in Balkan Romance is unremarkable, in that non-Balkan Romance languages also make regular use of a 'have' perfect. In general, this means of marking the perfect tense is found in many Indo-European languages. To illustrate this point, Asenova (2002) lists linguistic areas where Balkanisms are found, many of them outside the Balkans. Such a point may be carried to absurdity. Aronson (2007), working purely from the presence of diagnostic features that had been labeled 'Balkanisms,' argued facetiously for 'English as a Balkan language.' While Aronson did not seriously intend to include English among the set of Balkan languages, his analysis of the ways that English repeats many of the features considered distinctive of the Balkans illustrates how possession of diagnostic traits is insufficient to define a language as a member of a linguistic area without also taking into account the social and linguistic mechanisms by which the traits came to emerge.

If a feature is widely attested outside of the Balkans, then what makes it particular to the Balkans? Excluding genealogical factors can be complicated for the Balkan linguistic area, and any linguistic area in which related languages occur together. As Nichols and Bickel (2006:2) note, the list of classic Balkanisms is selected from a potentially open-ended set of syntactic, morphological, lexical, and phonological variables. Features that are relatively common across the world's languages are of little diagnostic value in distinguishing between grammatical convergence and coincidental independent development. Although the classic Balkanisms are not particularly common in the world's languages (Nichols and Bickel 2006:2), this still does not fully account for the shared grammar which authors have repeatedly noted as allowing for unusual intertranslatability among the Balkan languages of different Indo-European branches, when compared to their relatives outside the Balkans (Nichols and Bickel 2006:3-4).

One possible way around the problem is to rule out genealogy by definition. Emeneau (1980:1) thus takes a different circumstantialist tack and defines linguistic areas by contrast: languages from more than one family, possessing common traits that do not belong to other members of at least one of the families. Contrasting with an external relative is important, but not always possible, as when the only attested member(s) of a language group are within a proposed linguistic area. For the Balkans, this is particularly true of Albanian, which cannot be easily compared to an external, related member, since it is a divergent branch of Indo-European unto its own. What is attested of Albanian exists only within the Balkan linguistic area, although differences between Albanian dialects can be productive in determining intensity of contact and the time-depth of common features. The post-posed definite article, for example, has been long attested in Albanian, and toponymic

evidence suggests that it may have even been a feature of Albanian before contact with Balkan Latin, although such evidence is only suggestive. This feature, then, can be considered a very early Balkanism, since it does not appear in Late Latin or in Late Common Slavic, whose descendants only developed it later. The loss of the infinitive, in contrast, seems to have been a more recent phenomenon in Albanian, an innovation that appears in dialects most heavily in contact with Aromanian and Macedonian, but not in dialects spoken in more mountainous, remote, and monolingual portions of the Albanian-speaking geographic range (Joseph 1983:98-100). Contrast against outside members lends diagnostic features power and helps to clarify timelines of development for linguistic areas.

Even if features can be agreed upon as distinctive, there is a danger of circular reasoning in a feature-counting approach to linguistic area identification. When looking at broadly at features that have been noted as Balkanisms and languages that are considered Balkan languages, there has been a tendency to argue that a language is Balkan if it possesses more than one Balkan feature, but also, that a feature is Balkan if it is attested in more than one Balkan language (Desnickaja 1979:121). As noted above, many features that are attributed to a Balkan language as a whole, in fact, only exist in some dialects, and so an approach that reduces the patterns of convergence to the feature–language unit is insufficiently explanatory, particular for features with more limited distribution. A focus on numbering of features according to language obscures processes of development, necessary periods of contact, and the sequence of development as one feature builds on the next. Rather, defining a linguistic area such as the Balkans also requires us to provide a mechanism for how convergence could have occurred. A view of the Balkans as center and periphery, with gradual spread of innovations outward from a core over time, still does not account for some features that do not have a central peripheral distribution, like LDO in its modern distribution in Romanian, but also Macedonian, Albanian, and Aromanian. For this purpose, it is more informative to view linguistic areas as the product of layered, localized bindings between speech form, sometimes subject to social or geographical reconfiguration. With sufficient time-depth, this may produce a region that appears to show widespread convergence. But the historical details for each feature, and the pattern of spread between individual languages, must be examined in detail.

The historicist approach to linguistic areas

In contrast to the circumstantialist approach of feature counting and cross-comparison, Campbell promotes an approach he calls 'historicist' that insists on concrete evidence establishing that common traits in a linguistic area were the result of diffusion between languages. In his view, this is essential before designating languages as part of a linguistic area, to avoid the problems of determining whether sharing of features is a sufficient criterion. The earliest research into the Balkan linguistic area was only speculatively historic. Early authors observed commonalities, noted how those commonalities rendered Balkan languages distinct from their relatives, and proposed possible mechanisms that often depended on an assumption of one single language as the source of the putative convergences. With Sandfeld, mechanisms for contact were already a part of the analysis, and he proposed sources for the structures in an individual language, establishing a parallel between structural and lexical borrowing. This was surmised from synchronic evidence of outcomes, but relied on a reconstruction of historical processes. Indeed, ongoing contact need not necessarily be required to identify a linguistic area. While linguistic areas can be seen as synchronic phenomena, their strength and composition depends heavily on the social and linguistic mechanisms that

underlie their growth. Masica (1976:173) notes that some linguistic areas are the result of historical processes of speaker contact that are no longer active, but that have left their trace in the grammatical or phonological properties of the languages in an area, while other linguistic areas reflect ongoing processes of speaker contact driven by ongoing speaker interaction. In regards to the Balkans, Topolińska (1998) argues that the Balkan linguistic area is just such a historical artifact, rather than an ongoing process. Other authors, such as Friedman (2007) have argued that Balkan processes of convergence are ongoing in localized areas, even though convergence across the Balkans may no longer be ongoing. In the case of LDO, multiple layers of these traces must be tracked in order to account for geographical gaps, such as between Romanian and the other Balkan languages showing LDO. An undocumented process of speaker contact between Balkan Romance and an early form of Albanian appears to have provided the basis for the development of LDO in Balkan Romance, but a secondary and later period of convergence was necessary to account for the introduction of LDO into Albanian with the loss of the locative, and the introduction of LDO into Macedonian after the development of the post-posed definite article. Balkan Romance as a unitary whole, deriving from Balkan Late Latin, must be the unit of analysis for the period of contact suggested here wherein the Albanian locative provided the model for Balkan Romance LDO. Thus, the occurrence of LDO already in all Balkan Romance branches means that we do not have to account for its spread into Romanian, given that this was inherited from an earlier stage of Balkan Romance. However, the later spread of LDO into Albanian and Macedonian must be traced to the influence of Aromanian, in particular, given the different early modern distribution of Balkan speech forms at the end of the Ottoman Empire and the beginning of the age of Balkan nation-states.

As the circumstances of Balkan Romance illustrate, conditions that suggest a linguistic area synchronically may not be relevant to the original development of convergent features, or may only be relevant to some features, but not others. In light of this, Jacobson (1980:2) calls for a historical program of tracking diffusion to go beyond the "mere cataloging of the presence of absence of a category in a language," instead emphasizing means. Whether the impetus for convergence in the Balkans is viewed as historical or ongoing, a specific mechanism must be proposed that is relevant to each particular convergent feature (Hamp 1977:281). This may lead some features to be ruled out as the result of language contact, if they can be Linguistic areas must be defined by social relationships, not by genealogical relationships. The lack of genealogical explanation for commonalities is what marks out a linguistic area, and so the mechanism for its formation must be through social relationships mediated by two or more languages (Campbell 1985; Hamp 1977). Thomason (2001:99), summing up the consensus among those who make theoretical use of the concept of linguistic area, defines it as "a geographical region containing a group of three or more languages that share some structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor," with Muysken (2008) breaking this down into the key factors of (a) geographical region, (b) more than three languages, (c) shared structural features, (d) language contact, (e) ruling out accidental conjunction, and (f) ruling out genetic inheritance. It is thus a question of particular importance to defining a linguistic area whether a feature is a contact-induced innovation.

Any linguistic area must proceed from encounters between languages over time. But it is not languages themselves that engage in contact; rather, it is speakers of particularly languages that interact with one another and react to one another. As a cultural and psychological phenomenon, language contact is subject to interpersonal mediation. As such, language contact is actually speaker contact (Friedman 2006). This can be viewed as phenomenon occurring over a broad scale, between

social groups, or it can be viewed as an individual process, with a locus of contact in the mind of the speaker. Assessing the motivations and abilities of every speaker to have participated in the formation of a linguistic area is quite impossible. Nonetheless, something pertaining to their motivations, and corresponding language contact outcomes, can be assessed in the aggregate. This is essential to the determination of a historical process that fulfills Muysken's (2008) criteria of language contact and ruling out accidental conjunction. The mechanisms of language contact, then, are a necessary part of the study of linguistic areas such as the Balkans.

6.4 Mechanisms and motivations in language contact between speaker communities

There are multiple ways of looking at the mechanisms of language contact. The broadest view encompasses populations as a whole undergoing language shift and language imposition. In this way, contact is treated as between communities of speakers, and the properties of those communities in relation to one another are taken as a determining factor in the outcomes of contact. In some cases, linguistic contact implies linguistic imposition; an invading force speaking different language(s) requires a subject population to learn and use this language for communication. Or, an overwhelming majority subject population, in being more numerous than the conquerors, gradually impose their own language to a point at which it is uniformly spoken across all social classes of the resulting community. English, in its modern form, can be taken to be a result of both processes—imposition of Norman French by a conquering force in domains of law, religion, politics, the arts, and the court over a stratum of Germanic Old English spoken by the subject population, with the eventual spread of a highly Normanized from of English used in the wider community across the entire span of the social strata, resulting in Middle English. Dislocation of a community, too, can result in situations of language shift. Considered in another timeframe, language contact can be the result of ongoing coexistence and speech economies, with efficiency of communication and communal harmony exerting pressure toward convergent language changes. All of these processes have been applied to the study of the Balkans.

Prestige and language layers

A communal, historic approach to language contact, and one that was very prominent in the early work on the Balkan languages, was the idea of language strata. Terms such as "superstrate" and "substrate" refer to the state of speaker communities in contact prior to widespread language shift. In a situation of a substrate, speakers from a language community with lower prestige shift into a language with higher prestige—this shift leaves traces on the target language in the form of grammatical or phonological changes, which can be traced to the substrate language. In a situation of a superstrate, a higher prestige, usually smaller speaker community shifts into a lower-prestige target language, but the difference in power leads the target language to converge toward the superstrate language in key domains, often those associated with the original power-holders (Trask 2000). Substrate and superstrate serve as counterparts, with a substrate language contributing certain elements to the synchronic outcome (such as grammatical patterns) and the superstrate contributing other elements (such as vocabulary). While substrate and superstrate are often used to describe multilingual

dynamics that result in a single language at a later point, it is possible to have substrate/superstrate dynamics without shift (Goodman 1993:64-65).

In the aggregate, stratum is not a sufficient explanation for the grammatical commonalities of the Balkan languages, although there have been several proposals in this direction. Kopitar (1829) and Schleicher (1850) both assumed that the commonalities of the Balkan languages must have resulted from language shift, with a paleo-Balkan language (particularly Thracian) providing the grammatical template for changes in Late Latin and Late Common Slavic, with Albanian, Balkan Romance, and Balkan Slavic as the result. This is the 'substrate' hypothesis. But a substrate hypothesis of paleo-Balkan origin for Balkanisms only accounts for the outcome of Balkan Slavic if the timespan from the imposition of Latin with the Roman Empire to the settling of the Slavs is compressed. While Late Latin and proto-Albanian must have been in contact immediately from the introduction of Latin into the Balkans, as evidenced by toponyms, shared vocabulary, and deep grammatical convergences, Slavic arrived much later, and yet Balkan Slavic exhibits the full range of Balkanisms (Wahlström 2015:13). These grammatical features were not present in Slavic at the time of its first documentation in the Balkans, from the 10th century onwards. The earliest attestations of Old Church Slavonic preserve Late Common Slavic grammatical inheritances such as a synthetic case system, a where/whither distinction, separate marking for dative and genitive case. Only later did Balkan Slavic develop its grammatical commonalities with Albanian and Balkan Romance (Lindstedt 2000:235). A paleo-Balkan language could not have served as a substrate for the grammatical changes in Balkan Slavic. In addition, if we examine the other, earlier end of the timespan from Latin imposition to the Slavic invasions, there are also problems with identifying a paleo-Balkan substrate by name. We do not know what the grammars of ancient Balkan languages such as Thracian, Dacian, or Illyrian looked like (Katičić 1976) and we only have a vague sense of where these languages were spoken, if, indeed, they were different from one another, or cohered together as identifiable language groups. While they were mentioned in Roman works, the languages are scarcely attested. The proposal of a paleo-Balkan substrate for Balkan linguistic convergence rests on insufficient data and a confusion of the timeline for contact between the respective speaker communities of the linguistic area.

Working from the attested historical record, Late Latin has also been proposed as a substrate for Balkanisms (Demiraj 2004:39-42; Gołąb 1984:9-14; Gołąb 1997:15; Solta 1980:283; Steinke 1968:11). Kopitar (1829:86), in his own initial proposal, saw Albanian and Romance together as a Balkan substrate, layered atop an older Balkan form, perhaps ancestral to Albanian. There is considerable evidence that Albanian and Balkan Romance were in close contact with one another for a significant period before the introduction of Slavic into the Balkans, as attested by exclusively shared vocabulary and particular grammatical commonalities such as the agreeing adjectival article, which set these two members of the Balkan linguistic area apart from others. The full range of Balkanisms is not, however, attested in early Latin—they must have developed after the period of Roman colonization, and before Balkan Romance begins to be documented again as early Romanian. In contrast, Greek was proposed as a superstrate by (Sandfeld 1930:212) and Seliščev (1925), with the idea that Greek, as the prestige language of the Byzantine Empire and some social spheres of the Ottoman Empire, would have been a model for prestigious convergence. But the features of Greek that relate it to the other Balkan languages (such as the loss of the infinitive) appear to have developed at the same time as these features developed in other Balkan languages, rather than serving as a precedent. And features of Greek that distinguish it from the other Balkan languages, such as the pre-posed definite article, were never picked up, even when Latin in the Balkans was

developing a grammatical category of definiteness. Instead, Balkan Romance exhibits the post-posed definite article that has been frequently labeled a core Balkanism, alike with Albanian and Balkan Slavic. Looking for one language as a model for imposition of all Balkanisms does not account for the timeline of developments in this linguistic area, nor does it easily apply to every Balkan language feature.

Lindstedt (2000:235-238) rejects the idea of a single Balkan sub- or superstrate. He instead proposes an 'adstratum' explanation in which the languages of the Balkans mutually affected each other, with no pure directionality toward any one prestige language. This is a helpful perspective when considering the language prestige and agentivity configurations of Balkan languages over time.

Prestige refers to the relative power attributed to a language variety based on the social situation of its speakers. The prestige of languages affects the motivation for convergence, and social context affects language prestige (Myers-Scotton 1993; Poplack 1987:57). Often, it is assumed that speakers will shift toward higher prestige speech over time, in an attempt to acquire power through emblematic identification with the distinct speech habits of discrete community. But the degree to which social prestige overwhelms conserved structure will depend on complex factors, as Labov demonstrated in his famous study of American English speech habits in New York City (Labov 1966). Americans at both the lower and upper ends of the socioeconomic spectrum in his study would exhibit speech characteristics that diverged from standard English (the non-rhotacized accent local to New York City). But Americans from the middle class, when prompted, would aim for a rhotacized accent. It was only among the middle class that upward aspiration was forceful enough to trigger a divergence from the local, non-rhotacized norm. Thus, the prestige of one communal group depended on its position between groups of lower or greater prestige.

