SLAVICA HELSINGIENSIA

47

THE LOSS OF CASE INFLECTION IN BULGARIAN AND MACEDONIAN

Max Wahlström

HELSINKI 2015

SLAVICA HELSINGIENSIA 47

Series editors
Tomi Huttunen, Jouko Lindstedt, Ahti Nikunlassi

Published by:
Department of Modern Languages
P.O. Box 24 (Unioninkatu 40 B)
00014 University of Helsinki
Finland

Copyright © by Max Wahlström
ISBN 978-951-51-1185-2 (paperback)
ISBN 978-951-51-1186-9 (PDF)
ISSN-L 0780-3281, ISSN 0780-3281 (Print), ISSN 1799-5779 (Online)

Summary

Case inflection, characteristic of Slavic languages, was lost in Bulgarian and Macedonian approximately between the 11th and 16th centuries. My doctoral dissertation examines the process of this language change and sets out to find its causes and evaluate its consequences. In the earlier research literature, the case loss has been attributed either to language contacts or language internal sound changes, yet none of the theories based on a single explaining factor has proven satisfactory.

In this study, I argue that the previous researchers of the Late Medieval manuscripts have often tried to date changes in the language earlier than what is plausible in light of the textual evidence. Also, I propose that the high number of second language speakers is among the key factors that reduced the number of morphological categories in the language, but, at the same time, several minor developments related to the case loss—for instance, in the marking of possession—are likely to result from a specific contact mechanism known as the Balkan linguistic area. My main methodological argument is that the study of language contacts must take into account a general typological perspective to determine the uniqueness of the suspected contact-induced changes. Further, quantitative typological methods are also helpful in assessing, whether the co-occurrence of linguistic features within a linguistics area is truly independent and not explained by universal tendencies.

This dissertation is divided into three parts, each attaching to a different methodological approach. Through corpus methods, among other, the first of these examines the process of the loss of case inflection within the manuscript tradition that stems from Old Church Slavonic. The second approach is based on the study of language contacts. I compare the development of the Bulgarian and Macedonian case systems with the Albanian, Balkan Romance, and Greek case systems and their evolution. In addition, taking Romani language as an example, I analyze the effect of sociolinguistic setting on the type of contact-induced language change. The third approach studies the case systems of Balkan Slavic and the rest of the Balkan linguistic area in a typological connection to evaluate to which extent the phenomena related to the case loss can be attributed to universal tendencies, observed in the languages of the world.

Acknowledgments

I often feel that the things that led me to linguistics happened by chance, without any planning or conscious decision on my part. Yet the impression of coincidence is wrong. On my path to becoming a linguist, philologist, and Slavicist, I have been helped by many people whose roles have often proven decisive and whom I therefore wish to acknowledge here. Even if I fail to mention every name here, I assure you that no one has been forgotten.

First and foremost, I am indebted to my supervisor Jouko Lindstedt. His guidance in the writing of this study has been crucial in numerous ways. I am sure that through our conversations, many of his thoughts and ideas have found their way onto these pages, something that I am not ashamed to admit, since, as he himself once put it, I am the one who, after all, has to defend them. In addition to being one of the most educated linguists I have ever met, Jouko has also been my boss for almost a decade. During this time I have learned to trust his unwavering support, wisdom, and impartiality. This atmosphere of security has allowed me to pursue my goals and, through trial and error, learn many important lessons.

I consider myself extremely lucky to have had not one, but two extremely good supervisors. I found my second supervisor, Kaius Sinnemäki, three years into working on my doctoral thesis. Although I had very specific knowledge in mind when I contacted Kaius, I am still amazed at the depth, skill, and dedication with which he guided me through my entire study. Conversations with him have significantly broadened my perspective on scientific thinking. Moreover, whenever I encountered unexpected adversity, his empathy and encouragement were often exactly what was needed to restore my self-confidence.

In thinking about the career choices I never made deliberately, but which just happened, one person in particular, a teacher and mentor, stands out whom I must thank: Jussi Nuorluoto. His decision to hire me as a research assistant at a point when I did not have many academic credentials to boast about is no doubt the single most important event that caused me to pursue linguistics. With great joy, I felt the circle closing when Jussi offered his help in the final stages of writing this study, and he provided invaluable comments on the manuscript with the precision and intelligence I have come to expect from him.

I would like to thank my fellow students in Slavic philology during my undergraduate studies. To name only a handful of these dear colleagues,

Markus Sipilä, a beer with you was always a most welcome lesson in linguistics; Jaakko Turunen, I miss our conversations on social theory; and Laura Čupić, you taught me so much about the Balkans! I am also indebted to all my teachers of South Slavic languages, both in Helsinki and abroad: Thank you, Dragana Cvetanović, Marica Čilaš Mikulić, Mina Đurić, Kristina Kontkanen, Atina Nihtinen, Dinka Pasini, Andrijana Pavlova, Olga Sabo, and Staša Tofoska. In this connection, one person in particular deserves special mention, my teacher of the Croatian language, Rada Borić, who deeply inspired a whole generation of students in Helsinki, myself included. In fact, for me becoming a linguist has meant becoming a teacher as well. On this path, I have been able to draw on the examples of many great teachers throughout an almost uninterrupted quarter of a century of being educated. A warm thank you to all.

My godmother and my first academic role model, Tarja Sokka, dedicated a copy of her doctoral thesis to me when I was just fourteen, with the words: "to serve as an example." I wonder if she realized then how seriously I took that line. Embarking on postgraduate studies, I had three particular role models in mind who shaped my image of the kind of researcher I wanted to be: Eeva Luhtakallio, Janne Saarikivi, and Juha-Heikki Tihinen. Their broad knowledge and enthusiasm for things beyond their immediate fields of expertise still continue to inspire me. I would also like to thank Juha-Heikki, in addition to his greatly valued friendship, for the comments on my manuscript, as well as on several earlier works of writing. Another academic I continue to look up to is Victor Friedman, whom I had the pleasure to get to know in Helsinki in 2007. I remember asking him on our way into town from the airport why the cases were lost in Bulgarian and Macedonian. Although he did not give an answer, thereby making me find out for myself, he taught me more about everything during that week than probably anyone since. In the same vein, I am honored to have met two of the most gifted Balkanists in the younger generation, Maxim Makartsev and Eric Prendergast, and to call them my friends. Discussions with you have truly stimulated my motivation and curiosity throughout the years!

At the beginning of my doctoral studies, I was privileged to be accepted into the Finnish doctoral program in language studies, Langnet, which provided a number of readers for my academic writing in a productive and supportive environment. Much of creating such a fantastic place to mature as a linguist must be credited to Riho Grünthal, as well as to other heads of my sub-program. Here, I also want to thank all my fellow students in

Langnet—for your valuable comments and for putting up with my often extravagant enthusiasm. Many of you have become trustworthy colleagues and even friends, and I am looking forward to working with you in years to come. During my doctoral studies, I also had the opportunity to visit the Research Center for Areal Linguistics of the Macedonian Academy of Sciences and Arts in the spring of 2011. For this I would like to thank my welcoming hosts, Marjan Markoviḱ and Zuzanna Topolińska, as well as my always hospitable and helpful colleagues, both in matters personal and linguistic, Aneta Dučevska, Elka Jačeva-Ulčar, Maksim Karanfilovski, Ljudmil Spasov, and everyone in the Wednesday Round Table and in the office.

I would like to thank my pre-examiners Ronelle Alexander, Brian Joseph, and Mila Vulchanova for their comments and suggestions, many of which have greatly enhanced the overall quality of this study. In addition, parts of the manuscript as well as my conference presentations on the same themes have been commented on by many. I have also sought advice from a number of experts. I would therefore like to thank Petya Assenova, Maria Basdekis, Tat'jana Civ'jan, Sonja Dahlgren, Andrei Dumitrescu, Zlatka Guenchéva, Hilla Halla-aho, Hanna-Ilona Härmävaara, Laura Janda, Seppo Kittilä, Timo Korkiakangas, Ian Mac Eochagáin, Dionysios Mertyris, Tore Nesset, Pekka Posio, and Joanne Stolk. The English language of an earlier version of this manuscript was skillfully revised by Glenda Dawn Goss, who, along with all of those mentioned above, nevertheless bears no responsibility for any mistakes or inaccuracies in the present work. Nor do its contents necessarily reflect their recommendations.

For the past nine years, five of them as a doctoral student, I have enjoyed the pleasurable company of my colleagues in what was once known as the Department of Slavonic and Baltic Languages and Literatures at the University of Helsinki. Hearty thanks to all for these years and for making me feel at home. I would especially like to thank Johanna Virkkula, who has helped me with so many things that I lost count years ago. A colleague for several years whose significance to me is almost beyond words is Lotta Jalava, a fellow linguist and a very dear friend: where would I be without you? I would also like to thank Ilkka Järvinen, who, in addition to his lasting friendship, was my first critical reader.

Finally, I would like to thank the rest of my beloved extended family of friends for believing in me no matter what. You should know how much the things you have done for me have meant. Whether it was inviting me to

drink a beer or two or attend a party or a picnic or cooking a meal for me, washing my dishes, playing badminton and volleyball together—all of these things kept me sane throughout the most difficult periods of finishing my thesis. I am also one of those fortunates whose logical family coincides with the biological family: Äiti, isä, Ella, kiitos kaikesta, sanat eivät riitä!

Contents

Summary III

Acknowledgments IV

Abbreviations X

1 Introduction 1
- 1.1 Organization of the study 2
- 1.2 Data, methods, and theoretical background 3
- 1.3 Key concepts and terminology 5
  - 1.3.1 Bulgarian, Macedonian, and Balkan Slavic 5
  - 1.3.2 Case 7
    - *Is the vocative a case?* 9
    - *Case syncretism* 11
- 1.4 Earlier research on the loss of the Balkan Slavic case inflection 13
  - 1.4.1 Internal and external explanations 14
  - 1.4.2 Phonetic, syntactic, and semantic explanations 21
  - 1.4.3 Historical grammars, manuscript studies, and synchronic analyses 23

2 The diachrony of the Balkan Slavic case system 27
- 2.1 From the Proto-Indo-European to the Slavic case system 27
  - 2.1.1 The Proto-Slavic case system 28
  - 2.1.2 The dialectal developments of the case system 30
- 2.2 The case system during the Middle Balkan Slavic period 33
  - 2.2.1 On the written sources 35
  - 2.2.2 Nominative, accusative, and genitive 40
    - *Nominative* 41
    - *Accusative* 41
    - *Genitive in the dialects* 47
    - *Genitive replaced by other cases* 49
    - *Genitive replaced by prepositions* 52
  - 2.2.3 Dative 55
    - *Dative as the recipient* 58
    - *Dative as the addressee: Increase in the use of the preposition kъ?* 62
    - *Other uses of the dative as a verbal complement* 65
    - *Dative with prepositions* 70
  - 2.2.4 The locative and instrumental cases 72
    - *Locative* 73
    - *Instrumental* 74
- 2.3 Sound changes and the loss of case inflection 75
- 2.4 Conclusions: The Middle Balkan Slavic case system 81

3 The Balkan case systems 85
- 3.1 Language contact and areal linguistics 85
  - 3.1.1 Language contact and convergence: The Balkans as a linguistic area 86
    - *The Balkan linguistic area: A history of the notion* 87

Numbers and types of shared features in linguistic areas 88
Historical and sociolinguistic factors leading to convergence 92
3.1.2 Language contact and simplification: Resistant Romani 95
Romani: Retention of inflection 96
Simplification in language contact: The number of L2 speakers 97
3.2 The diachrony of the Balkan case systems 100
3.2.1 Albanian..100
3.2.2 Balkan Romance 105
3.2.3 Greek 110
3.3 Balkan argument marking beyond case inflection 114
3.3.1 The article as a carrier of case marking...115
3.3.2 Pronouns and the clitic doubling of objects..117
3.3.3 The Albanian and Balkan Romance "genitive" 119
3.3.4 Differential object marking in the Balkan languages...122
The accusative in the Balkan languages..124
Prepositional DOM in Macedonian of the Ohrid–Struga region 126
Effect of referential scales on argument marking 128
3.4 Conclusions: Is there a Balkan case system? 129

4 The loss of case inflection in Balkan Slavic from a typological perspective.........135
4.1 Linguistic typology and areal linguistics 137
4.1.1 Areal linguistics and linguistic typology...137
4.1.2 Areally-informed quantitative linguistic typology..139
4.1.3 Typological universals as explanations..144
4.2 The Balkan recipient–possessor merger 146
4.2.1 The merger of the morphological cases...147
4.2.2 Recipient–possessor merger: Between verbal and nominal syntax 150
4.2.3 Analytical marking of the recipient and possessor 154
4.2.4 Split-conditioned double marking of possession and the juxtapositional pseudo-partitive 157
4.2.5 Areal and typological limits of the merger...163
4.2.6 Conclusions..167
4.3 The fall of case inflection, the rise of the definite article: Evidence for a universal tendency? 170
4.3.1 The definite article in Europe 170
4.3.2 Statistical analysis of the data..173
4.3.3 Discussion...177
4.3.4 Conclusions..181

5 Conclusions 185
5.1 Sound changes and the case system: The inability to regenerate.......................186
5.2 The loss of case inflection as a Balkan development 187
5.3 The loss of case inflection from a cross-linguistic perspective...........................189

References 193

Abbreviations

A	agent argument of transitive verbs
ACC (acc.)	accusative
AOR	aorist
AUX	auxiliary verb
BCMS	Bosnian, Croatian, Montenegrin, Serbian
Bg.	Bulgarian
BLT	basic linguistic theory
CL	clitic
COMP	complementizer
DAT (dat.)	dative
DEF	definite article
F	feminine
GEN (gen.)	genitive
IMP	imperative
IND	indicative
INSTR (instr.)	instrumental
L2	second language
Lat.	Latin
LINK	linking article
LOC (loc.)	locative
M	masculine
Mac.	Macedonian
MBS	Middle Balkan Slavic
N	neuter
NOM (nom.)	nominative
NP	noun phrase
OCS	Old Church Slavonic
P	patient argument of transitive verbs
PAT	Predictive Areality Theory
PIE	Proto-Indo-European
PL	plural
PREP	preposition
PRS	present tense
PTCP	participle
REFL	reflexive pronoun
Ro.	Romanian

S	single argument of intransitive verbs
SAE	Standard Average European
SG	singular
VOC (voc.)	vocative
?	a form of uncertain identity

1 Introduction

Macedonian and Bulgarian differ from all other Slavic languages in several respects: For example, there is a postpositional definite article that attaches to the first word of the noun phrase. Also, the two languages have retained the past tense categories aorist and imperfect, which have been almost completely lost in the other Slavic languages. But perhaps most strikingly, while all other Slavic languages express the roles of noun phrases within a sentence by using both prepositions and case marking, Bulgarian and Macedonian have gradually lost their case inflection in its entirety. The last of these characteristics of Bulgarian and Macedonian—the loss of case inflection—is the subject of this study. The question belongs to the classical problems in the study of Slavic languages, no doubt because the lack of case inflection contrasts so strikingly with what is a characteristic and a relatively uniform feature of the other languages.