This concept has been applied to the Balkans in the discussion of the relative prestige of different languages. Turkish and Greek were positioned at the upper end of the prestige scale, as administrative languages of the Ottoman Empire. While there were many non-native speakers of Turkish or Greek in the Balkans, fewer native speakers of Greek or Turkish spoke other Balkan languages. Likewise, Balkan Romani and Judezmo speakers, occupying insular communities, were by default at least bilingual, but with Balkan Romani and Judezmo occurring low on the prestige hierarchy, very few non-native speakers engaged with these languages (Friedman 2001a:148). The languages of Roma and Jewish speaker communities were rarely learned by those who were not born to the communities. In contrast, Balkan Slavic, Albanian, and Balkan Romance existed at more or less the middle of the prestige scale, resulting in pervasive mutual multilingualism among these languages (Lindstedt 2000:242-243). This has been argued to be the explanation for the relatively greater degree of grammatical convergence between these languages than other languages of the Balkans, accounting for some core Balkanisms such as infinitive loss and the post-posed definite article. Occupying the middle of the prestige hierarchy reinforced their position at the core of the Balkan linguistic area, even during periods of migration or language shift and assimilation.

For some groups, a refusal to converge toward a higher prestige variety can serve to reinforce communal solidarity through overt contrast. Romani, in the Balkans, appears to have been protected from some of the convergences that affected other Balkan languages because its conservativisms served an emblematic function. Features such as distinctive aspiration and synthetic case remain in Balkan Romani despite its speaker community's long-term presence in the same multilingual social dynamics as Albanian, Balkan Romance, and Balkan Slavic. In this way, overt linguistic distinction can reinforce communal distinction (Friedman 1999). Social context has a complex effect on

motivations for shift and differences in prestige are not instrumental only in of themselves without consideration of the entire group dynamic of a community in which multiple speech varieties co-exist.

Alongside overt prestige, there can also be covert prestige (Matras 2002:249-50; Mesthrie 2001:381; Trudgill 1972); a language form spoken by a community with supposedly lower power nonetheless serves as a target for borrowing and convergence. This is a particularly important dynamic in secret or trade languages. In the Balkans, Romani shows some covert prestige, serving as a source for slang and colloquial borrowing into Balkan languages spoken by communities with more power. Aromanian vocabulary features very prominently in *purisht*, a trade language of Albanian craftsmen, indicating a covert prestige relationship between Aromanian speakers and this subsection of Albanian speakers. In particular, it is non-standard forms of Albanian where clear Aromanian influences on vocabulary appear; these forms resulted from the colloquial speech in which everyday multilingual interaction between Albanian speakers and Aromanian speakers was a part of regular life (Prifti 2012). In addition, while social prestige based on religious differences would operate against borrowing from Albanian into Macedonian (Gołąb 1959:424), Aromanian communities in Macedonia were urban, based upon a secondary migration from the collapse of Moscopole and other urban centers in southern Albania during the 18th century (Kyurkchiev 2006:16-17, Markovik 2007:51). Aromanians were able to produce literature specific to their community from relatively early in the documentation, and leveraged their fluency in Greek, a language of religion and administration, to promote education in Greek (Konstantaras 2006:96). Aromanians produced multilingual texts encouraging the use of Greek as a language of culture, and these texts reflected their ease with the other languages of middle-level prestige, Albanian and Macedonian. The establishment of an Aromanian *millet*, a distinctive constituent nation of the Ottoman Empire in 1905 supports this point (Schwandner-Sievers 1999:4). Aromanians also had linguistic access to the relatively organized Romanian state, with its own developing literature (Gica 2009). Rural Aromanians participated widely in transhumance and, later, in professions that required considerable mobility. Thus, both urban and rural Aromanian communities exerted an influence over the colloquial and local forms of Macedonian and Albanian in which they came into contact, and also served as a bridge between both communities that were otherwise distinguished by religion (Gołąb 1959:425-426). It is not only that the collectivity of Balkan Slavic, Albanian, and Balkan Romance exist at a middle level of social prestige which would facilitate convergence. In addition, Aromanian was positioned as a middle ground between Balkan Slavic and Albanian, able to exert influence over both through its unique position as a widely distributed, highly adaptive, and consummately multilingual population at the middle of the prestige hierarchy (Balamaci 1991).

The concept of agentivity is also important to the concept of substrate and superstrate—who initiates and who is the patient of language contact (van Coetsem 2000). Van Coetsem (2000:84) distinguishes source agentivity from recipient agentivity. In this framework, borrowing is a form of recipient agentivity, with speakers of the recipient language acting as agents for change to their own language by adopting elements from the source language. In contrast, interference can occur due to source agentivity, in which the speakers of the source language impose changes on the speakers of the recipient language. This could be the introduction of new elements of vocabulary, as would result from a superstrate, or the imposition of new grammatical patterns, as would result from a substrate. Work in areal linguistics can apply a matrix of agentivity to source and recipient languages to derive different contact outcomes.

During the Ottoman period, Turkish was the language to which the greatest prestige was at-

tributed, reflected in the extensive lexical borrowings from this language. Greek also had prestige, given that it was the ecclesiastical language of Orthodox Christianity under the Ottomans, as well as a second language of administration with a continuous literary tradition. In contrast, Balkan Slavic, Albanian, and Balkan Romance existed at roughly equal levels of prestige. Rather than tracing Balkan morphosyntactic features to language shift, Lindstedt attributes it to inter-communal 'negotiation' arriving at a common speech form over time. From among these Balkan languages, Balkan Slavic, Balkan Romance, and Albanian occupied the middle ground, but within this core, Aromanian was particularly influential due to the social position and relatively scattered distribution of Aromanian speakers, both in mountainous rural redoubts and urban centers. The neutral, intimate contact that defines interactions of the middle ground of the prestige hierarchy involves different mechanisms from language shift due to prestige—inter-speaker factors and discourse-bound variables are particular important as mechanisms (Poplack 1987:57). These can be connected from studies of bilingualism and multilingualism to the dynamics of the Balkan Sprachbund, and to other linguistic areas.

Multilingualism and code-switching

Social factors also play a role in code-switching; whether one language is dominant, neither language is dominant, or a mixed language is employed leads to different outcomes. Particularly, continuous, mutual code-switching can lead to the production of 'mixed' languages in which grammatical material has been extensively replicated and shared, along with vocabulary. Examples of code-switching language contact outcomes include Russian-Kazakh (Muhamedova 2004), Turkish-Romani (Friedman 2008b), and Kupwar village (Gumperz and Wilson 1971). The Balkan situation appears particularly akin to that described by Gumperz and Wilson for Kupwar village, in which speaker communities keep themselves separate but mutually interact in each other's languages in areas of public interface. As in the Balkans, the languages of Kupwar village diverge significantly from their relatives outside the zone of sustained, social multilingualism. And while the languages differ in vocabulary, they share a common set of phraseologies and grammatical patterns, which applies to the Balkans as well. Mutual multilingualism, then, is one of the key predictors of the synchronic outcome of languages that show mutual grammatical convergence.

The social conditions of the Balkans promoted multilingualism, rather than unidirectional monolingualism (Lindstedt 2000:241). Agentivity could be mutual and equal, particularly among the languages of Balkan Slavic, Balkan Romance, and Albanian, due to their middle position in the scale of social prestige. There was differentiation within speaker communities, as would be expected from research into the sociolinguistics of single languages, where there is always variety. Gender differentiation, for example, affected the degree to which speakers were multilingual. Men, with greater power and opportunity for contact outside the domestic community, tended to be more multilingual, a situation attested in Aromanian (Récatas 1934) as well as in Bulgarian (Grannes 1996: 4; Stojanov 1952: 218). But Brailsford (1904:85-86) provides a first-hand witness account of Slavic, Albanian, Romance, Greek, and Turkish being swapped interchangeably in a commercial marketplace in Skopje, Macedonia, by both men and women. Code-switching was frequent, and served its own emblematic function, reinforcing community solidarity across language lines (Friedman 1995). Macedonia appears to be the geographic heart of Balkan convergences, with a greater density than at any other point of the Balkan geographical range (Lindstedt 2000:234), but Balkan multilingualism is also attested by Capidan (1937) for Aromanian, which extended outside

of Macedonia, and by Mirčev (1952:124) for Eastern Bulgarian. Turkish and Greek appear to have been protected from extensive change by degrees of prestige that biased toward uni-directional multilingualism rather than mutual multilingualism. The multilingual dynamic in which speakers accommodated one another with equal agentivity was instrumental in the convergence of Slavic, Albanian and Romance. And in contrast to intimate multilingualism, one-way multilingualism yields different outcomes, as in Romani and Judezmo (Eloeva and Rusakov 1990; Friedman 2001; Friedman and Dankoff 1991), which stand at some distance from the other Balkan languages in their grammatical and phonological features.

Aikhenvald and Dixon (2006:26-34) separates linguistic factors from social factors as mechanisms for contact (Aikhenvald and Dixon 2006). If two languages exist in a social situation of mutual multilingualism, structural similarities can ease the process of grammatical convergence. From Aikhenvald's analysis (2006:26-34), I would particularly highlight the following structural factors that affect convergence between languages in a state of mutual multilingualism: (1) pragmatic salience, (2) pre-existing structural similarity, (3) a tendency to achieve word-for-word or morpheme-for-morpheme intertranslatability, and (4) morphotactic transparency and clarity of morpheme boundaries.

Pragmatic salience is important to the level of intimate, conversational interaction. Those linguistic structures which point to contextual variables available to both speaker and listener are easier to identify, copy, and take as a model for structural convergence. Discourse-level daily necessities are the area where multilingual contact occurs (Civ'jan 1965:9), and the surface patterns of discourse are the point where pattern-replication aids in understanding and easing communication (Matras and Sakel 2007). Sharing structures through pragmatic salience can lead to the retention of grammatical features and their reconfiguration toward pragmatic aims, even the preservation of grammatical complexity. Thus Balkan languages show a number of conservatisms when compared to relatives outside the Balkans in the verbal system. The marking of evidentiality in Albanian, Macedonian, Bulgarian, and some dialects of Aromanian is morphosyntactically complex, and distinct for each language, but makes reference to the same discursive category of vouching for evidence and taking responsibility for information—the day-to-day needs that informed Balkan linguistic convergence. More complex, mutual multilingualism results in the strengthening of pragmatic devices into grammaticalized features (Friedman 1994:108). As regards another Balkan factor relevant to LDO, the marking of definiteness, a shared system for marking definite reference in the form of post-posed definite articles depends on the pragmatic salience of definiteness. The wide distribution of the post-posed definite article in the core Balkan languages was facilitated by the function of this grammatical pattern for coding unique, mutually identifiable referents whose identity could be derived from shared context. Once consistent marking for definiteness was shared among the languages of the Balkans, exceptional patterns of behavior could be matched to shared semantic categories.

The feature of LDO, since it is marker of definiteness, is strongly discourse-linked. That brings LDO into the realm of discourse-bound variables, which depend on mutual speaker and listener identification of familiar, specific, or unique referents for which definiteness should be marked in the grammar. This is a pre-condition for the phenomenon of LDO. Without overt definite determiners, there is no structural potential for LDO. The development of definiteness in the Balkan languages, which occurred in all of Balkan Romance, Albanian, and Balkan Slavic before the spread of LDO, was a necessary precondition for the less extensive spread of LDO to Albanian and Macedonian. Macedonian speakers' preference for LDO with highly referential and highly

familiar nouns in the immediate environment speaks to the importance of this grammatical pattern for discourse conditions.

That Macedonian, Albanian, and Aromanian also shared a collapse of the semantic domains of goal and location also aided convergence in regards to LDO. When languages already possess the same category, it is easier to map morphosyntactic material into convergent forms. Thus a pre-existing structural similarity in the form of a shared semantic collapse in the distinction between goal and location aided the identification of a specific pattern of determiner omission in locative phrases. A pragmatic salient feature such as marking definiteness, even if in the unusual form of LDO, can be mapped between shared semantic domains. Mappability can be selected for by the processes of language contact Dombrowski (2013:33); if a language contains two grammatical ways of expressing a particular meaning, one of which allows for surface-to-surface copying of structure (an analytical morpheme rather than a synthetic declension, for example), this will be easier to identify and map from source to recipient language, and its use will be encouraged in situations of multilingualism and code-switching. Thus a specific case form for the locative semantic domain, as found in Albanian, could give way under pressure from Aromanian in which omission of the determiner distinguished the same semantic domain. Patterns of determiner omission with bare nouns and accusative case determiners with modified nouns easily map between Albanian and Aromanian, where accusative case and the locative semantic domain in prepositions is shared. A distinct morphological locative case does not as easily map between Albanian and Aromanian.

Here, the distinction between different kinds of language change proposed by Benveniste (1974:127) is a helpful theoretical model. Benveniste distinguishes between 'conservative' and 'innovative' changes in the transformation of linguistic categories. Conservative changes involve the preservation of a grammatical category, but with a formal substitution of surface expression, as in the replacement of case endings with prepositions for marking the same semantic case categories. In contrast, innovative changes involve changes leading to the loss of a grammatical category, such as the shift from a three- to a two-gender system of noun declension, or the creation of a new grammatical category, such as the development of evidentiality in some Balkan language verbal systems. The two preconditions for LDO of the goal/location merger and the category of semantic definiteness provide a preexisting grammatical category. In shifting from locative case to LDO, Albanian underwent a conservative change, in that the expression of definiteness in locative prepositional phrases by a distinct morphosyntactic means was fully preserved. Only the formal, surface expression of this grammatical category changed, from a distinct locative-case ending -*t* to the pattern of determiner omission for unmodified nouns and definite accusative articles for modified nouns that we see in Aromanian. Multilingualism and code-switching between these languages would have put pressure on the locative case and pushed Albanian toward a shift into grammatically integrated LDO, for which Albanian already had available morphological material. With the locative case already undergoing erosion throughout Albanian, and with Aromanian speakers rapidly shifting into southern Albanian dialects, there was scope for a conservative change to the expression of the category of definiteness in Albanian, without requiring the loss or innovation of a grammatical category.

Leveling of morphosyntactic distinctions can be conceptualized as a means of accommodating foreign speakers, easing mappability. One Balkanism, the loss of the infinitive, has been argued as such an accommodation, where the pressures of accommodation led speakers to opt for a more transparent morphosyntactic structure (Joseph 1983:179-213). Accommodation has been linked to the process of Koinezation (Kerswill 2002), the production of a simpler form of the language for

inter-communal speech. LDO, too, involves the introduction of a more transparent morphosyntactic structure to Albanian. Rather than a complex alternation between a definite accusative case and a variant definite locative case, Albanian developed a pattern of determiner omission/definite accusative alteration that easily mapped to the structure already present in Aromanian, drawing on case-marking resources shared by both languages, in that LDO applied only to accusative-selecting prepositions expressing locative semantics. Mufwene (2008:133-159) takes a substrate as a base for creolization and Topolińska applies the idea of creolization to the Balkans. This has been offered as a modern answer to the question of why the Balkan languages of 'degraded' grammatical structures, lacking particularly in synthetic noun cases, when compared to their neighbors or historical predecessors. It has been argued, significantly by McWhorter (2001:84), that creolized grammars are by nature simpler grammars (see also Wahlström 2015), in that they show greater morphemic transparency. Albanian, Balkan Slavic, and Balkan Romance, all underwent extensive contact, with each serving as a kind of substrate to the other under varying circumstances of intimate, conversational contact, code-switching, and language shift. In a last stage, particular to Aromanian and associated with the shift of Aromanian speakers into Albanian a further grammatical simplification occurred for Albanian, a conservative change that preserved the distinction of semantic definiteness in locative prepositional phrases, resulting in the development of LDO.

In Macedonian, where the complexities of case that define Albanian and Aromanian are lacking, the factor of morpheme-for-morpheme intertranslatability would exert sufficient pressure to introduce what is in fact more complex behavior. With a more general post-positive definite article lacking case distinctions, all that Macedonian needed to adopt from Aromanian was the distinct treatment of the locative semantic domain through determiner omission. Morpheme-to-morpheme mapping could be easily achieved in Macedonian without any shifts to the language's morphological structure; the accommodation of special behavior in the marking of definiteness in certain prepositional phrases would suffice. Intertranslatability between the morphemes of Aromanian and Macedonian in marking definiteness would be more transparent to speakers than separate systems for distributing definite articles between the two speech forms. The changes involved in all of these languages involve changes to surface patterns of the use of the definite article, in the form of a single article morpheme. This is the sort of surface-oriented change that argued to occur most easily in multilingual contact situations (Joseph 2001:23-24, Friedman 2006:671). Rather than developing entirely new, underlying grammatical categories as has sometimes been proposed to be a part of the language contact process (see Heine and Kuteva 2005), the development of a structure such as LDO builds on previously introduced similarities. The post-positive definite article, found in the same syntactic position in Balkan Slavic, Balkan Romance, and Albanian, is syntactically transparent across these languages. The locative semantic domain is assigned to the same prepositions which are intertranslatable with each other, thus along the trigger domain for determiner omission to be mapped surface morpheme to surface morpheme. The modification condition depends not on underlying nominal syntax but on the overt presence of a nominal modifier of any kind. Iterated periods of convergence and contact between the core Balkan languages provided the structural basis for mapping LDO among them.

6.5 Summing up the literature

As can be derived from this survey, the concept of the linguistic area cannot be defined solely by geographical extent and the sharing of features without also examining the localized interactions of languages within the linguistic area in concrete, historical and social detail. There are significant problems with the listing, circumstantialist approach to linguistic areas: it doesn't differentiate between weak linguistic areas such as the Eurasian zone of palatalization proposed by Jakobson and strong linguistic areas such as the Balkans or the Meso-American linguistic area. Without social context and comparative historical work, it is difficult to distinguish between independent developments and contact-induced innovations, particularly with features that are not typologically rare. It can be difficult to achieve consensus on what counts as a defining feature of a linguistic area; does it have to be shared between all languages in the linguistic area? How do you define a language as part of the linguistic area unless you've already established shared traits? What happens when languages only possess some of the traits, in a cline of overlapping isoglosses?

An answer to this question, briefly sketched, is that linguistic areas are not a spectrum, but rather clusters of clusters, where the density of localized interactions leads to specific contact-induced innovations. Individual clusters are interconnected with one another. Some innovations develop across the entire interconnected network, while others are confined to individual clusters, with dialects differentiated from one another; this three-dimensional mapping of the linguistic area better fits the mechanisms of language contact supposedly underlying the development of a linguistic area than a two-dimensional approach of counting and mapping, as if languages are in contact with one another rather than speaker communities.

The mechanisms for language contact are always localized to individual speaker communities: bilingualism, interference, prestige, and feature selection depend on local conditions specific to the discourse setting, rather than broad features of languages writ large. In the context of the Balkans, the defining feature of the linguistic area has been called a Balkanism, but often these are evaluated in terms of listing and documenting their extent or lack thereof. A Balkanism, however, needs to be defined precisely, with reference to the specific interactions of the languages/dialect that possess that feature—Balkanisms that extend across the entire linguistic area still exist as a result of localized interactions, and Balkanisms local to a specific cluster still help define the linguistic area through this approach that analyzes iterative layers of contact between specific speech forms and communities situated in the historical record. LDO, in particular, could not have developed without the layering of interactions between speech forms that formed the Balkan linguistic area over time. Previous periods of convergence and the spread of convergent innovations facilitated the development of LDO.

Chapter 7

The Development and Spread of Locative Determiner Omission as an Areal Feature

In the preceding chapters, I established the basis for analyzing languages of the Balkans as part of a linguistic area, and I examined the phenomenon of determiner omission in locative prepositional phrases synchronically and diachronically in Balkan languages belonging to Albanian, Romance, and Slavic branches. In chapter 1, I introduced locative determiner omission (LDO) as a grammatical phenomenon. In chapter 2, I provided a detailed analysis of LDO and its place in the Albanian grammatical system, which served as a basis of comparison for other languages. In chapter 3, I analyzed the same phenomenon of LDO in the Balkan Romance languages of Romanian and Aromanian, and in chapter 4, I confirmed that LDO exists in a less thoroughly grammaticized form in Balkan Slavic as well, although limited to dialects of Macedonian and not to be found in Bulgarian. In chapter 5, I examined the diachronic attestations of LDO in documentation from Balkan Romance, Albanian, and Balkan Slavic, comparing the points at which the phenomenon emerged in these Balkan language and tracking the grammatical predecessors to determiner omission. In chapter 6, I situated this data in the literature on the Balkan linguistic area and the processes of language contact that led to general patterns of grammatical resemblance between Albanian, Balkan Romance, and Balkan Slavic.

   With detailed data available from the relevant dialects and timepoints of these languages in contact, it is possible to bring the descriptive and theoretical data together to understand how LDO developed, how it came to be shared among the specific speech communities that show it, and why this feature was transmitted. The key aspects of LDO, analyzed in §2.3 are that it requires (1) a locative semantic trigger, (2) embodied by prepositions from a closed lexical set that normally selects the accusative case, (3) resulting in a morphological alternation in the marking of definiteness between determiner omission for bare nouns and a definite article for modified nouns, (4) which does not correspond to the marking of definiteness anywhere else in the grammatical system. In any one language, LDO is an unusual feature that would demand an explanation. But LDO is present in cluster of languages that already share a number of other similar grammatical properties that did not stem from a common ancestor. Therefore, viewing LDO as a feature that has resulted from processes of language contact, rather than an independent innovation in each language that shows it, better fits the facts of its distribution and the context in which it is shared. LDO satisfies the most stringent criteria that identify a shared feature of the Balkan linguistic area, or a 'Balkanism' as discussed in §6.3 (Joseph 1983:247): LDO is a similarity among Balkan languages that is an

innovation vis-à-vis their common ancestor language, resulting from language contact, which is unique to the Balkans.

The geographical distribution of LDO in the Balkan area suggests a specific path of transmission from an initial period of proto-Albanian and Latin contact into Balkan Romance, and thereafter, at a later point, from Aromanian into Albanian and Macedonian. That LDO is present throughout Balkan Romance, even in dialects such as Istro-Romanian no longer in direct contact with the core of the Balkan language contact zone (see §5.1), establishes a specific time point prior to the dispersal of Balkan Romance throughout the Balkans when LDO must have been present. In contrast, it has only recently become universal to the geographically contiguous zone of core Albanian dialects, while peripheral dialects of Albanian do not show LDO, as demonstrated in §5.2. Partially grammaticized LDO can only be found in colloquial Macedonian, not Bulgarian, and remains an unstable and developing part of the Macedonian grammatical system, as demonstrated in §5.3. The Macedonian speech community has existed and continues to exist in geographic proximity to the Albanian and Aromanian speech communities, reinforcing the prominence of this core zone for the understanding of LDO. But LDO can also be found in (Daco-)Romanian, well outside the core. If LDO is not to be treated as a coincidental common innovation of these particular languages—and the existence of other shared features among them strongly cautions against such an approach— then a diachronic analysis is necessary to account for the synchronic facts of distribution.

Given this unusual phenomenon and its distribution through several languages in the Balkans, I argue that the specific circumstances of language contact in this core zone of Balkan languages resulted in the spread of LDO from Balkan Romance to Albanian and Macedonian. LDO may have begun as an equivalent to the Albanian locative case in early Balkan Romance, resulting from a period of contact between proto-Albanian and Late Latin. But this initial period of contact did not result in the current state of LDO in Albanian. Rather, it was the more recent phenomenon stemming from multilingualism among Aromanian speakers and language shift of Aromanian speakers into Albanian and Macedonian that resulted in the spread of LDO in this cluster of languages. The grammatical system of Albanian, already showing proclivities to lose its morphological locative case, developed LDO after the model of the structure already present in Aromanian, with the spread of the feature matching patterns of Aromanian assimilation into Albanian speech communities. Aromanian multilingualism and language shift with Macedonian, too, led to the development of LDO in this form of Balkan Slavic, one of several influential features of Aromanian carried into Macedonian that distinguish Macedonian dialects from Bulgarian dialects. Aromanian was the key influence on both Albanian and Macedonian. The spread of LDO depends on its status as a discourse-bound grammatical variable that can be easily mapped between the morphemic structures of languages that already possess the same means of expressing definiteness through post-posed definite articles. Languages that have not engaged in recent extensive contact with Aromanian, such as Bulgarian, or whose noun phrase structure does not easily map with that of Aromanian, such as Greek, do not show LDO. Thus LDO is a grammatical borrowing, but it depended on the presence of other Balkan features such as the collapse of the grammatical marking of the location/goal semantic distinction, the reduction in nominal case, prepositional doubling, and the post-posed definite article to ease its spread. LDO, as an unusual feature, could only develop in languages that already existed in densely interconnected networks of multilingualism where a number of already established features allowed for extensive mappability. Analyzing LDO as an outcome of language contact elucidates the patterns of exchange and assimilation over the historical period of contact between Balkan languages, and also accounts for the unusual properties of this means of marking

definite reference in a discourse.

7.1   The spread of LDO from Balkan Romance

As analyzed in §5.2, Albanian and Balkan Romance have deep ties and may be two different outcomes of a pre-historic contact situation. Although it was once thought that Balkan Romance was the result of Late Latin imposed on 'Thracian' grammar, it is now thought that proto-Albanian and Late Latin were in a situation of extensive multilingualism very early in the period when Late Latin was developing into Balkan Romance. Albanian shows traces of deep, sustained Latin influence in both its grammar and lexicon, and innovations from Late Latin to modern Balkan Romance show convergence toward core grammatical properties of Albanian. The most persuasive proposal, following Hamp (1989), is that modern Albanian is the result of proto-Albanian speakers undergoing partial shift into Latin, while Balkan Romance is the outcome of a full shift of Albanian speakers into the Roman, Late Latin-speaking settlement community, resulting in a form of Balkan Latin that shared many grammatical and some lexical properties with Albanian.

I argue that what was transferred here in this early stage of contact between proto-Albanian and Late Latin is a distinct treatment of definiteness in locative prepositional phrases, rather than the exact means by which that distinct treatment was expressed. It is structural and not lexical similarities that define the shared feature of LDO in Albanian and Balkan Romance. The structure is not the outcome of a borrowed case system, or borrowed lexical items carrying their own rules for expressing definiteness within their domain, as demonstrated in §5.2. Several Albanian prepositions with no etymological relation to their Balkan Romance equivalents still trigger LDO. In addition, early Albanian did not exhibit LDO as is now observed in the modern core Balkan dialects. Rather, the semantic domains that triggered the appearance of the Albanian locative case in the earliest records (see §5.2) correspond to the environments that trigger LDO in the earliest Balkan Romance documentation. Goal and location are collapsed into a single semantic category, and accusative-selecting prepositions trigger a differential coding for definite reference when they express this collapsed locative category in both language groups. It is striking that LDO in Balkan Romance is confined to prepositions that normally select the accusative case, just as locative case only alternates with accusative case in Albanian. In Balkan Romance, the same set of prepositions that trigger LDO, when used to express non-locative semantics, fail to trigger LDO and govern the normal distribution of definite articles—this phenomenon does not have a parallel in Late Latin. Thus, comparing this with other examples of the influence of proto-Albanian on early Balkan Romance, such as the introduction of a post-posed definite article and particles of concord, a distinct treatment of locative semantic domains for accusative-selecting prepositions may be a feature borrowed from proto-Albanian into Balkan Romance. This locative semantic domain collapses the distinction between goal and location, an important general Balkanism and clearly one established very early among the Balkan languages. But the modification condition, in which the expression of definiteness alternates from a null article to a full accusative article depending on whether the head noun of the phrase is modified by any syntactic constituents, appears to have been a Balkan Romance innovation.

The means by which Balkan Romance arrived at LDO as the distinctive expression of definiteness in locative prepositional phrases is not available from the record. It must have happened after the introduction of consistent marking of definiteness with articles, a feature that was lacking in

Late Latin. But there are two possibilities for how Balkan Romance could have arrived from no morphemic marking of definiteness to a morphemic alternation between zero marking and a full article for nouns in locative prepositional phrases. One possibility is that Balkan Romance developed articles for nouns in all syntactic domains, like Albanian, but then lost them for the specific domain of locative prepositional phrases that select the accusative case. This would have been structurally parallel to the Albanian morphological locative case (which only alternates with accusative), but the morphemic outcome would have been different. An alternative possibility is that Balkan Romance gradually developed definite articles for all syntactic domains, beginning early with highly discourse-salient positions like sentential subject and object and extending these articles over time, but never fully extended the use of the article to unmodified nouns in prepositional phrases selecting the accusative. LDO in Balkan Romance would then represent a solitary hold-over from the original state of Late Latin. With the introduction of indefinite articles, the default state of the bare noun in prepositional phrases would then receive a definite interpretation, as is found in the first Balkan Romance attestations. In either case, Balkan Romance can be definitively designated as the first language to possess LDO. And it must have fully developed in Balkan Romance before dialect divergence between Daco-Romanian, Istro-Romanian, Aromanian, and Megleno-Romanian, since all of these forms of Balkan Romance show the same pattern of LDO. Having fixed the point at which Balkan Romance already developed LDO, we must look later, after the end of pre-historic Albanian/Late Latin contact and after the dispersal of Balkan Romance to various parts of the Balkan peninsula to account for LDO in modern Albanian and Macedonian.

Despite the possibility that Albanian LDO derived from the original Proto-Abanian-Latin contact situation, we have good evidence that Albanian LDO is the result of recent contact between Albanian and Balkan Romance, in particular Aromanian. While the locative semantic domain, a trigger for LDO in Balkan Romance, was also present as a salient grammatical category in Albanian, its relevance in the latter language was for the expression of locative case. Only with the later erosion and loss of the locative case in Albanian, in the modern period beginning from 19th century documentation into the 20th century, was there scope for the introduction of LDO through contact with Balkan Romance, specifically Aromanian. It is possible that Albanian could have simply lost the locative case of its own accord after the initial stage where Balkan Romance borrowed the locative triggering conditions for its own phenomenon of LDO. After all, Albanian has been undergoing a loss of case morphology since the time of its first documentation, as discussed in §5.2-§5.2. It has lost distinct markers for definite reference in the plural dative/genitive case (see §2.1) and it has collapsed the locative case into the accusative case. One could argue that the same process which led to the loss of the plural dative/genitive definite article in -*t* also affected the definite locative case article in -*t*. But this proposal of LDO as an independent, internally motivated innovation of Albanian would not account for the modification condition that makes modern Albanian LDO so similar to Balkan Romance LDO. Albanian did not just lose the locative case; it introduced the use of the accusative definite article only with modified nouns, while requiring unmodified nouns to undergo determiner omission to indicate definiteness, just as in Balkan Romance. Moreover, the prepositions that trigger LDO in modern Albanian differ in subtle ways from those that selected the locative case in early Albanian, as discussed in §5.2, and these differences suggest convergence toward Balkan Romance. While the prepositions 'with' and 'without' in early Albanian could select the locative case, in modern Albanian they fail to trigger LDO even though they are accusative-selecting prepositions, just as 'with' and 'without' fail to trigger LDO in Balkan Romance. Previously, locative case could occur with either modified or unmodified nouns. The

Table 7.1: Comparison of definite reference in locative prepositional phrases

Noun type	Early Albanian	Balkan Romance
Unmodified nouns, definite reference	Unmodified nouns appear with a distinct locative-case definite article	Unmodified nouns appear without a definite article, but are interpreted as definite
Modified nouns, definite reference	The same locative case definite article expresses definite reference, along with a locative particle of concord	Modified nouns take the accusative definite article

particularity of the prepositional triggers and the introduction of a modification condition for determiner omission strongly suggest Balkan Romance influence on Albanian and that LDO spread into Albanian from an Aromanian source, given that this was the Balkan Romance language that was spoken in intimate contact with Albanian.

The specific mechanism by which Albanian could have acquired LDO is important, however. After all, within the last three to four hundred years, Albanian speech communities have not been in direct contact with (Daco-)Romanian speech communities. Extensive Slavic speech communities intervene, and not all of these show LDO. Another form of Balkan Romance, however, has existed in intimate contact with Albanian in the modern period, and could have served as the vector for the introduction of LDO. The particular language involved here must have been Aromanian.

The differential treatment of definiteness marking in locative prepositional phrases overall is clear; however the details of dialect distribution of LDO require a more nuanced explanation. It cannot be traced directly to the period of intensive bilingualism between proto-Albanian and Late Latin. The transition of Albanian from the locative case to LDO must necessarily have included another period of contact between Balkan Romance and Albanian. Attributing the modern similarities between Albanian and Balkan Romance in the use of LDO to their early period of contact fails to account for important differences between them and complications in the timeline of development. While both Albanian and Balkan Romance share a tendency toward special marking of the objects of locative prepositions, there are numerous discrepancies in how they implement this special marking with the grammatical markers available to them. In particular, the earliest attestations of Albanian still feature a locative case used in prepositional phrases, and this distinct case form obeys different structural constraints from LDO. Table 7.1 provides a comparison of the state of Early Albanian and that of Aromanian (which has maintained roughly the same pattern of LDO throughout its attestation).