My main task is to answer the questions, when the cases were lost, how this loss took place, and, most important, why. To tackle the questions successfully, I have organized the study around three different approaches, each of which seeks to illuminate one crucial aspect of the destiny of the Bulgarian and Macedonian case inflection. The first approach is to analyze the written evidence and establish a picture of the historical process. The second is to evaluate the role of language contact in the historical process, given that Bulgarian and Macedonian belong to a group of languages that form what is called the Balkan linguistic area. These languages display a number of shared features, such as the definite article, and answers must therefore also be sought in their convergence. The third approach is based on the observation that the loss of case inflection is not limited to Bulgarian and Macedonian. In fact, almost all Indo-European languages spoken in Western Europe have undergone the same process. That a definite article has emerged in these languages as well compels us to ask whether these phenomena are interconnected and the result of some universal processes of language change. This question is addressed through two cross-linguistic case studies. A more detailed description of the study's organization and its main arguments are introduced below.

1.1 Organization of the study

The general preliminaries to this study are given in the present chapter, which includes a presentation of the data, methods, and theoretical background, offers some terminological considerations, and gives a summary of the earlier research on the subject of the loss of case inflection in the Bulgarian and Macedonian languages.

Chapter 2, entitled "The diachrony of the Balkan Slavic case system," examines the diachronic development of the Bulgarian and Macedonian case systems. **Section 2.1** serves as a starting point for an analysis of the subsequent changes in their case inflection. **Section 2.2** continues with an analysis of the previous research and the proposed theories regarding the changes in the case system. **Section 2.3** discusses the role of sound changes, in particular, in the case loss. In **Section 2.4**, I summarize the findings and evaluate the potential language-internal causes for the loss of case inflection. In the chapter, I argue for a cautious approach in analyzing the written sources. My main observation is that in previous research too much weight has been given to individual examples that are thought to represent particular changes in the spoken language. These individual instances may be indicative of the growing gap between the written and the spoken language, but the evidence is mostly indirect.

Chapter 3, entitled "The Balkan case system," seeks to establish an understanding of the similarities and differences among the individual case systems in the Balkan linguistic area by providing an areal point of comparison for the diachronic development of the Balkan Slavic case system. **Section 3.1** introduces the contact linguistic phenomena that are relevant to the convergence of the Balkan languages and the loss of case inflection in Balkan Slavic. In **Section 3.2**, I map the synchronic case systems and their development in Albanian, Balkan Romance, and Greek. **Section 3.3** addresses some key issues of argument marking in the Balkan linguistic area that reach beyond mere case inflection. **Section 3.4** presents concisely the similarities between the individual Balkan case systems and evaluates the potential contact-induced element in their development. My main claim in this chapter is that the recent advances in the study of contact linguistics offer potential explanations not only for the shared features between the languages, but also for such indiscriminate changes as the loss of case inflection.

Chapter 4, entitled "The loss of case inflection in Balkan Slavic from a typological perspective," analyzes the Balkan and Balkan Slavic case

systems with the help of two case studies that include a cross-linguistic component. **Section 4.1** serves as an introduction to the case studies by analyzing the relationship between linguistic typology and areal linguistics. In **Section 4.2**, I assess the dative–genitive merger, a feature of the Balkan linguistic area, in comparison with similar mergers in other, mostly European languages. I claim that despite the commonness of the merger, when the data is analyzed more carefully, the phenomenon stands out in the Balkans. **Section 4.3** seeks to understand the reduction of case inflection and the grammaticalization of the definite article. Using quantitative typological data I test a hypothesis regarding the correlation between the two phenomena in the languages of the world and propose a possible explanation for the potential universal tendency.

Chapter 5, ”Conclusions,” summarizes the findings of the three approaches and then presents these in a concise manner in a final discussion.

1.2 Data, methods, and theoretical background

The data used in this study include dialectal material, linguistic descriptions, written texts, reconstructed word forms, and a typological database. The variety of sources stems primarily from the temporal dimension of the research question: The development of the loss of case inflection has its starting point in the pre-literary era, accessible only through comparative linguistics, which enables the reconstruction of the earlier states of the language through sound correspondences. There are written sources from the following period, during which many key changes took place. These sources, however, often differ radically in their features from the spoken reality (the periodization of the written sources is discussed in **2.2**). Since the 19th century, dialectal records have also been available, which, ideally, should represent the spoken linguistic varieties without the interference of literary traditions as a medium. Finally, since many central questions in this study involve abstract diachronic, synchronic, and cross-linguistic comparisons, much of the data has been acquired from linguistic descriptions included in grammars, monographs, and scholarly articles.

The choice of methods in this study is dictated by the type of data: The comparative method is used to acquire data that are not immediately available through more direct sources (see **Sections 2.1**, **2.3**, and **3.2**). On the one hand, the role of written records as a source for the spoken language is evaluated vis-à-vis the philological tradition, which combines the knowledge of a vast array of subfields, such as book history and the study of

graphemics and scribal traditions (**2.2**). Corpus methods, on the other hand, are used to detect the occurrence and distribution of patterns in the texts, and to contrast different manuscripts (**2.2**). **Section 4.3** uses quantitative analysis of typological data to evaluate the role of language universals in explaining two co-occurring features in the Balkan languages.

This study does not adhere to any particular formal theoretical framework. Typically, such formal presentations are at their best with neatly outlined problems, whereby a formalization may convey the observations more elegantly than would be possible through prose. There are several reasons why I do not find such approaches feasible here: The loss of case inflection in Bulgarian and Macedonian is a very large topic and is approached from several different viewpoints in this study, which means that no single framework could be applied (for further arguments, see **1.3.2**). Yet at least some of these viewpoints could be seen as benefiting from a more structured, formal approach. In fact, similar problems have been approached from such theoretical frameworks; consider, for example, Wouter Kusters (2003, discussed in **3.1.2**), who employs optimality theory in the evaluation of changes in morphologic complexity. However, my impression has often been that what is achieved through the formal approaches does not necessarily amount to more clarity in the actual linguistic problems. This is also true in the case of Kusters's monograph, although his conclusions are nothing short of insightful and compelling. The more detrimental effect of some of the formal theoretical approaches is that, while they may provide models and hypotheses to be tested, they are also used to solidify claims about human cognition, which sometimes happens at the expense of empirical linguistic reality. This is a risk I am not willing to take.[1]

Apart from the terminological and conceptual traditions of historical Slavic linguistics and the study of language contacts, this study employs what is best described as a functional-typological approach, perhaps exemplified in the most concrete manner by the basic linguistic theory (BLT) of R.M.W. Dixon (2010a; 2010b; 2012). My understanding of BLT also coincides closely with Martin Haspelmath's (2010) proposition of framework-free grammatical theory, although Haspelmath does distance himself from BLT (ibid., 359–365). A functional-typological approach

[1] I have been given specific suggestions about particular frameworks on occasions where I have presented aspects of my current research. Although I eventually opted not use any of these suggestions, I am very grateful for these kind offers of help, especially as becoming familiar with these approaches has helped me to clarify the focus of the study.

employs terminology that allows the widest possible cross-linguistic identification of related linguistic phenomena, with the crucial advantage that each category and concept is based on the accumulated typological knowledge and is open to any new advances in the relevant fields of study. However, the result is that a functional-typological approach is not very theory-like (despite Dixon and Haspelmath's decision to call their approaches theories) in that it does not aim for the construction of a general theory of language. In the present study, the chosen approach means that the discussions and analyses seek purposely to be accessible to anyone familiar with the modern typological literature by avoiding field-specific shorthand. The potential drawback is that some scholars of Slavic languages and the Balkan linguistic area inevitably find the explicitness superfluous. This is, however, an unavoidable concession for the benefit of the cohesion of the entire work.

More specific theoretical questions will be addressed in the opening sections of **Chapters 2**, **3**, and **4**, adjacent to the analyses where they are discussed.

1.3 Key concepts and terminology

The key concepts and terminology will be defined mainly in the chapters and sections where they are introduced. In many instances, the concepts and terms are themselves debated, as are the phenomena to which they refer in this study. Therefore, it does not seem reasonable to treat them separately, especially as the study covers several very different approaches whose preliminaries would unnecessarily inflate the present chapter. Nevertheless, two general concepts are briefly presented here, since they are relevant to the whole study: First, the Slavic languages and linguistic varieties at the focus of this study are outlined, and, second, the concept of grammatical case is defined, with particular attention given to the vocative case and case syncretism.

1.3.1 Bulgarian, Macedonian, and Balkan Slavic

This study focuses on two South Slavic languages, Bulgarian and Macedonian, including their non-standard varieties or, in other words, their dialects. In older literature, Bulgarian was used to refer to both languages, for two main reasons: First, the Macedonian standard language was codified only after World War II; prior to that, references to the Macedonian language were rare, although the term "Macedonian dialects" was frequent,

indicating the geographic location of the varieties (on the codification of the Macedonian standard language, see Friedman 1998). The other reason is less self-explanatory. The geographical area of Macedonia was fought over after the collapse of the Ottoman Empire, with claims made on the region by Bulgaria, Serbia, and Greece. When the destiny of Macedonia as a constituent republic of the Socialist Federal Republic of Yugoslavia was finally settled in 1945, these old claims persisted in the socialist People's Republic of Bulgaria, taking the form of opposition to the Macedonian ethnic identity and language. This attitude was still very much alive in the 1990s, yet today, most scholars of the region and its languages do not deny the existence of the Macedonian language.

However, the dialects of Bulgarian and Macedonian form a dialectal continuum, meaning that no clear delimiting isoglosses exist, but the features gradually change (although, in Macedonian dialectological tradition, isogloss-based definitions do exist; see **Footnote 31**). Besides this fact, it is often claimed that the languages are "mutually intelligible." I have personally witnessed encounters between speakers of Macedonian and Bulgarian unaccustomed to such situations, where there has been a great deal of difficulty in the exchange of thoughts, despite a certain basic level of comprehension. In addition, mutual intelligibility is explained only partly by the "closeness" of the varieties and more by, for instance, the sociolinguistic settings (see, e.g., Gooskens & Hilton 2013). An oft-observed phenomenon is the limited mutual intelligibility between Finnish and Estonian. At least historically, the mutual intelligibility was often realized in an asymmetrical fashion, since the Estonians were much more knowledgeable about Finnish than *vice versa*. For several generations of Estonian speakers, the situation was explained by previous exposure to the language, such as Finnish television broadcasts, which were followed by those on the northern coast of Estonia during the Cold War (Verschik 2012, 274–275).

In addition to the names of the two languages, I frequently use the term Balkan Slavic in this study to denote both Macedonian and Bulgarian. This is not done solely to overcome the problems of assigning the dialects a status of pertaining to one or the other language, but also to address issues that involve both languages in a briefer manner. However, the way I use the term Balkan Slavic differs somewhat from what is perhaps its most common use, since in these pages, Balkan Slavic denotes Bulgarian and Macedonian only. Similar to the dialectal border between Bulgarian and Macedonian, no clear demarcation exists between Balkan Slavic and Serbian, although

northern dialects of Serbian are separated from Bulgarian and Macedonian by significantly larger differences than those found between the Bulgarian and Macedonian standard languages. Bosnian, Croatian, Montenegrin, and Serbian (abbreviated BCMS) also form a closely related set of languages. They were previously united by a uniform standard language, Serbo-Croatian or Croato-Serbian, and they share certain features considered to belong to the shared characteristics of the Balkan linguistic area (see **3.1.1**). The southernmost Serbian dialects, especially the Torlak dialects, show an even higher concentration of these features, including an advanced loss of case inflection by comparison with standard Serbian; consequently, the term Balkan Slavic is sometimes used for these dialects as well. These dialects, which can be seen as transitional between Balkan Slavic and Serbian, are largely omitted from the analyses here, owing to the need to limit the scope of the study. Investigating them vis-à-vis the case system is a task I intend to undertake at a later date as a follow-up to the present study.

1.3.2 Case

The definitions of grammatical case are usually dependent on two axes: The formal axis involves the status of the case marker as, for example, an affix attached to the head, and the functional axis describes the type of grammatical properties it expresses. Consider Barry Blake's (2001, 1) definition:

> Case is a system of marking dependent nouns for the type of relationship they bear to their heads. Traditionally the term refers to inflectional marking, and, typically, case marks the relationship of a noun to a verb at the clause level or of a noun to preposition, postposition or another noun at the phrase level.

Blake acknowledges the first axis by stating that, traditionally, expressing case involves inflection, meaning that the case marker is not a clitic, let alone an independent word. The second axis, the marker's functions, is reckoned with by noting that case marks dependent nouns, either on the clause or the phrase level. In other words, case typically does not mark adverbial expressions, much less NP-internal properties such as gender, number, or definiteness. The prototypical case that emerges from this kind of definition coincides well with the Slavic case, which is expressed by suffixes and chiefly marks dependent nouns (see below for the vocative). This is, of course, no surprise, as the common perception of case is Eurocentric and stems from the grammar traditions of Greek and Latin.

Although most general definitions of case operate only with affixes or other types of inflection (see, e.g., Comrie 1986; Malchukov & Spencer 2009), in the study of Balkan Slavic the concept of case is expanded time and again to include prepositions (see, e.g., the discussion of Civ'jan 1965 and Stölting 1970 in **3.4**) and linking articles, characteristic of Albanian and Balkan Romance languages (see **3.3.3** and Ljubenova 2001). In my view, this is often the result of authors not being willing to use functional terminology: It may seem more straightforward to talk about "datival functions" rather than recipients and benefactives, that is, semantic roles. Then, when the morphological dative is observed, let us say, diachronically, there is a certain temptation to refer to the analytical structures that replaced the dative as the use of the "prepositional dative case." The problems with these approaches are enormous, but perhaps the most obvious is distributional: If we observe, for instance, the preposition *na* in Balkan Slavic, which often replaced the dative, its distribution is very different from the original dative, which coincided only in certain functions. Synchronically, unless we call all prepositions cases, the only reason to call *na* a case marker would then be that a few of its functions are similar to the ancient dative. In addition, as Andrew Spencer (2008, 36) reminds us, a preposition can be described in terms of its distribution, whereas accounting for a morphological case often requires explaining why the form is different in the plural or with words belonging to a different declension paradigm, and why the form may sometimes be homonymous with another case form.[2]

There is, nevertheless, another reason why the notion of case is sometimes extended beyond inflection, which poses some challenges to this study as well. While the noun is the focus here, it is obvious that other nominals, like adjectives, pronouns, and numerals also express case. In Slavic, the adjectives display what Blake (2001, 7) calls the concordial use of case, where, as modifiers, they agree with their heads. More importantly, in the modern Balkan Slavic varieties, some personal pronouns still preserve some case distinctions. However, it is reasonable to ask whether the case forms of pronouns pertain to case inflection at all. With the Balkan personal pronouns often no separate case marker can be distinguished, the paradigms being suppletive and the forms largely lexicalized. In addition, pronouns, as

[2] It is well known, of course, that cases are often grammaticalized from adpositions and adverbials; see, e.g., Kulikov 2009. However, this fact poses problems for the definition of case mainly in more agglutinative languages (see Spencer 2008) or in languages that display two separate sets of case suffixes, characterized by different levels of bonding on the affix–clitic cline; see, e.g., **3.1.2** for the Romani case system.

a small and closed word class, often distinguish between a different number of cases than do nouns (see, e.g., Anderson 1985, 201). That the pronouns distinguish cases does not mean that the same categories exist with the nouns and thus entitle the noun system to be addressed in terms of grammatical case if such distinction can no longer be observed. The differences between the pronouns and other nominals will be discussed further in **3.3.2** and **3.3.3**.[3]

For purposes of this study, I settle for a definition of case, similar to Blake's: Case is an inflectional strategy to mark dependency. In Balkan Slavic, that is, in Bulgarian and Macedonian, case was lost and replaced by another system to mark the same grammatical relations. Yet this definition leaves out some problems, a few of which are still relevant for this study and must therefore be addressed below regarding the vocatice case and case syncretism.