It is only at the end of the 19th century that LDO in the form observed in Balkan Romance begins to spread through Albanian documentation. By the mid-20th century, the standardized form of Albanian shares the same structure for marking definiteness in locative prepositional phrases as Aromanian. The situation in the dialects was more fluid. Thus, Lambertz (1948:18) noted a tendency for Geg dialects to continue to use a form matching the indefinite accusative with nouns modified by an attribute. This would conform with the pattern of eliding the locative ending *-t* in the presence of the corresponding particle of concord *të*. The definite locative case, with elision, resembles the indefinite accusative in the form of the noun and its particle of concord. Thus, while these Geg dialects would have shown the loss of the locative case, the pattern of LDO was not yet fully established. In contrast to Geg, Lambertz asserts that Tosk shows the definite accusative with modified nouns. This matches the pattern found in Aromanian, and Balkan Romance generally. It also accords with the patterns of contact between speaker communities; Tosk dialects, particularly those of the southern Albania, were in continuous contact with Aromanian extending into the 20th century, whereas there is no direct evidence of continued contact between Geg dialects and Balkan Romance after the 9th century, the period in which Balkan Romance speakers in the northern areas of Albania, Kosovo, southern Serbia, and Macedonia appear to have been absorbed into either Albanian- or Slavic-speaking populations. Moreover, the standardized form of the language drew its grammatical form primarily from Tosk dialects, which suggests that locative determiner omission for unmodified nouns was a southerly Tosk feature by the mid-20th century.

There are, however, complications to this analysis when examining documents from individual dialects. Some dialects that fall within the transitional belt between northern Tosk and southern Geg in central Albania are the only areas within the contiguous Albanian-speaking community of the Balkans that still show sporadic attestations of the locative case. Thus, the innovation of locative loss occurred uniformly at both peripheral ends of the Albanian speaking community before it spread to all dialects in the transitional zone between Geg and Tosk. LDO must be considered an innovation in Albanian, since the locative case is widely attested in a variety of dialects in earlier Albanian documentation. But the standard pattern of spread for a structural innovation is expected to begin from a center and spread outwards. In this case, the pattern of spread for the loss of the locative appears to be from two discontinuous loci of innovation.

The sporadic preservation of the locative in the center of Albanian speaking territory and its loss at the peripheries is curious. Further still, the belt where the locative is still sporadically preserved in the mid-20th century is close to areas where Weigand (1894/1895) found communities of Aromanian speakers in his survey of Aromanians in the Balkans during the late 19th century. These communities were the result of migrations from a more southerly core of Aromanian speakers in the Epirus region spanning southern Albania and northwestern Greece. Many of the Aromanian speakers found by the time of Weigand's survey had scattered from the destruction of the urban center of Moscopole in 1770.
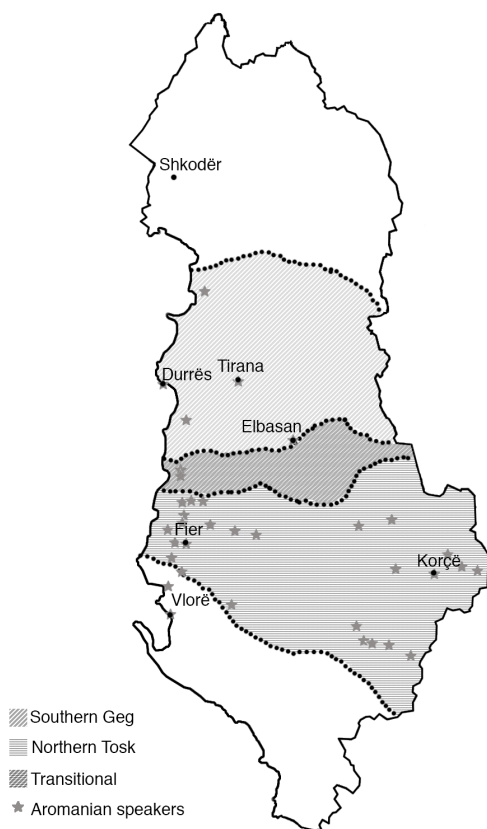
Figure 7.1: Map of northern Tosk and southern Geg dialects against Aromanian speaker communities in 1895

This suggests an intriguing pattern of transition from locative case to LDO in Albanian. The innovation of LDO appears to have begun at the southern range of the contiguous Albanian speaking region. Transition to LDO did not occur in non-contiguous Albanian dialects—Arvanitika and Arbëresh—which attests to this shift as an innovation. At the time that these two Tosk dialects separated from the main body of Albanian speakers, the locative was still present as part of the Albanian grammatical system. The form of LDO to which Albanian speakers transitioned following the departure of the diaspora dialects exactly parallels the form that has been in continuous use in Balkan Romance since earliest documentation. But this transition happened earliest and most uniformly among Tosk dialects where there were significant populations of Aromanian speakers, in the southern periphery of the Albanian -speaking zone. Aromanian-speaking villages were attested at the turn of the 19th century in the region south of where the old locative case was longest preserved. The preservation of a distinct, Aromanian speaker community resulting from recent migration, separate from the general Albanian community, lies just to the south of the area where some Albanian dialects preserve the old treatment of nouns in locative prepositional phrases. The shared structural features of LDO in Albanian and Balkan Romance have all the hallmarks of a contact-induced innovation in Albanian deriving from Balkan Romance. Thus presence of LDO appears to go hand-in-hand with presence of Aromanian communities, suggesting that assimilation of Aromanian speakers into Albanian is the source of this structure replacing the locative case. This argument requires further support, however, because the development of LDO in northern Geg dialects cannot be connected to a recent presence of significant numbers of Aromanian speakers or to the assimilation of Aromanian speakers into Albanian within the last century, as can be argued for southern Tosk. The loss of the locative in the northern Geg dialects must be motivated by reasons independent from its loss and development into LDO in the southern Tosk dialects. This argument can be further bolstered by addressing the data from Balkan Slavic, which supports the thesis that Aromanian is the source of LDO in Albanian and Balkan Slavic.

The geographic distribution of the Aromanian speech community provides good evidence for this path of transmission. Aromanian speakers were particularly in contact with southern dialects of Tosk, where LDO first appears consistently in Albanian, as shown in §5.2. Aromanian speakers were also in intimate contact with Macedonian speakers, particularly southwestern dialects of Macedonian, as shown in §4.1. These are the same Macedonian dialects where LDO is most entrenched, and also where prepositional doubling, a feature that associates with LDO in Macedonian and that is an especially strong trigger for LDO, has clearly been introduced from Aromanian. Aromanian would have already possessed a fully grammatically-integrated system of LDO from before the dispersal of Balkan Romance, and Aromanian speakers engaged in extensive multilingualism with their neighbors. The dialect distribution of LDO in Albanian and Macedonian, particularly the way in which it spread from an area centering around Lake Ohrid and the Epirus northward into either speech territory, points to Aromanian as the source of this phenomenon.

Aromanian speakers were once more widespread throughout the western Balkans, undergoing significant assimilation during the 18th through 20th centuries. I propose, as an explanation for this data, that assimilation of Aromanian speakers to the Albanian speaking community exerted influence on the grammatical structure of Albanian locative prepositional phrases, but only in the southern dialect range. In the area of Epirus and southern Albania, where Aromanian and Albanian speakers underwent significant communal contact, the Balkan Romance model for distinguishing nouns in locative prepositional phrases was carried into Albanian, triggering a shift from locative case to LDO. The scattering of Aromanian speakers from the communal center of Moscopole in

the late 18th century accelerated the trend toward dropping of the locative case in Albanian and the bilingualism of Aromanian speakers provided a model for a structural transition toward LDO. The area in which Aromanian speakers had not yet assimilated into the Albanian speaking community, marked by the ongoing presence of separate Aromanian villages, still contained Albanian dialects that sporadically preserved the old locative case. Delay in the assimilation of Aromanian speakers in the northerly reaches of Tosk dialects, or lack of an Aromanian speaking community entirely in the case of southerly Geg, resulted in the sporadic preservation of the locative among some Albanian dialects there into the 20th century. Otherwise, the convergence of southern Albanian dialects to a Balkan Romance model of LDO was complete by mid-20th century, aided by the imposition of southern Tosk norms as the standard language for the state of Albania following WWII. Albanian dialects that were not in contact with the main Albanian-speaking territory, Arbëresh (in Italy) and Arvanitika (in southern Greece) were not subject to Aromanian influence and correspondingly do not show LDO. Rather, they either show a preserved locative case, as in Arbëresh, or the elision of the locative-case definite article without a shift to the pattern of LDO, as in Arvanitika.

The loss of the locative in northern Geg Albanian dialects cannot be explained by a documented process of assimilation of Aromanian speakers into Albanian. There is no recorded presence of speakers of Aromanian or other Balkan Romance languages in the area of northern Geg dialects after the early Middle Ages. Extended contact with Balkan Romance did occur in this area, and must have been extensive and intimate up until at least the 12th century, but we do not have information on any further period of assimilation like that which was key to the development of LDO in southern Albanian dialects. Attention to the local details of speaker contact, dialect geography, and chronology are important to this account of the development of LDO. It may be that ongoing contact between pastoral Aromanians and Albanians of the mountainous region led to the elimination of the locative case and the introduction of LDO earlier in the northern Geg dialects than in the southern Geg dialects spoken in the plains areas of Albania, for which attestation still shows a locative case up to the late 19th century. It is only possible to determine from the documentation that the shift to LDO was already complete by the mid-20th century across the majority of Albanian dialects, with sporadic preserved locative case usage in central Albanian. The standardization of Albanian primarily on the basis of southern Tosk dialects may have played a role in the cementing of LDO across the whole range of contiguously spoken Albanian dialects. That the standard Albanian speech form for writing, public presentation, education, and inter-dialectal communication was drawn particularly from Tosk dialects would have supported the further transmission of LDO.

In sum, while the differential treatment of definiteness in locative prepositional phrases is a shared result of language contact between Balkan Romance and Albanian, the particular instantiation of this pattern as LDO does not find a uniform explanation across Albanian dialects. For southern Albanian dialects, the influence of Aromanian is clear. But lack of documentation for key periods of the development of Albanian hamper our ability to establish precisely how LDO became introduced as a structural alternative to the locative case in the far northern area of Albanian speakers. That Balkan Romance and Albanian underwent an intense period of contact is well understood and was important for laying the basis of the locative case in Albanian and LDO in Balkan Romance. The influence of Aromanian on southern Albanian dialects through a process of multilingualism and eventual assimilation matches the diachronic shift from locative case to LDO across Tosk dialects. That Aromanian was instrumental to this shift is further confirmed by the absence of LDO in the Tosk-derived dialects of Arvanitika and Arbëresh (see §5.2), which did not undergo the same extent of contact with Aromanian and were not subject to assimilation of Aro-

manian speakers. The specific period of speaker contact, between Tosk speakers and Aromanian speakers in the 19th and 20th centuries does not extend to northern Geg dialects. Since a language contact explanation is not sufficient to account for the structure across all of the contiguous Balkan dialects of Albanian (excluding Arvanitika and Arbëresh), I base my argument on further evidence that Aromanian assimilation brought with it the introduction of the LDO structure into Balkan Slavic. The distribution of LDO in Balkan Slavic, specifically in Macedonian, reinforces the role of Aromanian as the source of LDO for Tosk Albanian dialects, if not for Geg Albanian dialects. In both Albanian and Balkan Slavic, where Aromanian multilingualism and assimilation of Aromanian speakers into the broader speaking population was extensive, LDO became a part of the grammatical system of local Albanian or Balkan Slavic speech forms. From there, it was possible for LDO to spread to the broader body of Albanian and Macedonian speakers, in the form of an internally motivated change, a process which has also been noted for other borrowings from Aromanian such as the *have*-perfect (Alexander 1984-1985:41).

7.2 LDO through language shift

That shared features of the Balkan linguistic area spread through mutual multilingualism has been argued by many authors (cf. Lindstedt 2000, Friedman 2006, Topolińska 2007, Joseph 2010, *inter alia*). But in the case of LDO, the dynamics of contact between the source speaker community and the communities of Albanian and Macedonian speakers suggest that this feature came about through language shift of Aromanian speakers into these other languages. Not all language features in a linguistic area undergo convergence, and not all languages engage in the same kind of change. Languages that serve an emblematic purpose for their speaker communities may be protected from change and exhibit more conservativisms. And certain features, such as lexicon, are readily accessible to the conscious awareness of speakers, which makes them resistant to change even as other features that are less consciously emblematic undergo shift. Without particular configurations of social relations between speaker communities, in which languages have roughly equivalent social status, convergence may not occur at all. It is possible that the documented multilingualism between Macedonian, Albanian, and Aromanian communities could have yielded LDO. Longterm multilingualism was certainly necessary to provide the structural basis of LDO in the form of a locative semantic domain, synthetic case loss, and the post-positive definite article. But the particular pattern of development for LDO in Albanian suggests that it did not become common until widespread shift of Aromanian speakers into Albanian was underway. The ease with which Aromanians blended into their communities, often becoming 'the best Macedonians,' 'the best Albanians,' or 'the best Greeks' aided in this process (Schwandner-Sievers 2002:150). The recent innovation of LDO in Macedonian, likewise, seems to have been a result of language shift.

The social circumstances of contact between Macedonian and Aromanian would have enabled the adoption of locative determiner omission into Macedonian grammar relatively recently, within the same time period that determiner omission was consolidated in Albanian. A major migration of Aromanian speakers to southwestern Macedonia occurred following the collapse of the Moscopole center in the late 18th century, and migration continued into the 20th century (Markoviḱ 2007:51). These Aromanian speakers brought their particular dialect with them and established communities in urban centers such as Ohrid and Bitola, where they had significant influence on the local dialects of Macedonian (Koneski 1967:148). They also founded and expanded their own villages, such

as Kumanovo, and integrated into the multi-ethnic, multi-confessional, and multi-lingual communities that already characterized the Macedonian region in the Ottoman Empire. Multilingualism between Aromanian, Macedonian, Greek, Albanian, and Turkish was common during this period, as testified both by native residents to the region and by foreign observers. Not all languages participated in the dynamic of multilingualism. Social prestige of the native speaker community had an important effect on patterns of code-switching, mutual multilingualism, and accommodation. For example, while many non-Muslims and Muslims of non-Turkish linguistic background spoke Turkish as a second language, the prestige of Turkish and its significant structural divergence from the Indo-European languages of the Balkans mean that the structural influence of Balkan languages on Balkan Turkish was minimal. Greek, as well, possessed a level of social prestige greater than that of other Balkan languages. It was the language of Christian Ottoman administration, and a major language of literacy. Aromanian, Macedonian, and Albanian, at the local level of the southwest Balkans, were languages of middle-level social prestige and were employed in the most systematic mutual multilingualism (see §6.4).

In this milieu, the similarity in treatment of definiteness in prepositional phrases must be taken as part of a set of convergent grammatical features between Albanian, Aromanian, and Macedonian. The early attestation of LDO in Romanian and its distribution among Balkan Romance languages with an even earlier date of divergence establishes this feature as originating in Balkan Romance. The later, uneven attestation of LDO in Albanian can be accounted for by considering mutual multilingualism and Aromanian assimilation into the Albanian-speaking population as the source of this feature in Albanian. LDO replaced the Albanian locative case, which already offered a structural parallel with Balkan Romance LDO in that locative prepositional phrases required a specific treatment of definiteness separate from any other syntactic context. These parallels in the distinctive treatment of locative prepositional phrases between Albanian and Balkan Romance may be the result of longterm, earlier contact and exchange before the period of attestation. The specific adoption of LDO into Albanian, however, is the result of relatively recent contact. The spread of this feature in Albanian accords with the gradual assimilation of Aromanian speakers into the Albanian speaking population. This explanation finds further support in the distinctive treatment of definiteness in locative prepositional phrases in Macedonian. Macedonian shows a tendency to omit definite articles in locative prepositional phrases that distinguishes it from Bulgarian. Macedonian also shows numerous features that resulted from recent contact with Aromanian speakers through mutual multilingualism and assimilation of Aromanian speakers into the Macedonian speaking population. Parallels between Macedonian and Albanian, then, are the result of the shared influence of Aromanian on the grammatical structure of both languages. LDO can be counted as a Balkanism, but it is a local and specific outcome of the interaction between Balkan Romance on the one hand and Macedonian and Albanian on the other. LDO is neither an accidental parallelism nor is it an independent innovation in either Macedonian or Albanian. Summing together the data, diachronically and across the geographic span of the Balkan languages, it must be attributed to a specific period of contact with Balkan Romance.

LDO does not appear to have been an emblematic Aromanian feature, however. This would have made it easier to spread as language shift happened. Macedonian, for its part, had already undergone collapse of nominal case and the where/whither merge by the time when Aromanian influence would have applied. There was no case form to replace, the Slavic locative case having already been lost relatively early. Albanian poses a more complicated situation, since it did have a locative case, which appears to have served an emblematic function within Albanian. Use of

the locative in the late 19th century documents of the Rilindas appears to have been a conscious archaism. But omission of the locative case was already attested sporadically throughout Albanian, and appears in dialects such as Arvanitika that were not affected by extensive contact with Aromanian. There was great scope for variation in the realization of Albanian nouns in locative prepositional phrases, including omission of the locative case, use of the definite accusative, or combinations of the two, whether nouns were modified or not. Thus the particular configuration of LDO could have been normalized in Albanian within the scope of already existing variants. Dialects that eliminate the locative in favor of regularized LDO, in southern Tosk, had ongoing contact with Aromanian. Aromanians were often prosperous and educated members of the communities of southern Albanian cities, extending from the 18th century through to the early 20th century. When Aromanian speakers shifted into Albanian, it was these southern Tosk dialects with which they engaged. The speech habits of Aromanians, carrying LDO into Tosk Albanian, could have served as a model for the regularization of LDO in Albanian.

Standardization of Albanian would have played a further role in the spread of LDO throughout the language. The standardization of Albanian was already beginning during the late 19th century, but with a pluricentric model based on multiple dialects and literary centers. It was only after the Second World War, with the take-over and centralization of power under the Party of Labor of Albania, that a single literary model for education and public speech performance was selected. This was primarily based on the speech forms of Tosk, with much of the leadership of the PLA coming from southern Albania. Variability in the realization of the locative case, or its omission, was washed out in favor of consistent LDO.

Areas in Albanian where the locative case was sporadically preserved relatively late overlap with areas where distinct Aromanian speech communities appear to still be preserved. This provides evidence that the development of LDO was not a matter of convergence toward a grammatical pattern due to multilingualism alone. Aromanian communities certainly would have been multilingual in Albanian, even if they preserved their distinct speech forms. Instead, only when shift of Aromanians into the Albanian speech community was complete do we see the decisive loss of the locative case and its replacement with LDO. Language shift as the origin of LDO and its spread across Albanian through this form of grammatical change under conditions of language contact accounts well for this pattern of spread.

The loss of the locative and the introduction of LDO into northern Geg dialects of Albanian is more difficult to explain as a result of language contact. There has not been any recent community of Aromanian speakers in the northern, Geg-speaking areas of Albania. Nonetheless, by the mid-20th century, the locative appears to have been wholly eliminated from Geg dialects. The most feasible explanation for this phenomenon is that inter-Albanian contact led to the spread of LDO, starting from southern Tosk and then spreading throughout as the preferred variant to replace the loss of the locative case, which had already been sporadic in Geg.

Although the path of the loss of the locative case in Geg and its replacement with LDO is less clearly linked to Aromanian, evidence from Balkan Slavic provides further evidence for Aromanian language shift as the source of the development of LDO as a Balkan language feature. Macedonian, as well, shows patterns of LDO to be greatest in communities where Aromanians had been a key part of the speech community and then, recently, underwent shift into Macedonian. These are the same communities where other Macedonian features such as prepositional doubling and the have-perfect are most thoroughly grammatical integrated. The Aromanian language community in Macedonia has mostly undergone shift, with practically all remaining Aromanian speakers now

multilingual with Macedonian, while Macedonian native speakers rarely speak Aromanian. LDO now appears to be spreading throughout Macedonian, like other Aromanian-origin features such as the have-perfect, even to Macedonian communities where Aromanian speakers were not a significant presence. The attestation of LDO in Macedonian, where Aromanian language multilingualism and language shift was a factor, compared with Bulgarian, where Aromanian did not play an important role in language contact, reinforces this point. Bulgarian lacks LDO in constructions where it is common in Macedonian. Considering the diachronic evidence and placing it in context with developments in Albanian, Aromanian language shift is the clearest source for the development of this feature in those forms of Albanian and Balkan Slavic that have it.

7.3   LDO through structural borrowing

The development of LDO required structural changes in which existing morphology in Macedonian and Albanian was redistributed to accommodate a semantically restricted special domain, shared among all Balkan languages belonging to the Slavic, Romance, and Albanian groups. This new configuration of definiteness marking in Macedonian and Albanian, applied to a pre-existing postposed definite article, brought the grammatical systems of these languages into line with Aromanian. However, the structural changes necessary to yield LDO did not require the development or grammaticalization of new morphology. In Albanian, both case selection and prepositional semantics played a role: it was the accusative definite article, in particular, that developed a new distribution dependent on the modification status of the noun, determined by locative semantics in the prepositional phrase. In Macedonian, which lacked morphological nominal case at the time of contact with Aromanian, the locative semantics were sufficient to determine the new distribution of the definite article. The shared feature of LDO represents a convergence toward surface similarity in the distribution of grammatical elements, rather than convergence in underlying syntax. Transferring the practice of treating locative prepositional phrases in a particular way for the purposes of definiteness, and predicating this treatment on the modification status of the noun phrase, does not require the transfer of morphology or even lexical borrowing. It is a structural convergence outside the lexicon.

The interaction between LDO and the locative case collapse in Albanian is particularly intricate. In Albanian, we see both the loss of a morphological irregularity (in that the locative case appeared sporadically but with restricted domain) and its replacement with regularity (in that LDO applies consistently across the domain that once conditioned the locative). There is thus some regularization involved in the transfer of LDO into Albanian and its spread throughout Albanian dialects, but this depends upon previously existing complexity to which the LDO feature could map.

The semantic restriction on LDO—the locative semantic domain resulting from the collapse of the location/goal distinction inherited from Indo-European—is motivated by the appearance of the locative case with a closed set of prepositions that otherwise select the accusative. This can be contrasted with another change that looks similar, but occurred as an independent, Albanian-internal innovation: loss of the dative/genitive plural definite article. The development of LDO involves the same phonological process in the elision of final -t, particularly before the particle of concord in *të*. But there is no parallel for the loss of the dative/genitive plural definite article in other languages of the Balkans. In one case, the tendency toward elision of final -t resulted in a restricted elimination of the distinction between definite and indefinite. But in the case of LDO, the

tendency resulted in the introduction of a new pattern of marking definiteness, alternating between a zero marking and a full accusative case, which had not been previously been a regular pattern in the language. This new regular pattern matches the pattern of Aromanian and Balkan Romance in general. The Albanian-internal tendencies that resulted in dative/genitive article loss helped with the mapping of a new grammatical pattern to existing Albanian structure, which was transferred during Aromanian language shift into Albanian. Regularization, rather than simplification, is the key to LDO in Albanian.

Regularization does not account for the Macedonian pattern of LDO, however. LDO in Macedonian is highly variable, dependent on the register and speaker affinity for dialectal forms that are closer to Aromanian structure (western dialects) or more distant (eastern dialects, which are closer to the LDO-lacking Bulgarian). In Macedonian, LDO is an existing variant option, rather than a fully integrated part of the grammar. This may, too, have been the state of LDO in Albanian in the late 19th and early 20th centuries, before standardization spread the grammatical pattern of LDO throughout Albanian dialects. Macedonian grammatical marking of definiteness, in contrast, is made somewhat more complicated by the option for LDO applying to all lexical items. The environmental conditions of language contact have allowed for the development of a more complex structure in Macedonian, in comparison with a reduction in complexity in Albanian. Both of these directions of change match with a state of affairs already present in Aromanian resulting from a long-term grammatical integration of the LDO feature.

In all languages that possess LDO, a collapse of location/goal distinction into a single semantic domain, governed by the same set of prepositions, was important for facilitating the mapping of LDO. Without the locative semantic condition, it would not be possible to reproduce the structure of LDO even for those languages which express definiteness with articles. Collapse between where and whither is a very early phenomenon in Albanian and Balkan Romance, present from the beginning of the records and likely present well into the pre-history of these languages. The collapse of location/goal in Macedonian was a more recent phenomenon, traceable to the 1600s as morphological case collapse in general was proceeding. This collapse had to precede the introduction of LDO. There is also a locative/non-locative distinction between uses of prepositions that denote spatial locations and uses that denote more abstract relations. From its first documentation, LDO has consistently been used to distinguish between these semantic conditions. The potential to map LDO to a distinct semantic domain, indexed by a closed set of prepositions, allowed for the spread of LDO between these languages. This was particularly easy in Albanian and Balkan Romance, with their parallel nominal syntax. Even when the noun phrase structure of Macedonian showed fundamental distinctions in its structure (such as lack of a particle of concord and head-final orientation), LDO benefited from mappability based on the factors of the post-posed definite article and locative semantics.

In schematization, each of the Balkan languages can be represented by a map of morphemes to semantic categories. Aromanian provides the basis for the development of LDO in Albanian and Macedonian, and is represented by the schema in Figure 7.2. When a noun (N) is modified by another constituent such as an adjective (A), a preposition (Prep) that selects accusative case (acc) requires a definite article (Def) post-posed to the noun to select definite reference. The definite article agrees with the selection restriction of the preposition for case, and agreement for other features such as number and gender occurs between the noun and the modifier, to the extent that the modifier declines for these features. The semantics of definite reference is expressed by the overt morpheme of the definite article. The only relevant factor is the case-selection of the preposition;

whether the preposition expresses locative or non-locative semantics is not relevant in the case of modified nouns.

In contrast, with unmodified nouns, there is a split between accusative-selecting prepositions that express non-locative semantics and those that express locative semantics. Those that express non-locative semantics govern an unmodified noun in which definite reference is expressed by a post-posed definite article, as with modified nouns. But prepositions that express locative semantics select a bare noun, with no definite article. Rather, the fact that the noun is unmodified and that the preposition expresses locative semantics imposes a definite interpretation on the bare noun, which is not otherwise possible in the Aromanian system of nominal grammar.
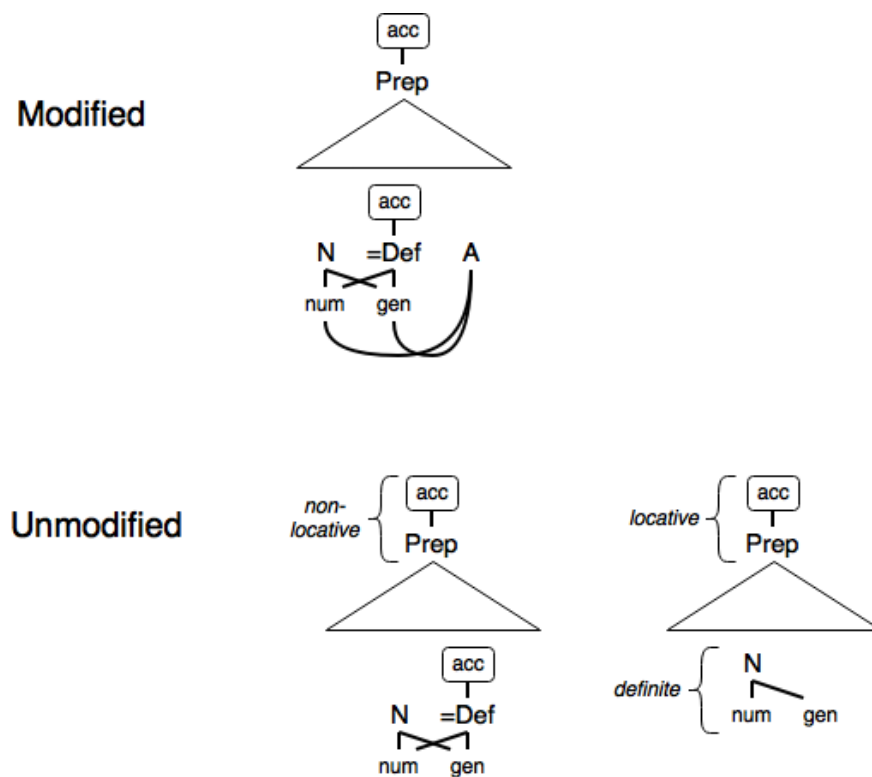
Figure 7.2: Schema of Aromanian locative determiner omission

This Aromanian schema can be compared against Macedonian and Albanian to show how LDO maps among grammatical categories in these languages, as shown in Figure 7.3. For modified nouns, Albanian and Aromanian show an explicit mapping between accusative-case selection and the use of a definite article with modified nouns. The presence of a modifying constituent in either language licenses the presence of the definite article, and accusative case is expressed on that definite article. Likewise, with unmodified nouns, non-locative semantics in prepositions that select the accusative case results in a definite article to express definite reference. The non-locative semantics as well as the presence of the accusative-case definite article maps between Aromanian and Albanian for unmodified nouns. LDO itself maps among all three of Macedonian, Albanian, and Aromanian, where the locative semantics of the preposition result in definite reference attributed to a bare noun. In this way, the coordination between morphology and semantics can be transparently encoded across all three languages through the shared structure of LDO.
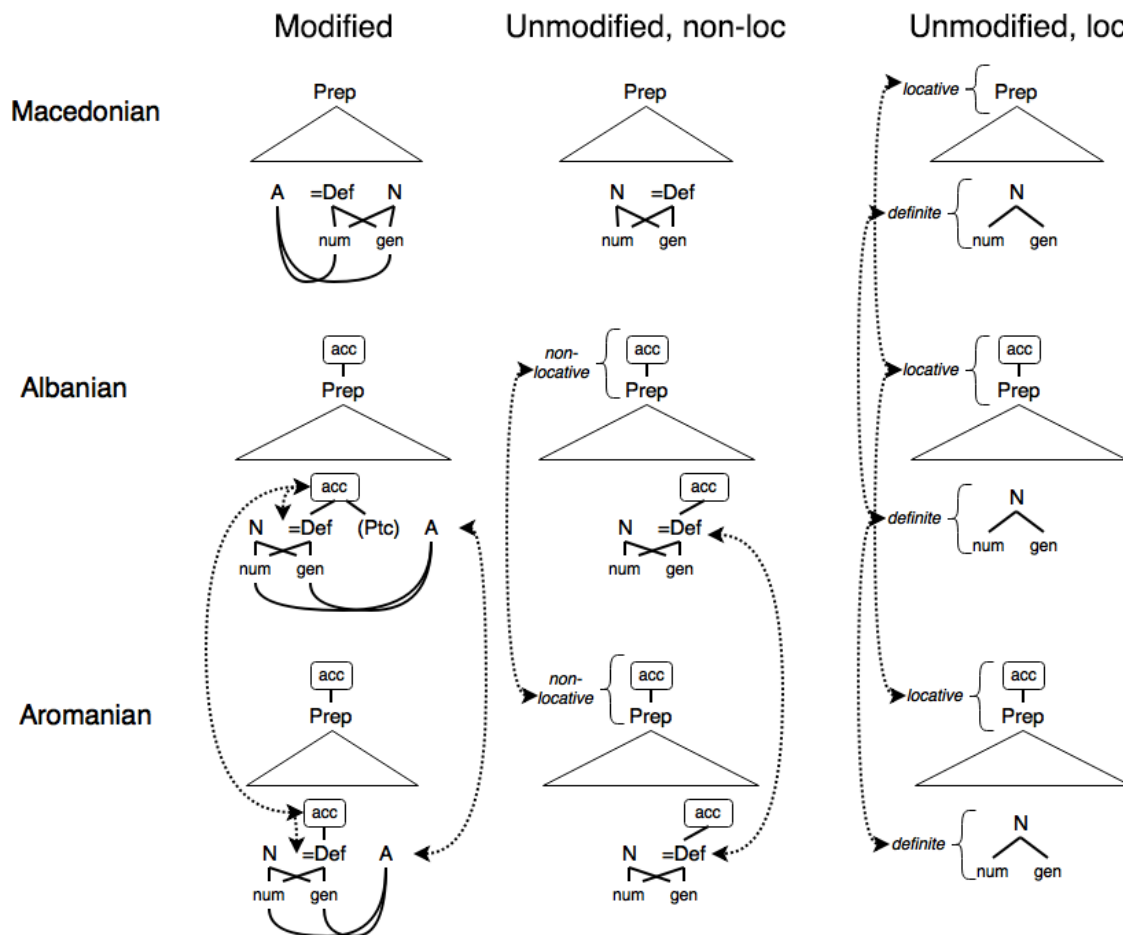
Figure 7.3: Schema of Macedonian, Albanian, and Aromanian locative determiner omission compared

The structure, function, stability, and mappability of LDO all played a role in allowing this unusual feature to spread. Its grammatical integration depended on the presence of other conditioning features that had already developed through convergence due to mutually multilingual language contact. Growing similarities between these languages eased the mapping of this particular feature in the process of language shift as Aromanian speakers increasingly engaged with Albanian and Macedonian. What resulted was a synchronic picture of convergence, although feature transfer predicated on pre-existing affinities was the process that yielded this outcome. As a result, Macedonian and Albanian were able to borrow grammatical structure from Aromanian and integrate it into their own grammatical systems.