Is the vocative a case?

The status of the vocative as a case is disputed by many. Typically, those who exclude the vocative from the category of case justify it with the observation that it does not mark grammatical relations and thererby does not act as a dependent; instead, it is used parenthetically as a form of address (for a summary of these views, see Kottum 1983, 135–137). In addition, Barry Blake (2001, 8) reminds us that modified nouns also occur as forms of address in languages without case inflection. Although the vocative is largely omitted from the analyses in this study, this is, nevertheless, not because of any theoretical *a priori* hypothesis. I try to show in what follows to which extent the vocative is relevant to the study of the Balkan Slavic case inflection.

In the traditional grammars of classical Greek and Latin, the vocative is listed as a case and presented together with other cases in an equal manner. In Indo-European languages, the vocative forms that stem from Proto-Indo-European (PIE) operate on a structurally similar level to the other cases; for instance, in Slavic, they may contribute to stem-final consonantal changes like other case forms (e.g. OCS *vlьkъ*[nom.]–*vlьcě*[loc.]–*vlьče*[voc.] 'wolf'). However, the PIE vocative, limited to the singular only, had no ending, unlike other cases, including the nominative, and therefore the distinct

[3] The relatively small role of pronouns in this study is the result of limitations of time and space, and an acknowledgment that they perhaps deserve a separate inquiry; it is not that their destiny is uninteresting from the point of view of loss of case inflection.

vocative forms in the daughter languages typically reflect either the stem vowel of a particular declension or are later innovations (Beekes 2011, 186–187). In fact, a similar strategy for forming vocatives is observed in spoken Russian, where the vocative is created by omission: *Saša*[nom.]–*Saš*[voc.].

The Balkan languages that are the focus of **Chapter 3** all use, with the exception of Albanian,[4] dedicated vocatives to some extent, including Balkan Slavic (discussed in **Chapter 2**), which mostly abandoned all other case inflection. Whenever such vocatives exist, however, they are often replaceable by the nominative and, occasionally, their use may be seen as dialectal, archaic, or even vulgar (see, e.g., Feuillet 1996, 149). While the Balkan Slavic and Greek vocatives largely continue the PIE vocative, Romanian differs from them in part with regard to the source of the forms. The feminine ending *-o* is borrowed from Slavic (cf. Ro. *fată*[nom.]–*fato*[voc.] 'girl' with OCS *žena–ženo* 'woman, wife') and the masculine ending *-e* seems to continue the Latin vocative, but its survival is probably also supported by the Slavic masculine vocative with *-e* (Dimitrescu et al. 1978, 209–212). There is also a vocative form that combines with the definite article (*domn-ul-e* 'sir!'). Interestingly, for plural reference, the form used is the definite dative–genitive plural.

Although often deemed irrelevant for syntax, the vocative case cannot be overlooked either in this respect because of its particular relationship with the nominative case, as was demonstrated above. At least in Indo-European languages, there seems to be significant dependency between the vocative and nominative, the default agentive case, although the dependence is asymmetrical, as the vocative is often replaced with the nominative, but not the other way around. Yet examples from the Balkan languages betray a more complex picture. In the South Slavic languages there are several male proper names, usually hypocorisms, that is, shorter or diminutive forms of words, whose nominative form derives from the vocative. In Serbian epic poetry, personal names sometimes appear in the vocative instead of the expected nominative, which may be partly due, however, to the requirements of the decasyllabic meter.[5] In addition, in Romanian some names of professions have two forms in the nominative, the second derived

[4] There is some ambiguity here, as the vocative is sometimes mentioned in descriptions of Albanian. However, no examples other than nominatives, both definite and indefinite, are given. Instead, Albanian has a number of preposed particles that are obligatory in indicating a vocative function (Hetzer 1978, 145).
[5] Jussi Nuorluoto, personal communication.

from the vocative: *meşter–meştre* ‘master’ (Dimitrescu et al. 1978, 203).[6] For these nominative forms to exist, there must have been contexts in which the vocatives also assumed the position of the nominative.[7]

Despite some compelling questions related to the status of the vocative—for instance, whether its retention can be attributed to language contact in the Balkans—it will not be discussed further in this volume. Rather than being a statement about the status of the vocative as a case, this choice is dictated by the necessity of limiting the scope of the study. There are also several issues regarding the vocative that, while contributing to the development of the Balkan case systems only marginally, remain to be studied in greater depth and perhaps also in a wider context. For example, the preference for the nominative or new analogical vocatives often seems to be driven by some unwanted connotations related to the old forms. Here, the question of an inflectional vocative is probably closely connected with the diffusion of the “unceremonious term of address” *bre* (see Joseph 2010, 626–627) and its other variants originating from Greek, which are ubiquitous in the Balkans. Also, the origin of the vocative-derived nominatives is still largely uncharted, and especially the birth of the Romanian plural vocative, originating from the dative–genitive, remains unsatisfactorily explained (for one such attempt, see Dimitrescu et al. 1978, 205).

Case syncretism

The basic notion of syncretism is straightforward: Within a linguistic variety, a single inflected form corresponds to two or more grammatical functions that are distinguished in some other inflected forms. Also, the loss of case inflection can be addressed in terms of syncretism: one by one, all forms inflected for case become syncretic with one another until no case inflection is left.

Most research that explicitly makes use of the term syncretism concentrates on synchronically-observed homonymy between inflected forms, for example, by mapping the semantic overlap among the categories that take part in the syncretism. Nevertheless, as is well known, syncretism

[6] Compare also the Bulgarian dialectal *majstore* (standard Bulgarian *majstor*), the single uninflected form, which is not believed to have originated in the vocative (Mirčev 1978, 164–165).

[7] Steinar Kottum (1983, 140–141) suggests on the basis of Polish dialectal material that some semantic properties of the vocatives offer a pathway to the subject position: The vocative implies a high level of individuation and familiarity, which gives them potential to be used as terms of endearment beyond their limited use as forms of address.

often results from indiscriminate phonological processes that cannot be attributed to dynamics of the similar semantic properties of the forms that become homonymous. Hence, some authors have proposed that syncretism resulting from sound changes should not be called syncretism at all (for discussion, see Baerman, Brown & Corbett 2005, 6). In my view, such syncretism is by no means less important from a functional viewpoint: If the syncretism of two grammatical categories resulting from an independent sound change proves to be stable, then it is indicative of the speakers' tolerance for that particular breach of the form-function correspondence. In fact, there are several ways that an unacceptable syncretism can be avoided: In historical linguistics, "reparatory" analogical change, which restores distinctions that existed previously, is abundantly evident. In addition, an otherwise regular sound change can sometimes be cancelled in a particular context to avoid the impending loss of a particular distinction (see, e.g.,Vidoeski 2005, 109).[8]

There is another relevant distinction between case homonymy and case syncretism which must be included in the definition of syncretism. Syncretism can refer only to intraparadigmatic homonymy. In the majority of Slavic languages (and, in fact, in many other Indo-European languages), the most frequent feminine nominative singular suffix is homonymous with the neuter nominative plural. The case form can be identified in most instances on the basis of the lexeme and without the help of context, because the respective plural and singular forms of that lexeme remain distinct. Yet this is not to say that interparadigmatic homonymy is an insignificant factor in language change, since language learners, for example, may not be able to identify correctly the paradigm to which the lexeme belongs. Confusion about a noun's identity regarding the paradigm to which it belongs is not, however, limited to language learners, which occasionally leads to problems also in the use of the term syncretism. For instance, Proto-Indo-European had several inflectional paradigms (see **2.1**), only some of which survived in the daughter languages. These paradigms in Slavic went through several

[8] Although syncretism may be tolerated, it can lead to changes outside the inflectional paradigm. It is a widely observed fact that the loss of case inflection is accompanied, for instance, by an increase in analytical marking of grammatical relations by adpositions. More complex corollaries have also been suggested: Riho Grünthal (2010) argues that the extensive homonymy between Estonian core cases led to an increase in the use of Germanic-styled phrasal verbs, available as a model through continuous contact with Low German. Although not corresponding exactly to any of the lost meanings, phrasal verbs often imply the telicity of the action, also expressed by the genitive–accusative, one of the case distinctions lost as a result of sound changes. See also **Section 4.3** in this volume.

changes, characterized, on the one hand, by the spread of certain endings to other paradigms and, on the other hand, by some nouns becoming inflected according to another paradigm. Synchronically, these changes would mean difficulty in identifying the paradigm to which a noun belonged, and, consequently, deciding, whether to treat the observed case homonymy in terms of syncretism.

In the present study, themes related to syncretism are ubiquitous. However, study of syncretism does not offer a particular unified framework that would help in the analyses. Perhaps with the exception of formal morphology, which is not the approach of choice here, it could be asked whether such frameworks exist at all. Nevertheless, the research dedicated to semantic mapping of case functions is a promising field (for an overview, see Narrog & Ito 2007), although this approach is not employed here either, at least as such. In my view, the biggest obstacle to the application of a more formal framework to the question of case loss raised in the present study is variation, observed on at least three different levels: dialectally, cross-linguistically, and diachronically. In light of the kinds of observations made in this volume, it is difficult to imagine an informative and illustrative formalization or visualization that would incorporate all these dimensions of the variation without leading to visual clutter or trying to depict something that is more easily put into words.

1.4 Earlier research on the loss of the Balkan Slavic case inflection

As a research topic, the fate of the Balkan Slavic nominal inflection has arisen within two separate, if not altogether independent, research traditions. The Slavists have been intrigued by the Bulgarian and Macedonian development, which diverges so strikingly from other Slavic languages. Meanwhile, researchers into the Balkan linguistic area phenomenon have analyzed the case systems of the Balkan languages, although many of these scholars are also conducting research into the Slavic languages. Yet the loss of case inflection in Balkan Slavic is not an undisputed Balkan feature for the clear reason that the other Balkan languages (or at least a majority of their varieties) do have cases, although not as many as during their earlier stages. This is also true of Romanian and other Balkan Romance languages, most of which have preserved some case inflection, unlike the rest of the Romance languages. However, as a tendency, the general increase in analyticity involving not only nouns is often mentioned in connection with the Balkan linguistic area languages (see **3.3**). In addition, some of the

acknowledged features of the Balkan linguistic area concern nominal syntax and are thus inseparably connected to its expression in Balkan Slavic, both synthetically and analytically.

Research into the development of the Balkan Slavic case system has gone on for more than two centuries. Much of it has aimed at determining the historical processes during the five-century-long "dark period," a phase with only indirect textual evidence, which extended roughly from the 11th to the 16th centuries. Other, synchronic approaches most often seek to compare and contrast the situation in Balkan Slavic with that of the other languages in the Balkan linguistic area or, in some cases, also with a wider variety of languages or general grammatical models. All approaches can be further divided into descriptive work, as in editions of individual manuscripts; chronological endeavors, which are typically the task of historical grammars; and explanatory analyses, which attempt to determine the reasons behind the loss of the case inflection. Nevertheless, all of these approaches are often somehow present in the majority of studies, making a concise and structured introduction to the field challenging.

In this section, I offer two, somewhat different presentations of earlier research and, through critical analysis, attempt to clarify the overall picture: What are the main explanations for the loss of cases? What are the key arguments? Which approaches require more serious scrutiny and which, if any, have already been successfully discarded? The discussion here is limited to more general questions only, and all relevant arguments will be addressed in detail in the subsequent chapters.

1.4.1 Internal and external explanations

One of the scholars of the case loss in Balkan Slavic, Klaus Steinke (1968, 3–34), divides in his monograph the research up to his time according to the models used to explain the loss of cases. He identifies two categories: researchers who support a theory based on a single explanatory factor and those who endorse multifactorial models. According to Steinke's classification, the unifactorial theories, characteristic of the research of the 19th and early 20th centuries, can be further divided into exogene and endogene, depending on whether the loss of case inflection is viewed as resulting from a language-external or language-internal cause, respectively. The exogene theories are often based on an alleged substrate effect on Balkan Slavic. Indeed, the first researcher into the Balkan linguistic area, Jernej Kopitar (1829, 86), viewed the loss of cases, as well as the rise of the

postposed definite article, as a result of a substrate of the Paleo-Balkan language Thracian, mediated by Balkan Romance. Fran Miklošič, an early scholar of Slavic linguistics, supported this view as well, although he assumed a kinship between the Thracians and modern-day Albanians (Steinke 1968, 6).[9]

A Bulgar substrate has also been proposed. Bulgar is a Turkic language spoken historically by a semi-nomadic people that, prior to their conquest of Bessarabia and parts of the Balkans in the late 7th century, inhabited the northern coast of Black Sea. Steinke (ibid., 9) counts Hannes Sköld among the supporters of the Bulgar hypothesis. However, I view Sköld's position as problematic: Although Sköld talks about the historical settings around the time of the First Bulgarian Empire (681–1018) and the role of the Bulgars, he explicitly states that "If I have so emphatically urged this point [the role of the Bulgars], it is not at all because I wish to make myself an advocate of the substratum theory in this particular case" (Sköld 1923, 24). In my view, all theories based on a Paleo-Balkan substrate fail to convince for the simple reason that the first written Slavic documents from the 9th century, which surfaced in copies a century later, exhibit a case system presumably very close to the Common Slavic system. By that time, three centuries after the arrival of the South Slavs in the Balkans, any Paleo-Balkan substrate would have already left its imprint on Balkan Slavic.

Uwe Hinrichs (2004, 235–236) considers the loss of case inflection part of the creolization of the Bulgarian language, a result of communication between the Bulgar and Slavic peoples. His approach is akin to the substrate theories, and its biggest shortcomings are based on a failure to demonstrate how the effects of the supposed creolization could take several centuries to manifest in any way, since, according to one common view, a creole is a pidgin language, or in Hinrichs's terminology, an interlanguage, which becomes the mother tongue of the next generation of speakers (for the definitions of a creole, see Thomason 2001, 159–162). For these theories to work, one should either suppose the existence of demographically significant speaker communities of Paleo-Balkan languages that survived until the second millennium, changing their language to Slavic only then or,

[9] A view that is still debated, with the descent of Albanian from Ancient Illyrian being the opinion that predominates among scholars (Katičić 1976, 184–188). Since the sources on Thracian and Illyrian are scarce, the main arguments revolve around the uncertain identity of the languages as either belonging to the Indo-European Centum or Satem group, Albanian being a member of the latter. For example, Ivan Popović (1960, 79–85) makes the case for the Thracian origin of Albanian based on the assumed Centum identity of Illyrian.

alternatively, posit a dramatic dialectal variation, which would have included a conservative dialect, reflected in Old Church Slavonic (OCS), and a coexisting progressive dialect whose influence would become evident only much later.