Chapter 8

Conclusion

This dissertation has analyzed a grammatical feature shared among the Balkan languages Macedonian, Albanian, Romanian, and Aromanian. This case study of determiner omission in locative prepositional phrases illustrates how 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. Without previous periods of extensive multilingualism, the structural basis for locative determiner omission would not have been available in Albanian, Balkan Romance, or Balkan Slavic. But without specific periods of assimilation and accommodation between Aromanian speakers and local forms of Albanian and Macedonian, locative determiner omission would not have appeared in these languages. This grammatical phenomenon must be understood as the outcome of contact between local speech forms, resulting in the grammaticalization of a pragmatically salient variable of definiteness that could be easily mapped across already similar noun phrase structure. The presence of locative determiner omission identifies a particular clusters of contact within the larger, layered clusters of contact that make up the Balkan linguistic area as a whole. 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. Locative determiner omission as an areal feature depends on the numerous previous periods of contact and convergence that resulted in shared Balkan grammatical features.

Clustering, iterated patterns of contact between local speech varieties is of particular importance in the account proposed here for the development of LDO. As seen in Figure 8.1, multiple stages of contact preceded the development of LDO in all the languages that now show it.

The first step of language contact leading to LDO occurred between Proto-Albanian and (Balkan) Late Latin, represented by the first level of Figure 8.1. At this time, Slavic languages were not yet spoken in the Balkans. In contrast, Latin speakers arriving with Roman settlers and Proto-Albanian speakers indigenous to the Balkans learned one another's languages in a zone of contact north of the Jireček line through northern Albania, Kosovo, northern Macedonia, and southern Serbia (Rusakov 2013). Language contact is evidenced by the numerous borrowings of Latin vocabulary into Albanian at an early time depth, and the presence in Balkan Romance of Albanian borrowings that do not appear in any other Romance language. This contact resulted in convergence toward important elements of shared structure. While we do not have direct evidence for the structure of Albanian prior to the 16th century, we know the changes that Latin underwent: the collapse of the distinction between goal and location, the introduction of particles of concord between adjectival or genitive modifiers and the head noun, and a post-posed definite article. Balkan Romance appears to have
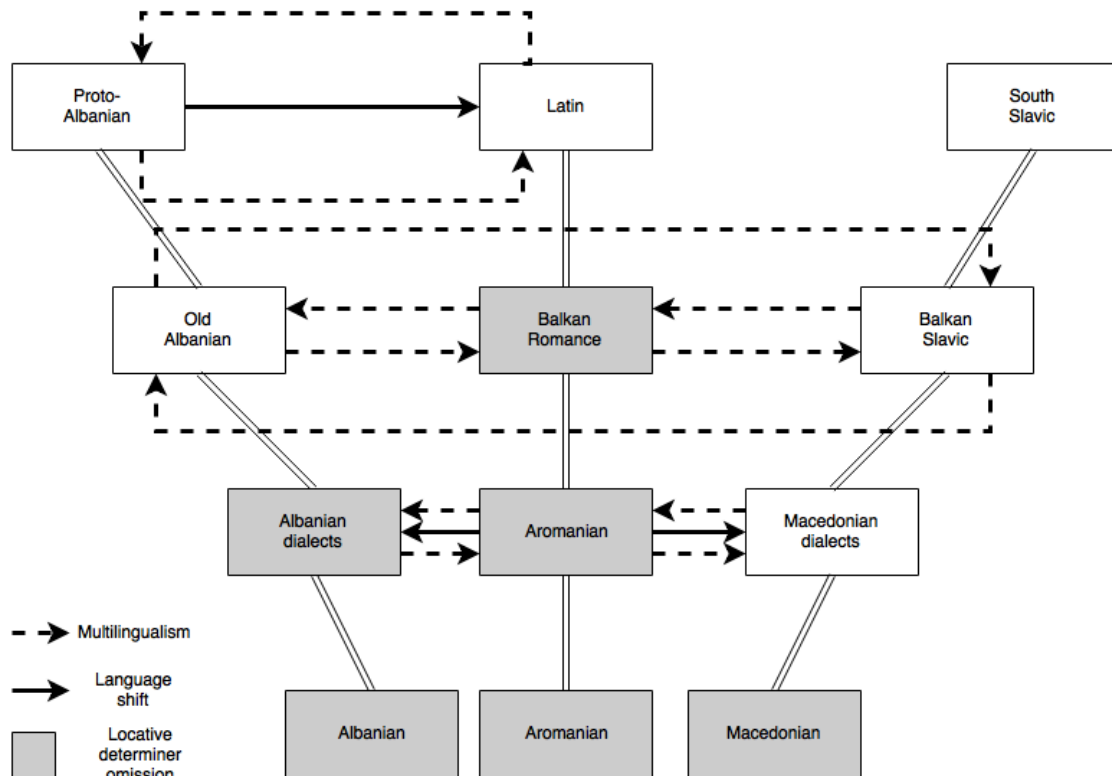
Figure 8.1: Diagram of the path of transmission for locative determiner omission

been the outcome of Albanian speakers shifting into Latin (Hamp 1989). The changes brought about by multilingual contact and shift laid the basis for the development of LDO. Without the development of the post-posed definite article, it is not possible to show a distinct means of expressing definite reference through the omission of an article. Convergence toward this syntactic and morphological means of expressing definiteness is a necessary basis for the development of LDO. A collapse in the distinction between goal and location also provides the semantic domain for triggering LDO—without this coherent semantic category, it would not be possible 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). This was the classic pattern of Balkan contact. This pattern extended throughout the Balkans where these three speaker populations were present. As languages at the middle of level of prestige, they were used widely by both native and non-native speakers. Similarities in grammar that already united Albanian and Balkan Romance affected Balkan Slavic as well, which diverged from non-Balkan South Slavic in precisely the features that are precursors to LDO. 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. While the classic pattern of Balkan contact brought Slavic languages into the fold of languages with the grammatical preconditions for the development of determiner omission, LDO did not develop 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 century, speaker dynamics changed as well. This is represented at the third level of Figure 8.1. 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. Aromanian, which was already in possession of LDO, provided a new grammatical model for expressing definiteness in locative prepositional phrases at precisely the time that these Albanian dialects were losing the locative case. Thus the early contact between proto-Albanian and Late Latin was a necessary step in the convergence of Albanian and Balkan Romance, but so was a much later period contact between particular, local dialects of Albanian and Aromanian. 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 where necessary for the development of LDO. An initial period of contact between proto-Albanian and Late Latin led to the development of core Balkan grammatical features. These were later reinforced during an extensive period of contact marked by mutual multilingualism among Albanian, Balkan Romance, and Balkan Slavic, with Slavic in particular converging toward the grammatical structures already present in Albanian and Balkan Romance. A further period of intensive, more localized contact between Aromanian and Albanian dialects, and Aromanian and Macedonian dialects, respectively, led to the spread of LDO from Balkan Romance into Albanian and Macedonian. Grammatical features already shared during the period of mutually multilingual Balkan language contact eased the way for the transfer of LDO during Aromanian language shift. The outcome, depicted in the last step of Figure 8.1, is the presence of LDO in Albanian, Aromanian, and Macedonian.

The kind of structural change involved here is important to the analysis of LDO. Whether LDO involves grammatical simplification is relevant to broader questions about what can be transferred through language contact and language shift. In the Balkans, the language contact that led to convergence among languages such as Aromanian, Macedonian, and Albanian has sometimes been characterized as creolization (Topolińska 2007), with morphosyntactic leveling and simplification of grammatical irregularities. It is not wholly accurate to say that LDO is morphosyntactic leveling. While Albanian did lose the morphological locative case, it did not lose the distinctive treatment of the locative semantic environment marked by this case. The pattern of Macedonian indicates greater complexity, in that the semantic interpretation of bare nouns is complicated by the conditioning factor of the preposition that governs them. Tracing back to the source language, Balkan Romance, it is unclear whether LDO was a simplification. It does not correspond to a particular state of Late Latin, which did not mark definiteness. Rather, it may have been the introduction of the Albanian pattern of distinct markers for the locative case into Balkan Romance. That would have been a transfer of greater grammatical distinction, to satisfy a requirement for the special marking of the locative.

This thesis has dealt primarily with data from Albanian dialects, Macedonian dialects, Aromanian, and Romanian. Further research is needed to identify whether other Balkan speech forms not discussed in this thesis may possess locative determiner omission. For example, research into

determiner omission in specific northern Bulgarian dialects might reveal whether Romanian could have had the same effect on local Bulgarian speech forms, or whether the social situation in this area was sufficiently different to mitigate against the development of determiner omission on a Balkan Romance model. Likewise, the Torlak dialects of Serbian are subject to extensive grammatical convergences toward the Balkan type and are spoken in an area where there was once extensive contact between Albanian and Balkan Romance. Tracing the extent or the existence of determiner omission in Torlak may shed light on the path of development for this feature in northern Albanian Geg dialects, for which more research is needed. As demonstrated in this dissertation, a historically informed focus on the spread of a grammatical feature through local speech forms provides new and productive insights into the constitution of a linguistic area. In the details of change, contextualized to individual languages, one can identify the diachronic and geographical layers of contact that yield remarkable outcomes of convergence among languages across familial lines.

References

Aikhenvald, Alexandra, and R. M. W. Dixon. 2006. Grammars in contact: A cross-linguistic perspective. In *Grammars in contact: A cross-linguistic typology*, 1–66. Oxford: Oxford University Press.

Ajeti, Idriz. 1985. *Studime gjuhësore në fushë të shqipes*. Prishtina: Rilindja.

Alexander, Ronelle. 1983. On the definition of Sprachbund boundaries: The place of Balkan Slavic. In *Ziele und Wege der Balkanlinguistik: Beiträge zur Tagung vom 2.-6. März 1981 in Berlin*, 13–26. Wiesbaden: Harrassowitz.

Alexander, Ronelle. 1984-1985. Areal linguistics and South Slavic linguistics: The case of Balkan Slavic. *Zbornik Matice srpske za filologiju i lingvistiku* 27-28:39–47.

Alexander, Ronelle. 2000a. *In honor of diversity: The linguistic resources of the Balkans*. Columbus, OH: Department of Slavic and East European Languages and Literatures, The Ohio State University.

Alexander, Ronelle. 2000b. Tracking Sprachbund boundaries: Word order in the Balkans. *Studies in Slavic and General Linguistics* 28:9–27.

Alibali, Agron. 2013. Arvanites. In *Encyclopedia of the world's minorities*, ed. Carl Skutsch, 138–139. Oxford: Routledge.

Altimari, Francesco. 1994. La distribuzione dell passato "prezuntivo" nell' albanese d'Italia. In *I dialetti italo-albanesi*, ed. Francesco Altimari and Leonardo Savoia, 211–221. Rome: Bulzoni.

Aronson, Howard. 2007. *The Balkan linguistic league, "orientalism," and linguistic typology*. Ann Arbor, MI: Beech Stave Press.

Asenova, Petja. 2002. *Balkansko ezikoznanie: Osnovni problemi na balkanskija ezikov sŭjuz*. Veliko Tărnovo: Faber.

van der Auwera, Johan. 1998. Revisiting the Balkan and Meso-American linguistic areas. *Language Science* 20:259–270.

Avram, Mioara. 1962. Scrisoarea lui neacșu. In *Crestomație romanică*, volume 1, 149–151. Bucharest: Editura Republicii Socialiste România.

Balamaci, Nicholas. 1991. Can the Vlachs write their own history? *Journal of the Hellenic Diaspora* 17:9–36.

Ballamaci, Iancu. 2010. *Metoda aromână/vlahă*. Bucharest: Predania.

Banfi, Emanuele. 1985. *Linguistica balcanica*. Bologna: Zanichelli.

Bauer, Brigitte. 2007. The definite article in Indo-European. In *Nominal determination: Typology, context constraints, and historical emergence*, ed. Elisabeth Stark, Elisabeth Leiss, and Werner Abraham, 103–140. Amsterdam: John Benjamins.

Benveniste, Émile. 1978. *Problèmes de linguistique générale II*. Paris: Gallimard.

Bien, Peter, John Rassias, and Chrusanthi Yiannakou-Bien. 1983. *Demotic Greek I*. Hanover: University Press of New England.

Boretzky, Norbert. 1968. Der bestimmte Artikel nach Präpositionen im Albanischen. *Zeitschrift für vergleichende Sprachforschung* 82:129–152.

Borgert, U. H. G., and C. A. Nyhan. 1976. *A German reference grammar*. Sydney: Sydney University Press.

Brailsford, Henry. 1904. *Macedonia: Its races and their future*. London: Methuen & Co.

Browne, Wayles. 1993. Serbo-Croat. In *The Slavonic languages*, ed. Comrie Bernard and Greville Corbett, 306–387. New York: Routledge.

Byron, Janet. 1973. Selection and evaluation of alternates in Albanian language standardization. Unpublished PhD dissertation, University of Pennsylvania, Philadelphia.

Byron, Janet. 1976. *Selection among alternates in language standardization: The case of Albanian*. The Hague: Mouton de Gruyter.

Cacaj-Shushka, Gjystina. 1999. *Gjuha në veprën e Naum Veqilharxhit*. Prishtina: Shtëpia Botuese Libri Shkollor.

Caciuperi, Apostol. 1985. *Gramatică armânească*. Riverdale, NY: Vlach Institute.

Campbell, Lyle. 1985. Areal linguistics and its implications for historical linguistics. In *Papers from the 6th international conference on historical linguistics*, ed. Jacek Fisiak. Amsterdam: John Benjamins.

Campbell, Lyle. 2006. Areal linguistics: A closer scrutiny. In *Linguistic areas: Convergence in historical and typological perspective*, 1–31. New York: Palgrave Macmillan.

Campbell, Lyle, Terrence Kaufman, and Thomas Smith-Stark. 1986. Meso-America as a linguistic area. *Language* 62:530–570.

Capidan, Theodor. 1925. *Meglenoromânii: Istoria şi graiul lor*. Bucharest: Monitorul Oficial.

Capidan, Theodor. 1937. *Les Macédoroumains*. Bucharest: Academia Română.

Caragiu-Matioţeanu, Matilda. 1968. *Dialectul aromân*. Bucharest: Academia română.

Çabej, Eqrem. 1971. Der Beitrag des Albanischen zum Balkansprachbund. In *Actes du premier congres international des études balkaniques et sud-est européennes VII*, 41–54. Sofia: Éditions de l'Académie Bulgare des Sciences.

Çamaj, Martin. 1984. *Albanian grammar*. Wiesbaden: Otto Harrassowitz.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6:339–405.

Ciobanu, Fulvia, and Isabela Nedelcu. 2005. Prepoziția. In *Gramatica limbii române*, ed. Valeria Guțu Romalo, 607–630. Bucharest: Editura Academiei Române.

Civ'jan, T. V. 1965. *Imja suščestviltel'noe v balkanskix jazykax*. Leningrad: Nauka.

van Coetsem, Frans. 2000. *A general and unifed theory of the transmission process in language contact*. Heidelberg: Universitätsverlag C. Winter.

Cyxun, G. A. 1981. *Tipologičeskie problemy balkanoslavjanskogo jazykovogo areala*. Moscow: Izdatel'stvo "Nauka i texnika".

Dahl, Östen. 2001. Principles of areal typology. *Language Typology and Language Universals* 2:1456–1470.

Demina, Evgenija. 1985. *Tixonravovskij damaskin: Bolgarskij pamjatnik xvii v. issledovanie i tekst, tom 2*. Sofia: Izdatelstvo na Bŭlgarskata Akademija na Naukite.

Demiraj, Bardhyl. 2006. *Gjon P. Nikollë Kazazi dhe "Doktrina" e tij*. Prishtina: Akademia e Shkencave dhe e Arteve e Kosovës.

Demiraj, Shaban. 1988. *Gramatikë historike e gjuhës shqipe*. Prishtina: Rilindja.

Demiraj, Shaban. 1993. *Historische Grammatik der albanischen Sprache*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Demiraj, Shaban. 1994. *Gjuhësi balkanike*. Skopje: Logos-A.

Demiraj, Shaban. 2002. *Gramatikë historike e gjuhës shqipe*. Tirana: Akademia e Shkencave e Shqipërisë.

Demiraj, Shaban. 2004. Romanizimi në Evropën juglindore - Kohëzgjatja, përmasat dhe pasojat. In *Studime II*, ed. Idriz Ajeti, 95–112. Prishtina: Akademia e Shkencave dhe e Arteve e Kosovës.

Desnickaja, Agnija. 1979. O sovremennoj teorii balkanističeskih issledovanij. In *Problemy sintaksisa jazykov balkanskogo areala*, ed. Agnija Desnickaja. Leningrad: Nauka.

Desnickaja, Agnija. 1990. *Osnovy balkanskogo jazykoznanija*. Leningrad: Nauka.

Dombrowski, Andrew. 2013. Feature selection as optimization: Phonological aspects of language contact along the Slavic periphery. Unpublished PhD dissertation, University of Chicago, Chicago.

Duridanov, Ivan. 1956. *Die entwicklung des Bulgarischen vom syntetischen zum analytischen Sprachtyp*. Wiesbaden: Otto Harrassowitz.

Elezi, Shpëtim. 2010. Sistemi rasor i shqipes në çetën e Bogdanit. *Gjurmime Albanologjike - Seria e shkencave filologjike* 40:147–166.

Eloeva, Fatima, and Aleksandr Rusakov. 1990. *Problemy jazykovoj interferencii. cyganskije dialekty Evropy*. Leningrad: Leningradskij gosudarstvennyj universitet.

Elsie, Robert. 2005. *Albanian literature: A short history*. London: I.B. Tauris.

Emeneau, Murray. 1956. India as a linguistic area. *Language* 32:3–16.

Emeneau, Murray. 1980. Linguistic area: Introduction and continuation. In *Language and linguistic area*, ed. Anwar S. Dil, 1–18. Stanford University Press.

Enfield, Nicholas. 2008. Transmission biases in linguistic epidemiology. *Journal of Language Contact* THEME 2:299–310.

Ferguson, Charles. 1970. The Ethiopian language area. *The Journal of Ethiopian Studies* 8:67–80.

Feuillet, Jack. 1986. *La linguistique balkanique*. Paris: Publications Langues'O.

Fine, John. 1983. *The early medieval Balkans: A critical survey from the sixth to the late twelfth century*. Ann Arbor, MI: University of Michigan Press.

Fine, John. 1987. *The late medieval Balkans: A critical survey from the late twelfth century to the Ottoman conquest*. Ann Arbor, MI: University of Michigan Press.

Franks, Steven. 2009. The evolving status of pronominal clitics in macedonian. In *XXXVI naučna konferencija na XLII Meǵunaroden seminar za makedonskiot jazik, literatura, i kultura*, ed. Maksim Karanfilovski, Ubavka Gajdova, and Blagica Veljanovska, 183–204. Ohrid: Ss. Cyril and Methodius University.

Friedman, Victor. 1993. Macedonian. In *The Slavonic languages*, ed. Bernard Comrie and Greville Corbett, 249–305. New York: Routledge.

Friedman, Victor. 1994. Variation and grammaticalizations in the development of Balkanisms. In *CLS 30: Papers from the 30th regional meeting of the Chicago Linguistic Society, Volume 2: The parasession on variation in linguistic theory*, 101–115. Chicago: Chicago Linguistic Society.

Friedman, Victor. 1995. Persistence and change in Ottoman patterns of codeswitching in the Republic of Macedonia: Nostalgia, duress and language shift in contemporary southeastern Europe. In *Summer school: Code-switching and language contact*, 58–67. Ljouwert/Leuwaarden: Fryske Akademy.

Friedman, Victor. 1999. Linguistic emblems and emblematic languages: On language as flag in the Balkans. In *Kenneth E. Naylor memorial lecture series in South Slavic linguistics, no. 1*. Columbus, OH: Department of Slavic and East European Languages and Literatures, The Ohio State University.

Friedman, Victor. 2001a. Romani multilingualism in its Balkan context. *Sprachtypologie und Universalienforschung* 54:148–161.

Friedman, Victor. 2001b. The Vlah minority in Macedonia: Language, identity, dialectology, and standardization. In *Selected papers in Slavic, Balkan, and Balkan studies*, ed. Juhani Nuorluoto, Martii Leiwo, and Jussi Halla-aho, 26–50. Helsinki: University of Helsinki.

Friedman, Victor. 2002/3. Boundaries and borders in Balkan Slavic. *International Journal of Slavic Linguistics and Poetics* 44-45:161–173.

Friedman, Victor. 2003. 'One' as an indefinite marker in Balkan and non-Balkan Slavic. In *American contributions to the 13th International Congress of Slavists*, ed. Alan Timberlake and Robert A. Maguire, 93–112. Bloomington, IN: Slavica.

Friedman, Victor. 2004. Albanian grammar. In *Studies on Albanian and other Balkan languages*, ed. Eqrem Basha, Robert Elsie, and Rexhep Ismajli, 19–58. Peja: Dukagjini.

Friedman, Victor. 2005. Albanian in the Balkan linguistic league: A reconsideration of theoretical implications. *Acta Studia Albanica* 1:33–44.

Friedman, Victor. 2006a. The Balkans as a linguistic area. In *Encyclopedia of langauge and linguistics*, ed. Keith Brown, 657–672. Elsevier.

Friedman, Victor. 2006b. Determination and doubling in Balkan borderlands. *Harvard Ukrainian Studies* 28:105–116.

Friedman, Victor. 2007. Balkanizing the Balkan Sprachbund: A closer look at grammatical permeability and feature distribution. In *Grammars in contact: A cross-linguistic typology*, ed. Alexandra Aikhenvald and R. F. W. Dixon, 201–219. Oxford: Oxford University Press.

Friedman, Victor. 2008a. Balkan Slavic dialectology and Balkan linguistics: Periphery as center. In *American contributions to the 14th International Congress of Slavists, Ohrid, September 2008*, ed. Christina Yurkiw Bethin and David M. Bethea, 131–148. Bloomington, IN: Slavica.

Friedman, Victor. 2008b. Codeswitching and code integration in Romani. *Proceedings from the Annual Meeting of the Chicago Linguistics Society* 44:123–138.

Friedman, Victor. 2011. The Balkan languages and Balkan linguistics. *Annual Review of Anthropology* 40:275–291.

Friedman, Victor, and Robert Dankoff. 1991. The earliest known text in Balkan (Rumelian) Romani: A passage from Evliya Çelebi's Seyāḥat-nāme. *Journal of the Gypsy Lore Society (Fifth Series)* 1:1–20.

Gica, Alexandru. 2009. The recent history of the Aromanians in Southeast Europe. *Newsletter of the Society Farsarotul* 31.

Gjinari, Jorgji. 1965. Toskërishtja veriorë dhe grupimi i të folmevet të saj. In *Konferenca e parë e studimeve albanologjike*, 241–254. Tirana: Universiteti Shtetëror i Tiranës, Inistituti i Historisë dhe i Gjuhësisë.

Gjinari, Jorgji. 1970. *Dialektologija shqiptare*. Prishtina: Enti i teksteve dhe i mjeteve mësimore i Krahinës Solcialiste Autonome të Kosovës.

Gjinari, Jorgji, and Gjovalin Shkurtaj. 2003. *Dialektologji*. Tirana: Shtypshkronja "Perlat Voshtina".

Gołąb, Zbigniew. 1959. Some Arumanian-Macedonian isogrammaticisms and the social background of their development. *Word* 15:415–435.

Gołąb, Zbigniew. 1984. *The Arumanian dialect of Kruševo in SR Macedonia SFR Yugoslavia*. Skopje: Makedonska akademija za naukite i umetnostite.

Gołąb, Zbigniew. 1997. The ethnic background and internal linguistic mechanism of the so-called Balkanization of Macedonian. *Balkanistica* 10:13–19.

Goodman, Morris. 1993. African substratum: Some cautionary words. In *Africanisms in Afro-American language varieties*, ed. Salikoko Mufwene and Nancy Conson, 64–73. Athens, GA: University of Georgia Press.

Grannes, Alf. 1996. *Turco-Bulgarica: Articles in English and French conerning Turkish influence on Bulgarian*. Wiesbaden: Otto Harrassowitz.

Gumperz, John, and Robert Wilson. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border in India. In *Pidginization and creolization in language*, ed. Dell Hymes, 151–167. Cambridge: Cambridge University Press.

Haarmann, Harald. 1978. *Balkanlinguistik*. Tübingen: Gunter Narr Verlag.

Hamp, Eric. 1971. La langue albanaise et ses voisins. In *Actes du premier congres international des études balkaniques et sud-est européennes VII*, 663–667. Sofia: Éditions de l'Académie bulgare des sciences.

Hamp, Eric. 1977. On some questions of areal linguistics. In *Proceedings of the 3rd annual meeting of the Berkeley Linguistics Society*, ed. Kenneth Whistler, 279–282. Berkeley, CA: Berkeley Linguistics Society.

Hamp, Eric. 1982. The oldest Albanian syntagma. *Balkansko ezikoznanie - Linquistique balkanique* 25:77–79.

Hamp, Eric. 1989. On signs of health and death. In *Investigating obsolescence. Studies in language contraction and death*, ed. Nancy Dorian, 197–210. Cambridge: Cambridge University Press.

Hamp, Eric. 1994. The position of Albanian in Indo-European. In *Ancient Indo-European dialects*, ed. Jaan Puhvel and Henrik Birnbaum, 97–121. Berkeley, CA: University of California Press.

Harris, Alice, and Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.

Harris, Martin. 1988. The Romance languages. In *The Romance languages*, ed. Nigel Vincent and Martin Harris, 1–25. New York: Routledge.

Heine, Bernd, and Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Hinrichs, Uwe, ed. 1999. *Handbuch der Südosteuropa-Linguistik*. Wiesbaden: Otto Harrassowitz.

Hock, Hans Henrich, and Brian Joseph. 2009. *Language history, language change, and language relationship: An introduction to historical and comparative linguistics*. The Hague: Mouton de Gruyter.

Ivănescu, G. 1980. *Istoria limbii române*. Iaşi: Juminea.

Jacobson, William. 1980. Inclusive/exclusive: A diffused pronominal category in native western North America. In *Papers from the parasession on pronouns and anaphora*, 204–227. Chicago: Chicago Linguistics Socieety.

Jakobson, Roman. 1931/1971. Über die phonologischen Sprachbünde. In: Travaux du cercle linguistique de Prague 4: 234-240. In *R. Jakobson: Selected writings, Vol. 1 Phonological Studies*, 137–148. The Hague: Mouton de Gruyter.

Jakobson, Roman. 1938. Sur la théorie des affinités phonologiques entre les langues. In *Actes du quatrième congrès international de linguistes tenu à Copenhague du 27 aout au 1er septembre, 1936*, 351–365. New York: Kraus Reprints.

Jakobson, Roman. 1958/1971. Typological studies and their contribution to comparative-historical linguistics. In *R. Jakobson: Selected writings, Vol. I Phonological Studies*, 523–532. The Hague: Mouton de Gruyter.

Jireček, Konstantin. 1911. *Geschichte der Serben*. Gotha: F. A. Perthes.

Joseph, Brian. 1983. *The synchrony and diachrony of the Balkan infinitive: A study in areal, historical, and general linguistics*. Cambridge: Cambridge University Press.

Joseph, Brian. 1999. Romanian and the Balkans: Some comparative perspectives. In *The emergence of the modern language sciences*, ed. Sheila Embleton, John E. Joseph, and Hans-Joseph Niederehe, 218–235. Amsterdam: John Benjamins.

Joseph, Brian. 2001. Is Balkan comparative syntax possible? In *Comparative syntax of Balkan languages*, ed. María Luisa Rivero and Angela Ralli, 17–43. Oxford: Oxford University Press.

Joseph, Brian. 2009. Broad vs. localistic dialectology, standard vs. dialect: The case of the Balkans and the drawing of linguistic boundaries. In *Language variation—European perspectives II: Selected papers from the 4th International Conference on Language Variation in Europe (ICLaVE 4)*, ed. Stavroula Tsiplakou, Marilena Karyolemou, and Pavlos Pavlou, 119–135. Amsterdam: John Benjamins.

Joseph, Brian. 2010. Language contact in the Balkans. In *The handbook of language contact*, 618–633. Malden, MA: Wiley-Blackwell.

Kallulli, Dalina. 2001. Direct object clitic doubling in Albanian and Greek. In *Comparative syntax of Balkan languages*, ed. María Luisa Rivero and Angela Ralli, 127–160. Oxford: Oxford University Press.

Kallulli, Dalina, and Liliane Tasmowski. 2008. *Clitic doubling in the Balkan languages*. Amsterdam: John Benjamins.

Karagjozi, Mimoza. 2005. *Parafjalët në "Mesharin" e Gjon Buzukut*. Tirana: Botimet Toena.

Katičić, Radoslav. 1976. *Ancient languages of the Balkans, part one*. The Hague: Mouton de Gruyter.

Katz, Hartmut. 1975. *Generative Phonologie und phonologische Sprachbünde des Ostjakischen und Samojedischen*. Munich: Wilhelm Fink Verlag.

Kerswill, Paul. 2002. Koineization and accommodation. In *The handbook of language variation and change*, ed. Peter Trudgill and Natalie Schilling-Estes, 669–702. Oxford: Blackwell.

Koneski, Blaže. 1967. *Istorija na makedonskiot jazik*. Skopje: Kultura.

Konstantaras, Dean. 2006. *Infamy and revolt: The rise of the national problem in early modern Greek throught*. Boulder: East European Monographs.

Kopitar, Jernej. 1829. Albanische, walachische, und bulgarische Sprache. *Jahrbücher der Literatur* 46:59–106.

Kristoforidi, Konstantin. 1872. *Kaţer unġilat' e Zotit eδe šęĺbŭesit t'ünę Iesu-Krištit*. Istanbul: A. H. Byoaǵiani.

Kurth, Richard. 1904. Der Gebrauch der Präpositionen im Rumänischen. Unpublished PhD dissertation, Universität Leipzig, Leipzig.

Kuz, Josef. 1938. K otázce členu v jazycích slovanských, se zvláštním zřetelem k staroslověnštině. *Byzantinoslavica* 7:212–340.

Kyurkchiev, Nikolai. 2006. The Aromânians: An ethnos and language with a 2000-year history. *International Journal of the Sociology of Language* 179:115–129.

Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Lambertz, Maximillian. 1948. *Albanisches Lesebuch*. Wiesbaden: Otto Harrassowitz.

Leafgren, John. 2002. *Degrees of explictness: Information structure and the packaging of Bulgarian subjects and objects*. Philadelphia: Benjamins.

Leumann, Manu, J. B. Hofmann, and Anton Szantyr. 1965. *Lateinische Grammatik 1*. Munich: Beck.

L'Huillier, Monique. 1999. *Advanced French grammar*. Cambridge: Cambridge University Press.

Liaku-Anovska, Kleanti. 2007. *Vlaškite narodni prikazni od istočna makedonija*. Skopje: Institut za folklor "Marko Cepenkov".