Both of these possibilities are improbable in light of the historical and textual evidence. Already in the canonical OCS corpus, there is dialectal variation that reflects differences in morphologic conservatism (Steinke 1968, 116). However, the differences are very subtle and do not demand an explanation based on contact influence. Furthermore, although the Bulgarian toponymy shows signs of direct contact with the speakers of Thracian, their number must have been very few even in the 6th century, owing to their Hellenization and Romanization (see, e.g., Katičić 1976, 136). Again, the Bulgars, whose conquest brought them into contact with the Slavs in the late 7th century and who subsequently formed a ruling elite, were totally "Slavicized" as early as the 9th century (Barford 2001, 93).[10]

Among the exogene theories, Steinke includes the views that consider one of the extant Balkan languages to be the source for the loss of case inflection. One of the most popular and influential of these views considers Balkan Romance responsible not only for the loss of nominal inflection in Balkan Slavic, but also a source for several other Balkanisms. As Steinke (ibid., 11) observes, the influential Slavist, Vatroslav Jagić (1894, 283), for instance, stated: "Ich halte den Verlust der Declination im Bulgarischen nebst dem Auftreten des postpositiven Artikels für ein Resultat der innigen Berührung der Bulgaren mit den Rumänen." In the same vein, the modern scholar Zbigniew Gołąb (1997, 15), who in his article contrasts the development of Balkan Romance and Balkan Slavic nominal declension paradigms, concludes: "…the internal linguistic mechanism acting as a driving force in the whole process of the Balkanization of Macedonian was the linguistic Slavicization of the original Balkan Romance (i.e. Arumanian) population…."

The role of the Greek language in the loss of case inflection has been pointed out by several scholars, Wilhelm Lettenbauer (1953, 162) among them, who stated the following: "Im Bulg. hat die expansiv wirkende griechische Sprache den Anstoß zum endgültigen Schwinden der synthetischen Deklination gegeben." As the last exogene approach, Steinke

[10] Interestingly, a parallel development took place two centuries later when the Scandinavian Varangians, who had founded Kievan Rus', were assimilated by the East Slavs.

lumps together the views regarding the source as an unspecified Balkan influence. According to Steinke (1968, 16–17), supporters of this view include Václav Vondrák, Antoine Meillet, and Johann Schröpfer. The influence of individual languages on the development of the Balkan Slavic case inflection will be discussed in more detail in **Chapter 3**. I argue in **3.1.1** and **3.4** that any single language cannot be found responsible for the majority of the shared linguistic features in the Balkans and the loss of case inflection.

Of the endogene arguments, the most influential, often called the theory of phonetic leveling, is based on the Balkan Slavic sound changes: the "confusion of the nasal vowels" and the merger of /y/ into /i/. Whereas many scholars in the field evaluate these mergers as a source of increased syncretism, according to Steinke (1968, 18–23), Pëtr Lavrov, Antoni Kalina, Ljubomir Miletič, and André Vaillant have been the chief proponents of this view.[11] The role of the sound changes will be discussed in **2.3**.

The first monograph dedicated entirely to the question, "Der Untergang der Deklination im Bulgarischen," by Karl Meyer (1920) was based on what Steinke calls an endogene syntactic argument: Meyer (ibid., 18–31), quite justifiably, refuted the earlier substrate hypotheses as indecisive and chronologically skewed. He was a skeptic of the Balkan influence as well, because in his view, Balkan Slavic lacks the tell-tale signs of an intensive language contact, namely, a significant number of loan words. While he acknowledged the significance of the accusative–nominative homonymy, increased by the partially merging nasal vowels and the fronting of /y/, these are not sufficient by themselves to explain the loss. Perhaps somewhat surprisingly, Meyer ended up claiming that, in fact, no specific explanation is needed, since the majority of European languages have undergone a similar process.[12] According to him, the process of case loss itself is characterized by an increased use of prepositional constructions, further thinning the case system. In addition to misunderstanding the nature and potential of language contact, Meyer's greatest shortcoming, as Steinke (1968, 24) points out, is his inability to explain why a similar initial situation

[11] Note that this theory has been attributed to Fran Miklošič, too, see, e.g., Ivanova-Mirčeva & Haralampiev 1999, 174.

[12] This could be seen as one of the first formulations of the idea that instead of asking why Balkan Slavic is so different from other Slavic languages, we should ask why other Slavic languages are so un-European (see, e.g., Aronson 2007, 31).

with the same syntactic preconditions did not lead to the loss of cases in the remaining Slavic languages.[13]

As the last of the unifactorial views, Steinke (1968, 24–25) presents some "undefined" endogene arguments by Bulgarian scholars Stefan Mladenov and Ljubomir Andrejčin. Both oppose suggestions of external influence on the Bulgarian language, the loss of cases being no exception. They propose ideas that, according to Steinke, should be viewed rather as pertaining to the philosophy of language, with Andrejčin referring to the loss of cases as part of the system of the Bulgarian language "reaching perfection" through general developments in the language. My view is that Andrejčin especially has often projected a sense of national pride into his discussions, which reveals regrettable albeit, especially in layman's thinking, not unusual linguistic attitudes: Changes through language contact are perceived as less prestigious than processes that can be explained autochthonously.

In addition, I suspect that these authors might also exhibit another common prejudice: not infrequently is the loss of morphologic inflection viewed as the "dilapidation" of the language. This attitude is documented, for example, by Benjo Conev (1984 [1919]b, 457), who comforted his fellow Bulgarians, who were "feeling bad when we compare our language of today with the old and the other related Slavic languages where there are still many case forms."[14] An example of an original solution to the feeling of inferiority caused by the lack of cases is Aleksandăr Teodorov-Balan's claim that Modern Bulgarian does have cases: in order to justify his view, he completely redefines the concept of grammatical case and claims that case is, in fact, a grammatical role that may or may not realize as a morphological case (on Teodorov-Balan's theory of cases, see Ljubenova 2001). However, while Teodorov-Balan's theories are not taken seriously in modern Bulgarian scholarship, they seem to represent, together with Conev, a reaction to the idea that equates (one interpretation of) linguistic complexity and the sophistication of the speakers, touched upon in **3.1.2**.

Finally, Steinke presents three plurifactorial models, of which the first is found in Benjo Conev's exhaustive, fifteen-hundread-page *Istorija na bălgarskija ezik* (History of the Bulgarian language, Conev 1984 [1919]a; 1984 [1919]b; 1985 [1919]). Conev (1984 [1919]b, 455–456) believes that the reasons leading to the loss of cases in Bulgarian are basically the same in

[13] Meyer's monograph was met with fierce and sometimes unjust criticism, especially from Miletič (1925). Steinke points out that Meyer's merits in the inclusion of the syntactic level in the analysis were not acknowledged until Duridanov (1956).
[14] All translations in this volume are the author's unless stated otherwise.

all modern languages that have lost case inflection. These reasons, he says, are both general and specific. The general ones include (1) the merger of sounds, (2) the analogical movement of words between declensional paradigms leading to fewer paradigms, and (3) synthetic change, for example, the merger of the genitive and the dative. The specific reasons include (1) external influence, acquired in foreign cultural contacts (as potential contacts for the Bulgarians, Conev mentions the Thraco-Illyrians, Goths, Bulgars, Cumans, Pechenegs, Avars, Greeks, and Turks), and (2) internal causes, mainly national or ethnic characteristics.[15]

Kiril Mirčev expresses his views on the loss of cases in his historical grammar of Bulgarian (1978), first published in 1958. According to him, given the contemporary level of knowledge of the Bulgarian language, a definite answer cannot be given. However, Mirčev (ibid., 292–294), who shows thorough familiarity with the previous research, summarizes the key reasons for the loss of case inflection as being the sound changes leading to increased homonymy, yet significantly, this process would have left the language as it was, namely, in a state in which the cases persisted, but with much syncretism were it not for the influence of the "environment." The vagueness of the reference to the language contacts is explained by Mirčev's disillusionment with what is clearly in his view, the nebulous field of language contacts (see, e.g., ibid., 65–66). Steinke (1968, 29) argues that Mirčev's observations about the dialectal variation of OCS displaying different levels of morphosyntactic conservatism was one of his most significant later contributions.[16]

Ivan Duridanov's monograph on the loss of case inflection (1956) is the second after Meyer's to deal with the question. Steinke (1968, 30) quotes passage from Duridanov (1956, 77): Its main point seems to be that internal

[15] Conev contrasts the Bulgarian *Volksseele* with the rest of the Slavs, stating that, as Bulgarian political life three decades (1880-1910) prior to his books had demonstrated, the Bulgarians, in contrast to their Slavic brothers, are parsimonious, hardworking, more pensive than joyful, and, most important, receptive, accommodating, and adaptable to everything new, all of which, in Conev's view, must be taken into account when assessing the morphologic changes in Bulgarian.

[16] Steinke (1968, 29) considers this a change of heart by Mirčev on the basis of a later article on early Balkanisms in OCS (Mirčev 1966), whereas in his historical grammar, first published 1958, Mirčev had stated, "It is unacceptable to assume that this dialect [the Salonika dialect, i.e., that of OCS] differed significantly from the rest of the Bulgarian dialects of the 9th century." This passage, however, remains unchanged in a later edition of the grammar, which was revised by the author himself (Mirčev 1978, 294). I believe that the passage is intended as a criticism of claims dating the loss of case inflection very early, either because of a substrate hypothesis or for some other reason. In his article in 1966, the dialectal differences Mirčev observed are relatively subtle.

contradictions within a language must be viewed as central to language change; however, one cannot underestimate the significance of external contradictions, which may sometimes become decisive. In my view, Steinke is not entirely fair to Duridanov in introducing this quote. Duridanov's work is defined by two very different approaches: On the one hand, he conducted what was at the time the most thorough and insightful analysis of the written evidence on the loss of case inflection that had ever been done. Yet on the other, the first part of his study pertains to the framework of dialectical materialism and Soviet linguistics that arose from the particular historical period and which was imposed on Bulgarian scholars as well, which had the unfortunate effect that Duridanov's views on the reasons behind the loss of case inflection remained unclear. However, as Steinke (1968, 30) observes, in his conclusion Duridanov urges researchers to study the chronology and nature of the phenomenon rather than hypothesize about its reasons, a challenge that Duridanov himself gladly accepted, thereby setting a solid example.

In his own monograph, Steinke (1968) conducts a survey of three apostle manuscripts from the 12th and 13th centuries and the variation of the case uses therein, with the idea of establishing a more precise chronology of the case loss. Steinke follows Duridanov's advice and does not participate in a discussion about the causes of the loss of cases except to state that studies must be conducted without prejudice, and the time for assessing the causes comes, if ever, only after determining the fundamentals of the phenomenon (ibid., 31).

Steinke's comprehensive overview provides a useful picture of the development of thinking in the field until the 1970s. However, the sources he cites are by no means equal in weight: Many of the opinions are based on researchers' overall view of the Balkan linguistic area phenomenon and its causes. The loss of case inflection, if mentioned, is included in the list of Balkanisms, and an explanation for the whole phenomenon devised to account for the case loss as well. These speculations, sometimes more philosophical than linguistic, are by no means comparable to the original research done by Meyer (1920) and Duridanov (1956), for example. Yet it is precisely the field of contact linguistics that has advanced the most since the 1960s, and contact linguistic approaches, following in the footsteps of Meillet and Schröpfer, for instance, who are only touched upon by Steinke, have gained more weight over the last decades (see **3.1**). As a result, some of the terminology in the overview requires attention: As Steinke (1968, 11)

himself notes, Miklošič views the Thracian substrate as affecting Balkan Slavic through contacts with Balkan Romance. Therefore, Miklošič's explanation is based as much on a substrate as on an adstrate effect. In the same vein, if in Jagić's view Albanian is a descendent of Thracian, there is no break-up of the continuum of the language, and the adstrate effect of Albanian would also be expected.[17]

1.4.2 Phonetic, syntactic, and semantic explanations

A different classification of earlier research is given in a more recent article by Andrej Sobolev (1991). First, he presents a bibliography on the theme and provides a list of sixty manuscripts between the 11th and 16th centuries, which is a representative corpus for the study of the Balkan Slavic case loss. Second, he presents what is perhaps the most detailed theory on the loss of case inflection, based on an intriguing synthesis of the previous approaches. In addition, his introduction to the field uses a different division than Steinke's. Sobolev groups the approaches into phonetic, syntactic, and semantic. Instead of a supposed ultimate cause, the division is based on the level of grammar that has been at the center of researchers' attention.

In Sobolev's division, Miletič is the key figure in phonetic theory: Sobolev (1991, 8) cites Samuil Bernštejn's (1948, 361) criticism that Russian too exhibits considerable syncretism in its case system without having reduced the number of cases. However, Bernštejn's further observations on the phonetic theory are interesting: He sees the Bulgarian linguists Conev, Miletič, Mladenov, and Mirčev as all being supporters of essentially the same view, although Mirčev's alleged stance is based on his very early article. For Bernštejn, the phonetic theory is basically part of the Bulgarian linguistic tradition, and it represents the reluctance to accept foreign influence as an impetus for change.[18]

As in Steinke's overview, the proponents of the syntactic approach include Meyer. Nevertheless, Sobolev (1991, 9) mentions another standpoint in this connection: Whereas Meyer emphasizes the role of the prepositions, Gennadij Tiraspol'skij (1980) highlights the invention of the definite article

[17] In fact, Gołąb's views, unlike Miklošič's and Jagić's, are based purely on the assumption of a substrate mechanism affecting Macedonian. Note, however, that many authors use the term "substrate" differently: namely, to describe the sociolinguistically less prestigious language and its effect, not necessarily as an assumption of its disappearance.

[18] It must be kept in mind that, as emerges from Steinke's overview, it is the non-Bulgarians Pëtr Lavrov and Antoni Kalina who must be seen as the first proponents of phonetic theory in the 1890s.

as a cause for the loss of case inflection (this view is discussed further in **4.3**). For Sobolev, Miklošič, Kopitar, and many others belong among the proponents of a syntactic approach. This is because of the popularity of the idea that language contact could bring about an increase in analytism. In fact, Sobolev calls this the most popular contemporary view in Slavic studies and Balkan linguistics. He criticizes these views for their inability to account satisfactorily for the spread of the oblique case (for a definition, see **2.2.2**) to new prepositional constructions that would replace other cases (Sobolev 1991, 10).