Lindstedt, Jouko. 2000. Linguistic Balkanization: Contact-induced change by mutual reinforcement. *Studies in Slavic and General Linguistics* 28:231–246.

Löfstedt, Einar. 1956. *Syntactica. Studien und Beiträge zur historischen Syntax des Lateins*, volume I. Gleerup: Lund.

Longobardi, Giuseppe. 1994. Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry* 25:609–665.

Lunt, Horace. 2001. *Old Church Slavonic Grammar, 7th rev. ed.*. New York: Mouton de Gruyter.

Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.

MacRobert, C. M. 1982. The Balkan type of linguistic contact. In *Papers in Slavonic linguistics I*, ed. F. E. Knowles and J. I. Press, 143–161. Birmingham: Department of Modern Languages, University of Aston.

Mallory, James, and Douglas Adams. 1997. *Encyclopedia of Indo-European*. Oxford: Taylor & Francis.

Mann, Stuart. 1977. *An Albanian historical grammar*. Hamburg: Buske.

Mansaku, Seit. 1994. Problems of the typology of the Albanian language in the context of the Balkan languages. *Studia Albanica* 1:75–85.

Markoviḱ, Marjan. 2007. About some Aromanian-Albanian parallelisms in nominal and verbal system (areal linguistic approach). *Prilozi* 32:51–58.

Markoviḱ, Marjan. 2011. Za udvojuvanje na predlozite vo ohridskiot govor. In *XXXVII Naučna konferencija, pri Meǵunarodniot seminar za makedonski jazik, literatura, i kultura, Univerzitet "Sv. Kiril i Metodij"*, 195–203. Skopje: Makedonska akademija za naukite i umetnostite.

Markoviḱ, Marjan. 2012a. Macedonian language from the perspective of its Balkan environment (language tendencies). *Colloquia Humanistica* 1:81–95.

Markoviḱ, Marjan. 2012b. Predloški sekvenci za prostorno doopredeluvanje vo makedonskiot jazik. In *Zbornik za čest na 80-godišninata na akad. Zuzana Topolinjska*, 189–200. Skopje: Makedonska akademija za naukite i umetnostite.

Markoviḱ, Marjan. 2013. Za nekoi funkcii na predlogot za vo makedonskiot jazik. In *XXXIX Naučna konferencija pri Meǵunarodniot seminar za makedonski jazik, literatura i kultura*, 29–35. Skopje: Univerzitet "Sv. Kiril i Metodij".

Masica, Colin. 1976. *Defining a linguistic area: South Asia*. Chicago: University of Chicago Press.

Matras, Yaron. 1994. Structural Balkanisms in Romani. In *Sprachlicher Standard und Substandard in Südosteuropa und Osteuropa*, ed. Nobert Reiter, Uwe Hinrichs, and Jiřina van Leeuwen-Turnovcová, 194–210. Wiesbaden: Otto Harrassowitz.

Matras, Yaron. 2002. *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.

Matras, Yaron, and Jeanette Sakel. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31:829–865.

McWhorter, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5:125–166.

Mëniku, Linda, and Hector Campos. 2011. *Discovering Albanian I*. Madison, WI: University of Wisconsin Press.

Mesthrie, Rajend. 2001. Sociolinguistic variation. In *Concise encyclopedia of sociolinguistics*, 377–389. Amsterdam: Elsevier.

Michov, D. M. 1904. Die Anwendung des bestimmten Artikels im Rumänischen, verglichen mit der im Albanesischen und Bulgarischen. *Jahresbericht des Instituts für Rumänische Sprache zu Leipzig* 14:6–28.

Miklosich, Franz. 1862. Die slavischen Elemente im Rumänischen. *Denkschriften der Kaiserlichen Akademie der Wissenschaften, philosophisch-historische Klasse* 12:1–70.

Mirčev, Kiril. 1952. *Za makedonskija literaturen ezik*. Sofia: Nauka i izkustvo.

Mladenova, Olga. 2007. *Definiteness in Bulgarian: Modeling the process of language change*. Berlin: Mouton de Gruyter.

Mufwene, Salikoko. 2008. *Language evolution: Contact, competition, and change*. London: Continuum.

Muhamedova, Raihan. 2004. Untersuchungen zum kasasich-russischen Code-Mixing (mit Ausblicken auf den uigurisch-russischen Sprachkontakt). Unpublished PhD dissertation, University of Freiburg, Freiburg.

Mulaku, Ragip. 2012. *Gjuha e shkrimeve të vjetra shqipe (shek. xv-xviii)*. Prishtina: Rilindja.

Muysken, Pieter. 2000. From linguistic areas to areal linguistics: A research proposal. *Studies in Slavic and General Linguistics* 28:263–275.

Muysken, Pieter. 2008. Introduction: Conceptual and methodological issues in areal linguistics. In *From linguistic areas to areal linguistics*, ed. Pieter Muysken, 1–24. Amsterdam: John Benjamins.

Myers-Scotton, Carol. 1993. *Social motivations for codeswitching: Evidence from Africa*. Oxford: Oxford University Press.

Neculce, Ion, and Iorgu Iordan, ed. 1975. *Letopisețul țării moldovei*. Bucharest: Editura Minerva.

Nedelcu, Isabela. 2013. Prepositions and prepositional phrases. In *The grammar of Romanian*, ed. Gabriela Pană Dindelegan, 451–465. Oxford: Oxford University Press.

Newmark, Leonard, Phillip Hubbard, and Peter Prifti. 1982. *Standard Albanian*. Redwood City, CA: Stanford University Press.

Nichols, Johanna, and Balthasar Bickel. 2006. Oceania, the Pacific Rim, and the theory of linguistics areas. In *Proceedings of the 32nd annual meeting of the Berkeley Linguistics Society*, 3–15. Berkeley, CA: University of California Press.

Nicolova, Ruselina. 2008. *Bŭlgarska gramatika: Morfologija*. Sofia: Universitetsko izdatelstvo Sv. Kliment Ohridski.

Orel, Vladimir. 2000. *A concise historical grammar of the Albanian language: Reconstruction of Proto-Albanian*. Leiden: Brill.

Pipa, Arshi. 1989. *The politics of language in socialist Albania*. Ann Arbor, MI: East European Monographs, University of Michigan.

Plank, Frans. 1995. *Double case: Agreement by suffixaufnahme*. New York: Oxford University Press.

Polomé, Edgar. 1982. Balkan languages (Illyrian, Thracian and Daco-Moesian). In *Cambridge ancient history III*. Cambridge: Cambridge University Press.

Pop, Sever. 1948. *Grammaire roumaine*. Bern: Francke.

Poplack, Shana. 1987. Contrasting patterns of code-switching in two communities. In *Aspects of multilingualism: Proceedings from the fourth nordic symposium on bilingualism 1984*, ed. Erling Wander, 51–77. Uppsala: Uppsala Universitet.

Price, Glanville. 2003. *A comprehensive French grammar*. Oxford: Blackwell.

Prifti, Elton. 2012. Aromanian elements of an Albanian argot. In *Balkanismen heute - Balkanisms today - balkanizmy segodnja*, ed. Thede Kahl, Michael Metzeltin, and Helmut Schaller, 245–263. Münster: Lit Verlag.

Pulaha, Selami. 1984. *Popullsia shqiptare e Kosovës gjatë shek. xv-xvi*. Tirana: Akademia e shkencave e RPS të Shqipërisë, Instituti i Historisë.

Récatas, Basile. 1934. *L'état actuel du bilinguisme chez les Macédo-Roumains du Pinde et le rôle de la femme dans le langage : avec cartes, photographies et textes macédo-roumains*. Geneva: E. Droz.

Reichenkron, Günther. 1962. Der Typus der Balkansprachen. *Zeitschrift für Balkanologie* 1:91–122.

Reinheimer Rîpeanu, Sanda. 2001. *Lingvistică romanică*. Bucharest: All universitar.

Reiter, Nobert. 1994. *Grundzüge der Balkanologie: Ein Schritt in die Eurolinguistik*. Wiesbaden: Otto Harrassowitz.

Riza, Selman. 2002. *Pesë autorët më të vjetër në gjuhën shqipe*. Tirana: Botimet Toena.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar*, ed. L. Haegeman. Dordrecht: Kluwer.

Rosén, Hannah. 1994. The definite article in the making, nominal constituent order, and related phenomena. In *Linguistic Studies on Latin: Selected Papers from the 6th International Colloquium on Latin Linguistics (Budapest, 23-27 March 1991)*, ed. József Herman, 129–150. John Benjamins.

Rosetti, Alexandru. 1968. *Istoria limbii române: De la origini pînă în secolul al XVII-lea*. Bucharest: Editura pentru literatură.

Rusakov, Aleksandr. 2013. Nektorye izoglossy na albanskoj diaklektnoj karte (k voprosu o vozniknovenii i rasprostranenii balkanizmov albanskogo jazyka). In *Issledovanija po topologii slavjanskix, balktijskix i balkanskix jazykov (preimuščesvenno v svete jazykovyx kontaktov*, ed. Vjačeslav Ivanov, 113–174. Mosco: Aleteija.

Sandfeld, Kristian. 1926. *Balkanfilologien: En oversigt over dens resultater og problemer*. Copenhagen: Lunp.

Sandfeld, Kristian. 1930. *Linguistique balkanique: Problèmes et résultats*. Paris: Champion.

Sasse, Hans-Jürgen. 1998. Small languages and small language communities 26: Arvanitika: The long Hellenic centuries of an Albanian variety. *International Journal of the Sociology of Language* 134:39–68.

Schaller, Helmut Wilhelm. 1975. *Die Balkansprachen: Eine Einführung in die Balkanphilologie*. Heidelberg: Universitätsverlag C. Winter.

Schaller, Helmut Wilhelm. 1997. Roman Jakobson's conception of 'Sprachbund'. In *Jakobson entre l'Est et l'Ouest, 1915-1939*, 207–212. Lausanne: Cahiers de l'ILSL 9.

Schleicher, August. 1850. *Die Sprache Europas in systematischer Übersicht*. Amsterdam: John Benjamins.

Schwandner-Sievers, Stephanie. 1999. *The Albanian Aromanians' awakening: Identity politics and conflicts in post-communist Albania*. Flensburg: European Centre for Minority Issues.

Schwandner-Sievers, Stephanie. 2002. Dawn for a 'sleeping beauty nation'. Aromanian identity politics and conflicts in post-communist Albania. In *Albania - A country in transition. Aspects of changing identities in a south-east European country.*, ed. Karl Kaser and Frank Kressing, 147–163. Baden-Baden: Nomos.

Seliščev, A. M. 1925. Des traits linguistiques communs aux langues balkaniques: Un balkanisme ancien en bulgare. *Revue des études slaves* 5:38–57.

Sherzer, Joel. 1976. Areal linguistics in North America. In *Native languages of the Americas*. Berlin: Springer.

Sobolev, Andrej. 1991. K istorii utraty balkanoslavjanskogo sklonenija. *Zbornik Matice srpske za filologiju i lingvistiku* 34:7–41.

Solta, Georg. 1980. *Einführung in die Balkanlinguistik mit besonderer Berücksichtigung des Substrats und des Balkanlateinischen*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Steinke, Klaus. 1968. *Studien über den Verfall der bulgarischen Deklination*. Munich: Otto Sanger.

Steinke, Klaus, and A. Vraciu. 1999. *Introducere in lingvistica balcanica*. Iasi: Editura Universității.

Stojanov, Ljudmil. 1952. Za turcizimite i dialketizimite v bŭlgarskija literaturen ezik. *Izvestija na Instituta za bŭlgarskija ezik* 2:218–219.

Stolz, Thomas. 2002. No Sprachbund beyond this line! On the age-old discussion of how to define a linguistic area. In *Mediterranean languages*, ed. Paolo Ramat, 259–281. Bochum: Universitätsverlag Dr. N. Brockmeyer.

Stowell, Tim. 1989. Subjects, specifiers and x'-theory. In *Alternative conceptions of phrase structure*, ed. Mark Baltin and Anthony Kroch, 232–262. Chicago: University of Chicago Press.

Szabolcsi, Anna. 1987. Functional categories in the noun phrase. In *Approaches to Hungarian 2: Theories and analyses*, ed. Istvan Kenesey, 167–189. Szeged: Jate.

Thomason, Sarah. 2001. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.

Thomason, Sarah, and Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley, CA: University of California Press.

Thunmann, Johann. 1774. *Untersuchungen über die Geschichte der östlichen europäischen Völker*. Leipzig: Siegfried Lebrecht Erusius.

Todoran, Romulus. 1977. Formarea dialectelor românești. In *Dialetologie Română*, ed. Aretia Dicu, 97–111. Bucharest: Editura științifică și encipclopedică.

Tomić, Olga Mišeska. 2006. *Balkan Sprachbund morpho-syntactic features*. Dordrecht: Spinger.

Topalli, Kolec. 2009a. *Nyjat e shqipes*. Tirana: Botimet Toena.

Topalli, Kolec. 2009b. *Sistemi rasor i emrave të shqipes*. Tirana: Plejad.

Topalli, Kolec. 2011. *Bazat e gramatikës historike të gjuhës shqipe*. Tirana: Shtëpia botuese e librit universitar.

Topolińska, Zuzanna. 1991. Quelle est l'interprétation la plus plausible du term 'syntaxe balkanique'? In *VI Meǵunaroden kongres za proučuvanje na jugoistočna Evropa (Sofija, 1989). Prilozi na učesnicite od Makedonija*. Skopje: Makedonska akademija za naukite i umetnostite.

Topolińska, Zuzanna. 1998. In place of a foreword: Facts about the Republic of Macedonia and the Macedonian language. *International Journal of the Sociology of Language* 131:1–11.

Topolińska, Zuzanna. 2007. Towards a reformulation of the inventory of Balkanisms. *Prilozi* 32:65–72.

Tosco, Mauro. 2008. What to do when you are unhappy with language areas but you do not want to quit. *Journal of Language Contact* 1:112–123.

Trask, Robert. 2000. *The dictionary of historical and comparative linguistics*. Edinburgh: Edinburgh University Press.

Trubetzkoy, Nikolay. 1923. Vavilonskaja bašnja i smešenie jazykov. *Evrazijskij vremennik* 3:107–124.

Trubetzkoy, Nikolay. 1928. [Proposition 16]. In *Actes du premier congrès international des linguistes à la Haye, du 10-15 avril 1928*, 17–18. A. W. Sejthoff.

Trubetzkoy, Nikolay. 1931. Phonologie und sprachgeographie. *Travaux du cercle linguistique de Prague* 4:228–234.

Trudgill, Peter. 1972. Sex, covert prestige, and linguistic change in the Urban British English of Norwich. *Language in Society* 1:179–195.

Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.

Tsitsipis, Lukas. 1998. *A linguistic anthropology of praxis and language shift: Arvanítika (Albanian) and Greek in contact*. Gloucester: Clarendon Press.

Velten, Harry. 1943. The Nez Perce verb. *The Pacific Northwest Quarterly* 34:271–292.

Wahlström, Max. 2015. The loss of case inflection in Bulgarian and Macedonian. Unpublished PhD dissertation, University of Helsinki, Helsinki.

Weigand, Gustav. 1894/1895. *Die Aromunen: Ethnographisch-philologisch-historische Untersuchungen über das Volk der sogennanten Makedo-Romanen oder Zinzaren*. Leipzig: J. A. Barth.

Weinreich, Uriel. 1953. *Languages in contact*. Berlin: Walter de Gruyter.

Weiss, Daniel. 1996. Die Geburt eines Artikels: Zum Status von makedonisch *eden*. In *Slavistische Linguistik 1995: Referate des XXI Konstanzer Slavistischen Arbeitstreffens*, ed. W. Girke, 421–455. Munich: O. Sagner.

Woordard, Roger. 2004. *The Cambridge encyclopedia of the world's ancient languages*. Cambridge: Cambridge University Press.

Xhaferi, Haredin. 2011. *Morfologji e gjuhës shqipe*. Tirana: Shtëpia Botuese Botart.

Xhuvani, Aleksandër. 1957. Çështja e fjalorit të gjuhës shqipe. *Nëndori* 1:151–163.

Zegrean, Iulia-Georgiana. 2012. Balkan Romance: Aspects on the syntax of Istro-Romanian. Unpublished PhD dissertation, Università Ca'Foscari, Venice.