In his third, semantic approach, Sobolev (1991, 9) surprisingly includes Conev. It is not clear what exactly in Conev's views should be regarded as semantic, unless it is perhaps what Conev called synthetic change. As with the syntactic approaches, the semantic approaches too are very different one from another. Here, Sobolev mentions Herbert Galton's (1967) article on the "evolution of Bulgarian syntax," a synchronic study attempting to find macro-level reasons for the analytical expression of nominal syntax and yet a large array of finite verbal forms. Galton's partly very confusing work explores the semantic properties of the preposition *na*, for example, and reaches the conclusion that, by comparison with other Slavic languages, Balkan Slavic represents a "concrete" way of thinking, the cases and the infinitive representing the abstract in language and the prepositions, the finite verb forms, and the definite article representing the concrete. Sobolev dismisses both approaches—Conev's for representing "national romantic psychologism" and Galton's for resorting only to hindsight reasoning and not describing the dynamics of the change. Somewhat similar to Galton, Jørn Ivar Qvonje (1979) explores the semantic properties of the mergers of the expression of location–goal and dative–genitive in the Balkan languages. While criticizing Qvonje for not presenting the causes leading to the merger of goal and location and not justifying the alleged connectedness of the two mergers convincingly enough, Sobolev (1991, 10) nevertheless praises Qvonje's proposition that perhaps one should first consider the merger of semantics and only afterwards the form of the dative and genitive.

Steinke and Sobolev's differing overviews illustrate the common problem in these types of endeavors: To obtain a general picture, one needs to simplify the situation to a point that does not always do justice to the individual researchers. It is quite clear that the many proponents who Steinke and Sobolev have grouped in the same category would fiercely oppose each other's positions and be more accepting of the ideas of

someone from another group. For example, there are very few scholars who would entirely oppose the role of language contacts in the loss of case inflection, although some of the contact explanations are mutually exclusive. Nor does anyone deny that the sound mergers also contributed to the process, although they were not necessarily its ultimate cause.

Sobolev himself presents a theory that answers the main criticisms. He begins with the outcome of the development, the oblique case, and attempts to explain the contexts and conditions that enabled its development. He acknowledges the roles of both language contacts and sound changes, but sets clear limits on their influence. According to Sobolev (1991, 12), language contact can only eliminate oppositions—for instance, yielding the syncretism of goal and location—but it cannot create new, competing structures.[19] Therefore, the source for the spread of the oblique case must be sought in the historical corpus of Balkan Slavic within the prepositional structures that commanded only one case, different from the accusative (which also served, together with the often homonymous nominative, as a source for the oblique case). In fact, as many others have done, Sobolev too evades the question of the decisive reasons for the loss of case inflection.

1.4.3 Historical grammars, manuscript studies, and synchronic analyses

Although a number of important contributions to the study of case loss in Balkan Slavic were discussed in the two preceding subsections, here I would like to present some further studies representing approaches that are essential to understanding the field. A number of historical grammars, in addition to those of Conev, Mladenov, and Mirčev, have addressed the question of the loss of cases in Balkan Slavic. Many of these merely make reference to the studies already mentioned. However, an insightful, balanced, and dispassionate overview in regard to the role of internal change, dialectal variation, and external influence can be found in Blaže Koneski's *Istorija na makedonskiot jazik* (History of the Macedonian language, Koneski 1965, 130–142). One of the questions raised by Koneski is the role of clitic doubling of indirect objects in contributing to the weakening of the dative case (clitic doubling is discussed in **3.3.2**).

Many studies of Balkan Slavic manuscripts have been conducted in order to chart the particular case system they reflect. These monographs include

[19] Since this part of Sobolev's argument will not be discussed further in this volume, it must be pointed out that he is much mistaken in his assertion; see, e.g., Johanson 2009 for numerous counterexamples regarding grammatical case.

Samuil Bernštejn's (1948) study on the *Wallachian Letters* of the 14th and15th centuries, Jerzy Rusek's (1964) work on *Hludov Triodion* from the 13th century, Elena Češko's (1970) monograph on the history of Bulgarian declension, a study based on several manuscripts, and Horst Kämmerer's (1977) analysis of the noun inflection in the 14th and 15th century copies of the translations by Evtimij of Tărnovo. A more recent study is Ute Šin's (1996) survey of the case system of the *Chronicle of Manasses* from the 14th century. While Rusek and Češko's approaches are similar to Steinke's, Kämmerer and Šin do not, in fact, contribute directly to the study of the loss of case inflection, since they most often observe changes that are characteristic of other Slavic languages as well (see, e.g., Šin 1996, 240–241). Similarly, Češko devotes a good deal of space to the analysis of the loss of dative absolute, a participial construction that did not survive in the other Slavic languages either.

The synchronic situation in Balkan Slavic has been addressed, among other scholars, by Antoaneta Popova (2008) and, from the Balkan perspective, by Tat'jana Civ'jan (1965). Certain subsystems within the case inflection have also attracted the attention of scholars of both the Slavic and the Balkan languages. The dative–genitive merger, as it is most often called, has been studied from the Balkan Slavic perspective by Angelina Minčeva (1964) and from the Balkan perspective, but with an emphasis on Slavic by Jørn Ivar Qvonje (1979; 1980) and Nicholas Catasso (2011). Other synchronic approaches include the overviews of the Balkan linguistic area phenomenon, such as those by Helmut Schaller (1975), Georg Solta (1980), and Petja Asenova (2002).

***

The loss of case inflection in Balkan Slavic has attracted a great deal of attention over a long period of time. The complexity of the phenomenon is appreciated by the majority of scholars of the area, and perhaps understandably, many have refused to contribute to the contest of suggesting its ultimate causes, directing their attention instead to more restricted problems. It is, however, safe to say that the overwhelming consensus in the field seems to be that the effect of a Paleo-Balkan substrate, including the effect of Bulgar, is a very unlikely source for the changes. And even in the case that these languages acted as substrata, their effect would be impossible to evaluate convincingly for purposes of a scholarly argument. It seems equally clear that both internal and external factors do have a role in the

change. This division is also reflected in the structure of the present work, in which separate chapters are dedicated to each approach (**Chapters 2** and **3**).

2 The diachrony of the Balkan Slavic case system

In this chapter, I examine the diachronic development of the Balkan Slavic case system. In the first of the four sections below, in **Section 2.1**, I introduce the early developments of the Slavic case system. This presentation serves as a starting point for analyzing the subsequent changes in the case inflection and the eventual case loss. In **Section 2.2**, I conduct a meta-analysis of the previous research by evaluating the proposed theories and analyses of the changes in the case system. Most of the research deals with texts from the Late Mediaeval era dating between the 12th and 16th centuries; these Middle Balkan Slavic (MBS) manuscripts were created after the Old Church Slavonic (OCS) period (the mid 10th – the 11th centuries), but before the first texts appeared in the vernacular. It is characteristic of the Middle Balkan Slavic texts to use a highly archaizing grammar, which therefore does not reflect the actual situation in the spoken language. In **Section 2.3** I provide an analysis of the role of sound changes in the loss of case inflection in Balkan Slavic. Finally in **Section 2.4**, I summarize the findings of this chapter and propose an interim MBS case system.

2.1 From the Proto-Indo-European to the Slavic case system

Some of the developments in the Proto-Slavic and OCS case systems cannot be properly understood without a short introduction to the relationship between the Proto-Indo-European (PIE) and the Proto-Slavic case systems. During the post-Proto-Slavic dialectal period, certain features of the case system continued to evolve, with slightly different outcomes in the individual Slavic languages. I will address these divergent developments in **2.1.2**.

It is often said that the Proto-Slavic case system preserved most of the features of the Proto-Indo-European system. Nevertheless, exactly what characteristics of the PIE case system Proto-Slavic preserved must be defined in greater detail. The main characteristics of both systems are three grammatical genders, three numbers, and shared cases, that is, nominative, accusative, genitive, dative, locative and instrumental, and the vocative. [20] The nominative–accusative, genitive–locative, and dative–instrumental syncretism of the Proto-Slavic dual is inherited from the PIE case system. PIE also had an ablative case, which displays some syncretism with the

[20] The status of the vocative as a case was discussed in **1.3.2**.

genitive. In Proto-Slavic, the ablative merged with genitive, which is visible in the genitive singular of the *ŏ*-stem masculines, where **-ā* is derived from the PIE ablative instead of from the genitive, as in the other declensions (Matasović 2008, 180–181).[21] The old functions of the ablative are visible in the use of the Slavic prepositions of separation *iz*, *otъ*, and *sъ*, which all govern the genitive.

Unlike in Proto-Slavic, in early PIE the morphological distinctions between the grammatical genders of nouns were not yet based on formal distinctions. In fact, based on the fact that Hittite lacks the feminine gender, it is believed that the three grammatical genders were a late development in PIE. According to one popular theory, the origins of the gender distinction lie in early PIE, which was an ergative language, marking prototypically animate agents of transitive actions with an ergative case (Beekes 2011, 215). This would explain why the nominative and accusative of neuters are always homophonic and non-marked, including in Slavic, and why the masculines originally differed from the neuters in the nominative, but not in the accusative case: The absolutive case of the ergative system, marking the patients of transitive verbs and the agents of intransitive verbs, is continued by the syncretic neuter nominative and accusative, whereas the nominative of the masculines reflects the ergative case, which marked the agents of transitive verbs. Thus the grammatical masculine gender with its distinctive nominative form developed from words that were more likely to appear as agents and be therefore used in the ergative case, such as animates. (Ibid.)

2.1.1 The Proto-Slavic case system

The subsequent developments in Proto-Slavic included an increase in the fusional characteristics of the case system, as the number and case often became morphologically inseparable. One development that further contributed to the fusional type of Proto-Slavic nominal morphology was the reanalysis of some of the PIE nominal stems as parts of the case markers, leading to more variation in the case desinences (Schenker 1996, 106). Also, the number of declensions was in decline, which is apparent as early as in the OCS period, thus favoring the alignment of the natural gender with the three dominant inflectional paradigms, the *ŏ*-stem declension for the

[21] In addition, it has been suggested that the genitive singular of masculine/neuter pronouns (e.g., OCS demonstrative m.sg. pronoun *togo*) reflects the ablative ending, as in PIE *tād and the particle -go (Arumaa 1985, 175). However, in most accounts the form is thought to derive from the bare Proto-Slavic stem **to-* and the particle **-go* (see, e.g., Matasović 2008, 229).

masculines and neuters and the *ā*- and *ĭ*-stem declensions for the feminines (see, e.g., Halla-aho 2006a, 66–76). **Table 1** illustrates the most common declension paradigms in OCS.

		'slave'	'man'	'son'	'woman'	'earth, land'	'bone'
		***ŏ*-stem**	***jŏ*-stem**	***ŭ*-stem**	***ā*-stem**	***jā*-stem**	***ĭ*-stem**
Singular	**nom.**	*rab-ъ*	*mǫž-ь*	*syn-ъ*	*žen-a*	*zemĺ-a*	*kost-ь*
	acc.	*rab-ъ*	*mǫž-ь*	*syn-ъ*	*žen-ǫ*	*zemĺ-ǫ*	*kost-ь*
	gen.	*rab-a*	*mǫž-a*	*syn-u*	*žen-y*	*zemĺ-ę*	*kost-i*
	dat.	*rab-u*	*mǫž-u*	*syn-ovi*	*žen-ě*	*zemĺ-i*	*kost-i*
	loc.	*rab-ě*	*mǫž-i*	*syn-u*	*žen-ě*	*zemĺ-i*	*kost-i*
	instr.	*rab-omь*	*mǫž-emь*	*syn-ъmь*	*žen-ojǫ*	*zemĺ-ejǫ*	*kost-ьjǫ*

		***ŏ*-stem**	***jŏ*-stem**	***ŭ*-stem**	***ā*-stem**	***jā*-stem**	***ĭ*-stem**
Plural	**nom.**	*rab-i*	*mǫž-i*	*syn-ove*	*žen-y*	*zemĺ-ę*	*kost-i*
	acc.	*rab-y*	*mǫž-ę*	*syn-y*	*žen-y*	*zemĺ-ę*	*kost-i*
	gen.	*rab-ъ*	*mǫž-ь*	*syn-ovъ*	*žen-ъ*	*zemĺ-ь*	*kost-ьji*
	dat.	*rab-omъ*	*mǫž-emъ*	*syn-omъ*	*žen-amъ*	*zemĺ-amъ*	*kost-ьmъ*
	loc.	*rab-ěhъ*	*mǫž-ihъ*	*syn-ъhъ*	*žen-ahъ*	*zemĺ-ahъ*	*kost-ьhъ*
	instr.	*rab-y*	*mǫž-i*	*syn-ъmi*	*žen-ami*	*zemĺ-ami*	*kost-ьmi*

Table 1. The most common declension paradigms in OCS, including the partly reconstructed *ŭ*-stem declension. Note the soft-stem variants (*jŏ*- and *jā*-stem) of the I and II declinations. For the relationship between Late Proto-Slavic and OCS, see 2.2.1.

The adjective in early PIE was not formally separable from the noun unless it had a derivational suffix. Nor did adjectives exhibit grammatical gender, which emerged only later through suffixation. (Schenker 1996, 219–221.) The most commonly found stem type in PIE adjectives was the *ŏ*-stem, and in Late Proto-Slavic it was this declension that all masculine and neuter adjectives followed. The *ŏ*-stem declension's role as the "non-feminine" inflectional paradigm was further strengthened in Late Proto-Slavic, since feminine adjectives begun to employ the forms of the *ā*-stem declension only.

PIE pronouns differed from nouns partly because of different case desinences. As in PIE, Proto-Slavic pronouns can be divided into two groups: non-gendered (first- and second-person personal pronouns, reflexive pronouns, interrogatives "who" and "what") and gendered (demonstrative, most interrogative, and quantitative pronouns etc.). The demonstrative

pronoun *jь*[22] spread the pronominal case endings to adjectives, after it started to appear as an article-like suffix attached to the adjective, often indicating definiteness. This development, shared with Baltic, bears a resemblance to the Germanic strong adjectival inflection and had the effect that in most modern Slavic varieties, both derived and non-derived adjectives are also a morphologically distinctive word class. Of the numerals, one and two follow the pronominal declension, whereas three and four follow two different noun paradigms. Proto-Slavic numerals are declinable, number two being declined in dual, and from one to four they are gendered. The remaining cardinals, with the exception of 11–14,100, and 1000 follow the *ĭ*-stem declension.

The resulting Proto-Slavic case system was fusional, characterized by several inflectional paradigms into which words were increasingly aligned according to their gender, a process that further cemented the formal distinction between the grammatical genders. In addition, as a result of sound changes and other developments, there was homonymy and syncretism between different case endings, perhaps most notably between the genitive singular of the *ŭ*-stem and the dative singular of *ŏ*-stem, and between the nominative singular of the *ŭ*- and *ŏ*-stem masculines and the genitive plural in the majority of the paradigms (for a thorough discussion on the origin of the *ŏ*-stem masculine nominative singular, see Halla-aho 2006b, 111–141).[23] Extension between the declensions worked in two ways: on the one hand, nouns of rarer paradigms shifted to more dominant paradigms, and, on the other hand, individual case endings spread to foreign paradigms. The most notable instances of case syncretism were either inherited, such as between the nominative and accusative of neuters and the three-case system of the dual, or they resulted from early Proto-Slavic developments, such as the syncretism between the nominative and accusative of the *ŭ*- and *ŏ*-stems.

2.1.2 The dialectal developments of the case system

Like the sound system of Proto-Slavic, its case system too underwent many more or less common changes during its later dialectal stage, for instance,

[22] Not attested independently in the nominative case, except for the first part of the relative pronoun *jь-že*.

[23] Note that the homonymy was not necessarily total in all instances, since there may have been differences in the accentuation.

the loss of the dual.[24] A considerable part of the research into the history of the Balkan Slavic case morphology deals with observations of phenomena that are also present in many other varieties of Slavic. A brief summary of these phenomena will highlight the areas of the Balkan Slavic historical case system in which, compared with other Slavic languages, changes would be anticipated, regardless of the eventual loss of case inflection.

Although losing ground as an independent declension, the *ŭ*-stem case forms served as building blocks for new declensions in the individual Slavic languages emerging from the Proto-Slavic (for an overview of this development, see Janda 1996). In all Slavic languages, the *ŭ*-stem genitive plural *-ovъ* spread to other stems, most significantly, to the *ŏ*-stem, whose genitive plural had become homonymous with the masculine nominative singular.[25] In South Slavic languages, the plural marker for a great number of monosyllabic masculines developed from the *ŭ*-stem nominative plural *-ove*. In addition, West Slavic employs to various degrees the *ŭ*-stem desinences for distinctive animate and inanimate masculine declension paradigms.

A semantic criterion that affects the choice of case in all Slavic languages is animacy, which belongs to a set of phenomena known in typological literature as the effect of referential scales.[26] For the modern languages, a rough generalization is that animate singular masculines employ the genitive when other nouns appear in the accusative. The feature is also present in certain modern Balkan Slavic dialects that retain some case inflection. This use of case, traditionally referred to as the genitive–accusative, is evident, though not fully developed in OCS: The singular *ŏ*-stem nouns denoting humans sometimes, although not always, take the genitive ending as objects of transitive verbs (for some semantic criteria, see Lunt 2001, 56):

[24] However, the Upper and Lower Sorbian and Slovenian still preserve the dual. For retention of archaisms on the periphery of a dialectal continuum, see subsections **2.1.2**, 2.2.3: *Dative with prepositions*, and **4.1.1**.
[25] However, in BCMS (Bosnian, Croatian, Montenegrin, Serbian) Štokavian dialects the genitive plural was later replaced by *-ā* (for perhaps the most credible theory of its origin, see Matasović 2008, 185–186).
[26] Here, the term “referential scales” has been adopted following Bickel & Witzlack-Makarevich 2008. These phenomena have been dealt with under several headings, including person, participant, topic, empathy, referential, nominal, indexability, agency and animacy hierarchy; see, e.g., Nichols 2001, 516. These themes are discussed further with regard to Balkan Slavic in **Chapters 3** and **4**.

(1) Codex Marianus,[27] Luke 7:3

da	*prišedъ*	*sъpasitъ*	*rab-a*	*jego*
that	coming	saves	servant-M.GEN.SG	his

'that [Jesus] comes and saves his servant'

But:

(2) Codex Marianus, Mark 12:2

posъla	*kъ*	*tęžatel̑-emъ*	*rab-ъ*
sent	PREP	tenants-M.DAT.PL	servant-M.ACC.SG

'[He] sent a servant to the tenants'

In OCS, attributive pronouns and adjectives agree with the genitive–accusative by appearing in the genitive, but with nouns, only one genitive allomorph occurs in this function. The only *ŭ*-stem noun in the OCS corpus denoting a human, *synъ* 'son', accepts only the *ŏ*-stem genitive singular ending *-a* as a direct object. Also, an originally *ĭ*-stem *gospodь* 'lord' takes *-a* as a direct object. For both non-*ŏ*-stem nouns, the original genitives *-u* and *-i* are also attested, but never in the function of a direct object.

Eastern Slavic also developed a differential marking for the accusative plural: Similar to the genitive–accusative in masculine singular, it extended the differential use of the genitive to mark the plural animates in the accusative as well (for an excellent diachronic analysis, see Klenin 1987). This development is also shared by West Slavic Polish and Slovak but there it is limited to masculines with a human referent. South Slavic, with the exception of Bulgarian and Macedonian, maintained the nominative–accusative distinction in the masculine plural. After the earlier accusative had combined with the nominative, a new accusative was created through the adoption of a differential ending, *-e*, limited originally to a subgroup of the *ŏ*-stem nouns with a palatalized stem-final consonant, the *jŏ*-stem. However, unlike in North Slavic (i.e., West and East Slavic), in South Slavic the new differential accusative covers all masculine plural nouns, not only those denoting animate beings. Another South Slavic feature of the plural case forms, limited to the BCMS Štokavian dialects (including the standard languages), is the syncretism of the dative, locative, and instrumental cases, the common form originating from the dative–instrumental dual desinences.

[27] From this point on, the OCS text passages are Latinized and normalized in traditional terms (see, e.g., Schenker 1996). Later texts are Latinized, but elements, not represented by the orthography, are not reconstructed.

Also, the same dialects display an almost complete syncretism between the dative and locative in the singular.

2.2 The case system during the Middle Balkan Slavic period

In this section, the outcome of the loss of case inflection, as reflected in the 17th-century vernacular-based Damascene literature and the modern dialects, is contrasted with the earlier written sources. This is done primarily through a meta-analysis of earlier research, whose main features were presented in **Section 1.4**. For an earlier point of reference, I use the OCS parts of the annotated PROIEL[28] corpus, which is based on the *Corpus Cyrillo-Methodianum Helsingiense*.[29]

Most of the research into the loss of case inflection in Balkan Slavic is done on the basis of texts created after the canonical OCS period, but before the appearance of the vernacular-based literature in the 17th century. Variation in the use of case forms is observed not only between various translations, but also between different copies of the same texts. Especially valued are the texts between the 11th and 13th centuries, because of an increase in uniformity and the conscious attempts to archaize the literary norm in the 14th century. The primary method of most studies is to find uses of cases that differ from the OCS norm. While this research is often conducted in the most diligent manner, it typically concentrates on instances relevant only to the scholar’s own theory, a fact that also calls for comparison with the earlier research.

The alternative hypotheses for the variation between OCS and these manuscripts are the following: 1) the interference of the spoken language, that is, that the non-normative use of cases is evidence of similar changes in the spoken language; 2) the scribes’ inability to use the cases correctly, owing to the grammatical distance between the written norm and the spoken language, which led to indiscriminate vacillation or “mistakes.” There is also a third possibility: 3) an emerging new literary norm. These hypotheses are often found in the previous literature, although usually they are merely implied. I seek to compare the types of variation that have been observed by scholars and discuss their interpretations. In line with the hypotheses presented above, the main questions are which aspect of the variation can be interpreted as directly indicative of the developments in the spoken language

[28] http://foni.uio.no:3000
[29] http://www.helsinki.fi/slaavilaiset/ccmh/

of the time, and, on the basis of this, what can be said about the case system of the spoken language.

To evaluate the written historical data, a synchronic point of comparison must be established through the analysis of dialectal data. Yet studies that deal with the remaining case inflection in the Balkan Slavic dialects are unfortunately very few and plagued by certain problems. Although the descriptions of individual dialects often include remnants of the case inflection if these exist in the dialect, these studies usually lack a deeper analysis of the syntax, and the examples, often consisting of individual words taken out of their context, are presented only from a morphological point of view. A classical study, discussing the vestiges of the case inflection in Balkan Slavic is the "Old declension in the Bulgarian dialects of today" by Ljubomir Miletič (1890). The article introduces an overwhelming list with hundreds of examples of inflected forms, allegedly still in use at the time it was published. However, according to Samuil Bernštejn (1948, 359), along the inflected nouns Miletič gives a number of petrified forms that no longer represented inflection. Bernštejn warns further that Miletič does this with the intention of proving that the loss of case inflection was not due to external influences, but rather the result of internal processes in the Bulgarian language. Fortunately, Stojko Stojkov (1968) presents a modern and concise account of the remaining case inflections, albeit limited to the dialects spoken in Bulgaria. Nevertheless, Stojkov's article includes the Rhodopian dialects, which display the largest amount of retained case inflection among the Balkan Slavic varieties.

The structure adopted by most previous research addresses the development of one OCS or Proto-Slavic case at a time, and this is the chosen approach here too. After an analysis of the historical written sources in **2.2.1**, the development of the Proto-Slavic case system is observed by first establishing the synchronic status of the case, that is, whether it is still used dialectally, and, second, the development of the expression of the functions that the case expressed previously. To give an account of these processes poses certain challenges, since the destiny of the individual historical cases offers a logical starting point, but the modern phenomena usually involve the forms and functions of several historical cases. Note, for example, that the genitive–accusative is treated under *Accusative* and that some of the questions relevant here are addressed in **Section 2.3**, which discusses the role of sound changes in the loss of case inflection.

2.2.1 On the written sources

The historical sources that inform us about the development of the Balkan Slavic noun phrase can be divided into three groups. The first group comprises the canonical Old Church Slavonic corpus (called Old Bulgarian in the Bulgarian and sometimes also in the German tradition). These manuscripts originated between the second half of the 10th century and the end of the 11th century. A second group of texts was written after the OCS period and up until late medieval times, that is, during the 12th to the 16th centuries, while the third set of texts comes from the 17th century on. This division coincides roughly with the periodization of the history of the Bulgarian language into Old Bulgarian, Middle Bulgarian and Modern Bulgarian.[30] Here, based on the discussion in **1.3.1**, the term Middle Balkan Slavic (MBS) is used to indicate that the analysis also involves manuscripts regarded as Macedonian (on these manuscripts, see Koneski 1983, 4–5). Like the term Middle Bulgarian, MBS is used to denote both the writing produced during the 12th to the 16th centuries and the non-attested spoken language of the same era, whose greater understanding is also the object of this study. Therefore, the MBS literature coincides in time, but not in its features with spoken MBS.

The end of the Old Church Slavonic period is characterized by developments in phonology. The two jers, Late Proto-Slavic short high vowels **ŭ* and **ĭ*, were elided in "weak" positions and preserved in "strong" positions (counting from the end of the phonological word, the last jer was weak, the jer before a weak jer was strong, and the jer before a strong jer or any other vowel in the following syllable was weak): the word **dĭnĭsĭ* 'today' would eliminate the first and the last jer, preserving the middle, as reflected in the modern Bulgarian cognate *dnes*. While the canonical OCS texts do sometimes show this change, these phonological changes among

[30] For the periodization, see, e.g., Feuillet 1999, 11–15. An alternative, two-part division is proposed in Georgiev 1952. A more detailed periodization of OCS is discussed in Schaeken & Birnbaum 1999, 13–21. However, a note must be added regarding their objection to the use of the term *Altmakedonisch*. They argue that, unlike the term *Altbulgarisch*, which is used synonymously with OCS, the use of *Altmakedonisch* in reference to those OCS manuscripts deemed Macedonian is not advisable, because there was no Macedonian state at the time. However, the term *Altbulgarisch* was not coined on the basis of the First Bulgarian Empire, rather it is a projection of the name of the modern language into the past, when the language was called simply Slavic, in OCS *slověnьskъjь* (*językъ*). It is true that the use of both names in an analogous way is detrimental, as *Altbulgarisch* is typically used as an umbrella term, see also **1.3.1**. A similar problem is the use of Old Russian in the sense of Old East Slavic. In this volume, the term Balkan Slavic is preferred for the historical linguistic varieties after the OCS period, partly to overcome the problems outlined here.

others are found increasingly in later texts where the letters <ъ> *back jer* and <ь> *front jer* are often either omitted in a weak position, preserved in a strong (non-weak) position, or their distribution does not correspond to what is expected in the norm.[31] Nevertheless, the Church Slavonic tradition persisted and preserved the main characteristics of OCS, surviving in new recensions of Serbian and Russian origin until the 19th century as the literary and ecclesiastical language of the Orthodox Slavs of the eastern parts of the South Slavic dialectal area. Thus, the so-called Damascenes, the first texts that clearly reflect the actual developments in the vernacular, emerged in the early 17th century. The Damascenes form a separate and unparalleled phase in the history of Balkan Slavic literature.

The division of the texts into three groups is needed mainly to assign them a status as evidence for the spoken language. The OCS period must be seen to a great extent as representing the actual spoken language at the time of the creation of the first texts, at least in its phonology and morphology.[32] Despite the famous testimony of the monk Hrabr (see Schenker 1996, 173) from the end of the 9th century, suggesting that the Slavs made use of letters to render their speech in writing before the invention of the Slavic scripts, Glagolitic and Cyrillic, without prior written evidence it is unwise to presuppose a tradition that preserved much older archaisms. Thus, the first translation of ecclesiastical texts in Slavic during the Moravian mission, conducted by Cyril and Methodius starting in the year 862, could not have been based on a language much different from their own, namely, a South Slavic dialect from the Thessaloniki area. However, precisely because of these dialectal features, which distance the language from the other Slavic languages, OCS is not a written manifestation of Late Proto-Slavic, although OCS amounts to the closest available textual source. On the other hand, the syntactic evaluation of the OCS texts is much more difficult, since the great majority of texts are Biblical translations of Greek origin, often in a word-for-word correspondence with the Greek original, as example **(3)** shows:

[31] In fact, in the Macedonian tradition, the emergence of the different reflexes of the back jer in a strong position constitutes a watershed after which the development of the Bulgarian and Macedonian languages are treated separately; ъ is rendered *o* in Macedonian and *ă* in Bulgarian, while ь is realized as *e* in both languages (Koneski 1983, 5–6).

[32] Nevertheless, the first preserved manuscripts date from almost a century after the Moravian mission of Cyril and Methodius, meaning that some caution must be observed not to overplay the unity of the spoken and written languages of the time: see, e.g., Nuorluoto 2012b, 37.

(3) A comparison between the *Codex Marianus* and the Greek Bible,[33] Mark 1:4.

bystъ	*joanъ*		*krъstę*	*vъ*		*pustyni*
was.AOR.IND.3SG	John.NOM.SG		baptize.PRS.PTCP.NOM.SG	in		wilderness.LOC.SG
egéneto	*Iōánnēs*	*ho*	*baptízōn*	*en*	*tê̂*	*erḗmō*
came.AOR.IND.3SG	John.NOM.SG	DEF	baptize.PRS.PTCP.NOM.SG	in	DEF	wilderness.DAT.SG

i	*propovědaję*	*krъštenьje*	*pokajaniju*	*vъ*	*otъpuštenьje*	*grěhomъ*
and	proclaim.PRS.PTCP.NOM.SG	baptism.ACC.SG	repentance.DAT.SG	in	forgiveness.ACC.SG	sin.DAT.PL
kaì	*kērýssōn*	*báptisma*	*metanoías*	*eis*	*áphesin*	*hamartiô̂n*
and	proclaim.PRS.PTCP.NOM.SG	baptism.ACC.SG	repentance.GEN.SG	in	forgiveness.ACC.SG	sin.GEN.PL

'John was baptizing in the wilderness, proclaiming a baptism about repentance for the forgiveness of sins'.

The second group of texts, the Middle Balkan Slavic corpus, differs from OCS in that its scribes and authors spoke a language that was increasingly different from the language they wrote. While some of the features reflecting the changes in the spoken language, such as the loss of weak jers and variation in the use of the two nasal vowel graphemes <ѫ> *back nasal* and <ѧ> *front nasal*, are evident, the morphosyntactic reality of the vernacular is not represented in the texts to the same extent. The rise of the Second Bulgarian Empire in the 13th century meant a development toward a more uniform written style. In the second half of the 14th century, Evtimij of Tărnovo, who later became the Patriarch of Bulgaria, initiated orthographic and stylistic standardization. The reforms of the so-called Tărnovo School were not aimed at modernizing the language, however, but rather at cementing the basic features of OCS. The majority of the manuscripts preserved from this period are also translations of Greek texts or copies of older OCS manuscripts.

A special group of texts in the MBS corpus consists of the so-called *Wallachian Letters*, the correspondence between the rulers of the Danubian Principalities, situated in today's Romania and Moldova. This correspondence, the earliest of which was written in the 14th century, continues the Balkan Slavic ecclesiastical literary tradition stemming from OCS. Yet these texts display many features that betray the contemporary spoken language, including some changes in the case inflection. However, the study of the *Wallachian Letters* is complicated by the fact that it is not known with certainty to what degree the scribes spoke Slavic as their mother

[33] The annotated examples from the Greek Bible are from Bible hub, an online parallel bible project, found at http://biblehub.com; unless stated otherwise, the English translations rely on the *International Standard Edition*.

tongue. Also, features are often found that point to Serbian influence on the language in the letters. (Bernštejn 1948, 15–16, 128–129.)

In the end of the 14th century, the Ottoman conquest of the Balkans and the resulting collapse of the Second Bulgarian Empire lessened the literary activities among the Slavs of the Eastern Balkans. The third group of texts consists of a type of literature that begun to be popular in the area at the end of the 16th century. The Damascene literature, named after Damaskinos Stouditis (died ca. 1580), a Greek clergyman, consists of collections of various kinds of stories, mainly but not exclusively religious, translated mostly from Greek (Božkov et al. 1963, 405–407). However, this genre employed vernacular-based language, including in the Slavic translations. Like the OCS texts, this literature can be seen as more direct evidence of spoken language. While still containing many characteristics of the old tradition, the grammar is clearly very different from the earlier tradition, to the degree that according to Roger Gyllin (1991, 106), the Damascenes, whose translators were characteristically meticulous and linguistically conscious, must be taken into account when dating the beginning of the modern Bulgarian literary language. The following example from the 17th century *Tihonravov Damascene* shows, among other things, the use of the uninflected form similar to the old nominative in the feminine singular and the postposed definite article:

(4) *Tihonravov Damascene* (Dëmina 1985b, 322)

i	*toj*	*go*	*ostavi*	*da*	*mu*
and	he	him	let.AOR.3SG	COMP	him.DAT

služi	*na*	*trapeza*	*ta*	*njegova*
serve.PRS.3SG	at	table.F.SG	DEF.F.SG	his.F.SG

‘and he[1] let him[2] to serve him[1] at his[1] table’

***

A short excursus is needed here to the so-called *Cserged Prayers*, a set of texts from modern-day Romania like the *Wallachian Letters*. The *Cserged Prayers* consist of Protestant Lutheran hymns and other texts translated from German and contained in three manuscripts dating from the early 19th century. These texts are often taken at face value as representing a linguistic situation more than 400 years earlier than is justifiable to assume, which may lead to detrimental results in analyzing the loss of case inflection in Balkan Slavic. Sometimes, only on the basis of the *Cserged Prayers*, the rise

of the definite article and the greatly diminished case inflection in Balkan Slavic are dated to the 12th and 13th centuries.

The Slavic inhabitants of a village called Bolgárcserged in Hungarian, or Cergău Mic in Romanian, were the descendants of Bulgarians who were forced to move from northwestern Bulgaria to Transylvania in the 13th century (Miletič 1987 [1896–1900], 101–111). Ljubomir Miletič, who studied the Cserged Bulgarians intensively, changed his view of the reasons for the departure of the villagers. First, he proposed that they must have belonged to the sect of Bogomils, deemed heretic by the Orthodox Church, and were therefore forced to flee Bulgaria (ibid.). Later, he was inclined to believe instead that they were taken as slaves between 1260–1266, during Stephen V of Hungary's campaign against the Bulgarians (Miletič 1926, 7). Because for centuries the villagers lived in isolation from other Bulgarian speakers, it has been thought that their language, as evidenced by the 19th century manuscripts, must represent the linguistic situation in Bulgaria at the time of the villagers' departure.

In his critical edition of some of the texts, Fran Miklosič (Miklosich 1856) states that their language stands closer to OCS than to Modern Bulgarian. Samuil Bernštejn (1948, 233) goes as far as to treat the language of these texts as if they represented 13th-century spoken Bulgarian, a view shared by Gunnar Svane (1961, 235). According to Svane, parts of the Cserged Prayers were based on liturgical texts that date back to the 16th century. However, according to Miletič (1987 [1896–1900], 116), whose edition Svane used, the translations were based, at the earliest, on a hymnal first published in 1680. Therefore, we should assume a minimum of 400 years between the translations and the era whose linguistic situation they are said to represent. It is true that the language of *Cserged Prayers* displays some archaic features, most notably the retention of distinctive reflexes for the Proto-Slavic nasal vowels. However, this is not unforeseen in the Balkan Slavic dialects, whether it be the geographically-isolated Rhodopian dialects or the Kostur-Korča dialect found in the southwestern periphery of Balkan Slavic. In the case of Bolgárcserged, the village is both isolated and situated in a peripheral location.

Sometimes the parallel development of related linguistic varieties which are separated geographically is called a drift. To the degree that the language of the Cserged Bulgarians of the 17th and 18th centuries shared similar innovations that in other Balkan Slavic varieties are attested only after the separation, a drift in this case may be the result of processes that begun

before the varieties lost contact (on drift as a reflection of variation in the proto-language, see Joseph 2006).[34] Here, the differential accusatives for animates in Slavic, discussed in **2.1.2**, might serve as an example of a similar chain of events. In addition, Brian Joseph (2013, 45) argues that the emergence of a definite article from a demonstrative during the attested history of West Germanic need not to be accounted for by dialectal variation in Proto-West-Germanic, owing to the commonness of such a development. However, if we assume that the definite article developed at least partly through language contacts, then this criterion is also met by the Cserged Bulgarians, since all of their contact languages had a definite article. In addition, the situation of Istro-Romanian, a Balkan Romance language spoken in today's Croatia discussed in **3.2.2**, shows that isolation from the speakers of a related variety may be insignificant for the retention or loss of case inflection: Of the two varieties of Istro-Romanian, one preserves case inflection, whereas the other does not (Zegrean 2012, 29).

A more suitable place for the *Cserged Prayers* in the history of Balkan Slavic literature is alongside the Damascenes, or perhaps even more appropriately, the Slavic evangeliaries of the 18th and 19th centuries from modern Northern Greece, written in the vernacular and with Greek letters (for two such evangeliaries, see Lindstedt, Spasov & Nuorluoto [eds.] 2008; Mazon & Vaillant 1938). This is not to say that the *Cserged Prayers* are insignificant for the study of case inflection (see, e.g., **2.2.3**). Nevertheless, they cannot be used as evidence for Middle Balkan Slavic.

2.2.2 Nominative, accusative, and genitive

The destiny of the nominative, accusative, and genitive cases is intertwined and therefore treated under the same heading. The Proto-Slavic nominatives served as the source for the new, uninflected nominal form of Balkan Slavic, but it is sometimes thought that the accusative too competed with the nominative to express this function (see, e.g., Miletič 1890, 234), although the evidence for this involves only a limited number of modern dialects. Some of the forms of the Proto-Slavic genitive were preserved dialectally, although not in the original function, but as the accusative or the oblique

[34] For a slightly different use of the term "drift" in linguistics, see the famous definition in Sapir 1979 [1921], 147–170. Interestingly, Edward Sapir (ibid., 163–164) also treated case loss as a particular drift, shared by a number of Indo-European languages.

case.[35] The destiny of the accusative is more complex, although the problems involve only a limited number of modern dialects.

Nominative

The nominative was generalized as the single, uninflected nominal form, with one exception introduced in this subsection in connection with the accusative.[36] In those dialects of Balkan Slavic that still display some case inflection, the historical nominative also serves as the unmarked case, except for one form, introduced in example **(10)**. There are several interesting questions regarding the development of the uninflected form: For example, there is dialectal variation in the gender of some former feminines, ending in a consonant, some of which became analyzed as masculines. Also, as was mentioned in **2.1.2**, a significant number of monosyllabic masculines developed a plural ending from the *ŭ*-stem nominative plural *-ove*, this too being subject to great dialectal variation. However, since this study concentrates on the loss of case inflection and the old nominative was not lost, the use of the nominative is discussed only in connection with other cases when it replaces the expected non-nominative case form.

Accusative

The Proto-Slavic accusative did not survive in Balkan Slavic, with the exception of dialectally attested *ā*-stem singular nouns (Stojkov 1968, 32–34). One type of this inflected accusative is limited to certain kinship terms:

[35] Oblique is somewhat problematic as a term. Most often it refers to the marked, non-subject case of a two-case system. Since the unmarked case of a two-case system is characterized by its only role as the case of the subject (and, in Balkan Slavic, the predicative) and the marked case or oblique is characterized simply by its complementary distribution with the unmarked case. Therefore, in purely synchronic terms, it would be misleading to call the marked case the accusative. Other terms have also been used for Balkan Slavic: In the Bulgarian descriptive tradition the term *obšta forma*, the general form, is used (see, e.g., Mirčev 1978, 289). In addition, Stojkov (1968, 28) uses the term *Agglomerativ*. To complicate things, in some modern Balkan Slavic varieties there is a third inflected case, the dative. For a three-case system, the traditional terms for the cases are more justifiable. The terminological solution adopted here is the following: The oblique case is used only to refer to the synchronic case form that became the marked case of a (mainly) two-case system, especially in the sense of becoming the default case that prepositions take, even though there may be some vestiges of a distinct dative case. However, in glosses of dialectal data in the present chapter and in describing the development of the Balkan case systems in **Chapters 3** and **4**, the historical terms such as genitive–accusative are used. The single uninflected form, on the other hand, is not a case, since a language can have either two or more cases or, alternatively, none at all.

[36] However, a number of nominatives originate from the Proto-Slavic non-nominative cases, but these forms are common to other Slavic languages as well, cf. Mac. *smokva* 'fig' < Late Proto-Slavic **smokъv-*[non-nominative stem], not from **smoky*[nom.].

(5) A Balkan[37] dialect of the Veliko Tărnovo region (Stojkov 1968, 32–33).

na	*bašt-ə́*	*mi*	*bašt-á*	*mu*	*umr'a*
PREP	father-ACC	I.DAT	father-NOM	he-DAT	died

'The father of my father died'[38]

In this type of accusative inflection, the case is differentiated only when stressed. There are indications that this pattern represents the earlier situation in a great number of Balkan Slavic dialects. This is supported mainly by 17th century Damascene literature (see, e.g., the *Tihonravov Damascene* in Dëmina 1985b and the *Damascene of Trojan* in Ivanova 1967), where, if the stress falls on the syllable bearing the case ending in an *ā*-stem noun in an oblique position, the ending is often written with the back jer <ъ>. In these manuscripts, *ə* may be written with <a>, but an *a* that falls under stress is never written with <ъ>, and <ъ> must therefore represent *ə*. The language in the *Damascene of Trojan* is based on the northeastern Bulgarian dialects (Ivanova 1967, 6), where words of the type *žena* 'woman, wife' are always stressed on the final syllable (Stojkov [ed.] 1966, Map 118):

(6) *Damascene of Trojan* (Ivanova 1967, 19)

Ti	*e*	*dalъ*	*bogъ*	***žen-ъ***	*da*
you.DAT	is	given	God	woman-ACC	COMP

raždašь	*děca*	*ně*	*zarad*	*da*	*kurvišь*
give.birth.2SG	children	not	so.that	COMP	you.fornicate.2SG

'God has given you a woman for you to procreate not to fornicate'

It is probable that type of inflection in examples **(5)** and **(6)** previously had a significantly wider distribution in those dialects that display distinctive reflexes of **a* and **ǫ* only in stressed positions. In most dialects, this accusative was later replaced by the earlier nominative, and by analogy, the case inflection ceased to exist. Judging by the fact that this differential accusative is today limited only to kinship terms, the use of the accusative probably ceased first with the nouns denoting inanimates or non-humans. This differential accusative shows a split along a referential scale, discussed in more detail in **3.3.4**, and an impetus to the retention of the form may have been given by the typical structure of an NP with a kinship term as the head. Kinship terms are treated in a special way in Balkan Slavic, as the familial relationships are expressed adnominally with a postposed dative clitic,

[37] The Balkan dialects of Bulgarian are named after their location in the region of the Balkan Mountains and are not to be confused with other uses of "Balkan."
[38] Glosses and marking of stress are by the author.

indicating whose relationship is in question, as in example **(5)**. Whereas other nouns often appear with the definite article or are accompanied by adjective modifiers or possessive pronouns, all of which agree in gender, case, and number with the head, in this structure, particular to the kinship words, the definite article is most often omitted and possession is expressed through the clitic that does not agree with the head. In addition, adjective modifiers appear less often in this structure. Therefore, it may have been easier to retain the inflection in NPs with a kinship noun as the head, where the problem of case agreement did not need to be resolved with the already uninflected modifiers and the definite article.

Another phenomenon is also connected to the stressed ending of the *ā*-stem nouns: It is the only context where the uninflected nominal form is said to continue a case other than the nominative (see, however, **Footnote 36**). There are dialects in northern and southeastern Bulgaria with no case differentiation in the historical *ā*-stem nouns, but when stressed, the ending is *-ə* instead of *-a*, apparently reflecting the Proto-Slavic accusative ending **-ǫ* (Miletič 1890, 234). For instance, in the Nevrokop (Goče Delčev) dialect, one encounters uninflected forms like *žéna* – *ženə́-ta* 'woman – the woman' (Mirčev 1936, 66). This phenomenon is no doubt related to the even more widespread Bulgarian pronunciation of the feminine nouns with the definite article whose stem ends in a consonant. For example, in the word *radostta* 'the happiness', the stress moves to the final syllable when the definite article *-ta* is added, but the pronunciation varies between [rədosˈta] and [rədosˈtə].

In addition to the stressed endings, there are also less exclusive contexts in which a differential accusative of the historical *ā*-stem nouns occurs. A fully preserved accusative singular of the *ā*-stem nouns is limited to the transitional dialects between Balkan Slavic and Serbian, located in western Bulgaria, southern and southeastern Serbia, and Northern Macedonia (*žena-ta*[nom.] – *ženu-tu*[acc.] 'the woman, wife'), and the Rhodopian dialect of Tihomir (Stojkov 1968, 33–34):

(7) Rhodopian dialect of Tihomir (Stojkov 1968, 33)[39]

a)

ann-á *žén-a* *dójde*
one-NOM.F.SG woman-NOM.F.SG came
'a woman came'

b)

sréšnax *ann-ó* *žén-ʊ*
I.met one-ACC.F.SG woman-ACC.F.SG
'I met a woman'

Both dialects that preserve the accusative always have a reflex of the Proto-Slavic *ǫ realized differently from the nominative ending -*a*, independently of the place of stress. The role of sound changes in the retention of the accusative is discussed further in **2.3**.

During the dialectal stage of the Late Proto-Slavic, the genitive was adopted as the accusative form of the animate masculine singulars (see **2.1.2**). This genitive–accusative form is the most widely retained inflected case form throughout the different varieties of Balkan Slavic, although its use is limited to proper male names and kinship terms. It is not, however, part of the norm of standard Bulgarian, although it was still included in most of the 19th-century grammars (see, e.g., Gruev 1983 [1858]; Bogorov 1986 [1848]). Although it is widely attested in the dialects, the genitive–accusative is increasingly ostracized in Macedonia too, despite its inclusion in the norm of the standard language (see Friedman 2001, 22). In a limited number of dialects, usually coinciding with the full retention of the accusative singular of the *ā*-stem nouns, the adjective and the definite article may also be declined, as the following example shows:

(8) Rhodopian dialect of Tihomir (Stojkov 1968, 31)

na *visók-e-tok* *čilě́k-a*
PREP tall-M.GEN.SG/ACC-DEF.M.GEN/ACC.SG person-M.GEN/ACC.SG
'to/of the tall man'

The definite article -*tok* continues **togo*, the masculine/neuter genitive singular of the demonstrative pronoun **tъ*.

In addition to the most common genitive–accusative ending -*a*, there is another type of desinence that uses the Proto-Slavic consonantal *t*-stem genitive singular ending. This form is used with male personal names ending

[39] Note the penultimate stress on *žena*, unlike in example **(6)**.

in *e* or *o*: *Stojančo-te* (Stojkov 1968, 31). The dialectal spread of these genitive–accusative forms is limited to only a few dialects in southwestern Bulgaria. However, toward the west, a similar form is widespread throughout the dialects: *Goče-ta* (see, e.g., Cvetanovski 2010, 52 for the Western Prespa dialect spoken in Albania). However, *a* in the desinence is not etymological and must be explained by the analogy of the ending -*a*.

Furthermore, there is also a form that, despite its unclear status, must be dealt with here in connection with the genitive–accusative, owing to the similarity of functions. The form corresponds historically to the *ŏ*-stem locative plural (see **Table 1**), which seems to have been analyzed first as the genitive, aided by the similarity between the locative plural and the genitive plural of adjectives. However, this plural case form exists very marginally in a Rhodopian dialect of the Smoljan region, and, according to Stojkov (1968, 34), it is generally limited to pluralized family names or kinship terms indicating an entity such as a household. Yet in Stojkov's account the syntactic relations the form may express is puzzling. He addresses the form together with the *ā*-stem accusative singular and the dative–genitive and maintains that all the forms represent the *Agglomerativ*, the oblique case, a term he contrasts with *Dativ*; nevertheless, he gives the following example in which the form clearly functions to mark the recipient:

(9) A Rhodopian dialect of the Smoljan region (Stojkov 1968, 34)

Pódaj	*Mándofc-**ɛh***	*ájsva*	*písmo*
give.IMP.2SG	Mandofci.PL-?	this	letter

'give this letter to the Mandofci'

This example seems, however, erroneus, since the apparent source of the sentence cited by Stojkov, a study of Momčilovci dialect by Stojko Kabasanov (1952), introduces the same example with the preposition *na*:

(10) The dialect of Momčilovci (Kabasanov 1952, 39)

Pódaj	*na*	*Mándofc-**ɛh***	*ájsva*	*písmo*
give.IMP.2SG	PREP	Mandofci.PL-?	this	letter

'give this letter to the Mandofci'

It turns out that this particular plural form functions, indeed, as the oblique case. This is extremely significant, because it is the only instance where Balkan Slavic has resorted to another case form to maintain the nominative–oblique distinction with plural nouns.

In the Bulgarian standard language, the definite article has two cases, nominative -(*j*)*ăt* and oblique -(*j*)*a*, limited to the masculine singular nouns

and adjectives. In Bulgarian grammars, the oblique definite article is called *kratăk člen* and the nominative article *pălen člen*, meaning the short and the full article, respectively. However, this form does not result from organic linguistic development and is not attested in any of the dialects, but was introduced into prescriptive grammars of Bulgarian in the 19th century, as can be inferred from the wording of Stefan Mladenov (1979, 250–251): “It is an invention of the grammarians that the nominative article is written *-ăt* and the accusative article *-a*.” In the Bulgarian dialects, either the article -(*j*)*ət* or the article -(*j*)*ə* is found, both uninflected for case. However, owing to the normative efforts, today this is a living feature of some formal varieties of modern spoken Bulgarian.

It is widely believed that during the MBS period the accusative case was increasingly used instead of the other non-nominative cases, thus finally becoming the source of the oblique case (see, e.g., Steinke 1968, 109; Duridanov 1956, 155). However, the destiny of the accusative in Balkan Slavic is closely connected with sound changes, discussed in detail in **2.3**. These changes meant that many accusative forms became homonymous with the nominative. Therefore vacillation between these cases is expected in the MBS manuscripts in instances where sound changes rendered the forms homonymous early on. While the distinction still existed in the writing system for the most part, the relationship between the accusative and nominative during the MBS period remains problematic for another reason: There are relatively numerous examples of mixing between the two cases, even with the *ā*-stem singular forms, which must be presumed to have remained at least partly distinct in a great number of dialects until the Modern Balkan Slavic era. It must therefore be asked whether the nominative could have been used as the oblique case as well, even if a distinct accusative form existed.

In Ivan Duridanov’s (1956, 156) material, which consists of Balkan Slavic manuscripts written before the 14th century, the case that dominates as the replacement for the non-nominative or non-accusative cases, that is, when the forms are distinct, is the accusative by 82.3 percent over the nominative. However, when these cases are used erroneously as the subject or the direct object, in other words, to replace each other, then the nominative is observed more often as the object rather than the other way around. Interestingly, Duridanov observes that the *Hludov Parimejnik*, a manuscript originating in Macedonia in the late 13th or early 14th century, shows the most instances of the replacement of the accusative by the

nominative among the texts in his corpus. The manuscript was written in a region where the reflexes of *ǫ and *a eventually merged into *a*. Although the 14th century seems early for this change to have taken place, as we must first assume the change of the *ǫ into ə, and only then into *a* (see Koneski 1983, 44), as will be argued in **2.3**, the accusative singular forms seem to be preserved remarkably well if they have remained distinct from the nominative. Therefore it could be argued, as Duridanov (1956, 156) assumes, that this relatively early vacillation is indeed the result of homonymy between *ǫ and *a, at least in some dialects.

Genitive in the dialects

In OCS, the genitive functioned as the predominant adnominal case with its central function being to express the possessor. In addition, the genitive was used as the direct object with certain transitive verbs, such as *iskati* 'to want, to seek' and *zьrěti* 'to watch', and it could alternate as the direct object of a number of other transitive verbs with the default direct object case, the accusative (see, e.g., Klenin 1987). As the direct object, the genitive expressed partitivity, a function that is demonstrated also by its role as the obligatory case of the object of negated transitive verbs. A number of prepositions also took the genitive. While a few Proto-Slavic genitive forms are preserved dialectally in Balkan Slavic, the genitive–accusative form cannot be used adnominally.[40] However, the genitive forms of the non-gendered pronouns were also preserved in all dialects as the non-clitic oblique case pronouns, although the non-gendered accusative pronouns, as in OCS, were either fully homonymous with the genitive pronouns or had an alternative form, which was homonymous with the genitive pronouns.

[40] The evidence is inconclusive on exactly when the genitive ceased to be used adnominally. While modern dialects do not employ the genitive adnominally, it does occur sporadically in the vernacular-based Damascenes of the 17th century, although this could be due to the influence of Church Slavonic. The case of the *Cserged Prayers*, however, is significant, since they do not continue the Church Slavonic tradition. Miletič (1926, 55) observes a handful of these genitives, two of which appear with the words "God" and "Holy Ghost," which may, however, indicate that the adnominal use of the genitive may have persisted with formulaic, religious expressions. Nevertheless, the examples with a pronoun, like *tiah-to dobro* 'their good', are not indicative of the use of the genitive, but of the genitive form of the personal pronoun having become a possessive adjective, because the definite article becomes attached to the pronoun. The examples Miletič gives of the feminine forms with the ending *-i*, albeit seeming to continue the hard *ā*-stem genitive singular *-*y*, result most likely from an earlier merger of the *-*y* and the soft-stem dative desinence *-*i* into a single dative case, as the use of the dative definite article in an example given by Miletič shows: *nebentski-tui horí* 'of the heavenly host'. See also the discussion in this subsection.

By far the most frequently preserved genitive form of the nouns and adjectives, the genitive–accusative, was already discussed above. There is also a case form that is very limited both in its occurrence in the dialects and in the contexts in which it can be used. In the Tihomir dialect spoken in the Krumovgrad region, there is a form used only with numerals and quantifiers, which reflects the genitive plural of the *ŏ*- and *ā*-stem nouns (Stojkov 1968, 39). While the feminine words are distinguished by the omission of the word-final vowel, the masculines are homonymous with the nominative singular.

It is generally assumed that the source of the non-nominative or non-accusative case, the dative of the dialects, was always the historical dative. Also, as a general tendency observed in the MBS manuscripts, it was the hard-stem endings that were generalized at the expense of the soft endings, not the other way around. The expected *ā*-stem dative singular form is -*e* (or in some dialects, -*ɛ*, see **2.2.3**), reflecting the hard ā-stem dative ending *-*ě*. There is, however, a dialectally preserved *ā*-stem singular dative case ending -*i*, discussed further in **2.2.3**, which is in contradiction with one or other of the two tenets: *čerka*[nom.] – *čerki*[dat.] 'daughter' (Miletič 1890, 256–257). The two alternative sources seem to be either the soft stem dative ending *-*i* or the hard-stem genitive ending *-*y*. Miletič gives also the form *majci*, reflecting the second palatalization of the velars, an expected development, since the results of the second palatalization of velars varies from dialect to dialect. Also, in most varieties of BCMS and Slovenian the dative singular of the hard *ā*-stem nouns is -*i*, which most often appears with the palatalization of the preceding velar (Matasović 2008, 192–193). To account for the dative ending -*i* of the hard-stem nouns, it must be borne in mind that both **i* and **y* were rendered homonymous rather early, and therefore both the dative and the genitive case desinences may have contributed to the form. However, there is yet another explanation that is not in contradiction with either the source of the new datives being in the historical dative or only the hard-stem endings being generalized, namely, that here the ending -*i* may result also from vowel reduction, discussed in **Section 2.3**. However, without more data, the role of vowel reduction remains inconclusive.

This (mainly historical) dialectal variation between the -*e* and -*i* dative forms may imply that the genitive and the dative forms of both the soft- and hard-stem declensions contributed to the dialectal dative forms as a result of the early confusion between the two cases. A hypothesis regarding this

confusion is discussed below when I address the accusative replacing the genitive in the Middle Balkan Slavic manuscripts. This confusion is not, however, noticed in other declensions: For example, the genitive does not generally appear to replace the dative as the marker of recipient. The similarities in the desinences must have therefore played a major role with the *ā*-stem nouns. It must be noted too that the mixed origins of the dative–genitive case forms are more or less the rule in the Balkan languages (see **3.2**).

Genitive replaced by other cases

The type of vacillation expected in the MBS manuscripts on the basis of assumed changes in the spoken language is the replacement of the genitive by the dative in adnominal constructions and by the emerging oblique case in other contexts. Yet in OCS too, the dative could be used as an adnominal case expressing possession (Večerka 1984, 116)[41]; for example, in the *Codex Marianus* the dative appears as the adnominal case in 24 percent of instances instead of the genitive.[42] However, the increase in the use of the adnominal dative in the MBS manuscripts must be seen as reflecting the situation in the spoken language (see, e.g., Steinke 1968, 62–66). The dative gradually became the default adnominal case, as is evident in the dialects that preserve the dative: the genitive–accusative cannot be used to express possession, but the dative can be. This observation seems to be crucial evidence of the steps of grammaticalization of the *na*-construction to mark adnominal possession: First, it replaced the dative, and second, it began to appear in adnominal constructions (for this view, see, e.g., Minčeva 1964, 3). Yet the dative assuredly did not first replace the genitive entirely and then be itself replaced by the prepositional constructions, but the processes must have overlapped. Whether the genitive could still appear adnominally when the *na*-constructions began to be used in the adnominal position cannot be said with certainty. This may have been possible, especially if one accepts the hypothesis I propose in **2.4** about the grammaticalization of the

[41] Since this variation is very often attested in OCS, but is more or less absent in the Slavic languages other than Bulgarian and Macedonian, it is possible that this is a dialectal feature of OCS, not of Proto-Slavic. Perhaps a bit daringly, it could be even asked whether this is in fact a very early Balkanism in OCS. In addition, given the relative novelty of the genitive in Slavic, which is heir both to some forms and some functions of the ablative, it is possible that the possessive function of the genitive was never firmly established in Balkan Slavic before it was lost, and other Slavic languages merely grammaticalized the genitive as the only case in this function.

[42] N = 363; source: the annotations in the PROIEL corpus; see **2.2**.

na-construction into the adnominal position through an additional path, surpassing the recipient use as an intermediary phase.

Klaus Steinke (1968, 69) investigated the instances of variation between the uses of the dative and the genitive in the adnominal position in three apostle manuscripts from the late 12th and early 13th centuries, which thus predate the wave of normativization of the language during the second Bulgarian Empire. One of the specific constructions under Steinke's scrutiny is a translation of the phrase *hē áphesis amartíōn* 'forgiveness of sins'.

(11) Acts 10:43

otdajanie grěhovъ[GEN.M.PL]/grěhomь[DAT.M.PL]/grěhomь[DAT.M.PL]

(12) Rom. 3:25

za otdanie byvъšihъ prěžde grěhovъ[GEN.M.PL]/grěhovъ[GEN.M.PL]/grěsěhь[LOC.M.PL]

In Steinke's corpus, the phrase appears a total of 19 times, 10 of these times with the dative, supposedly showing the increase in the use of the adnominal dative by comparison with OCS. Yet this construction may be ill suited to demonstrate the phenomenon, because we find the same phrase in the *Codex Marianus* (see example **(3)**), where there are three instances with the dative and two with the genitive. However, Steinke's example **(12)** shows an interesting use of case, because the locative plural appears to replace the expected genitive or dative. As Mirčev (1978, 162) observed, this phenomenon is also evident in the *Evangeliary of Dobrejšo* from the early 13th century, no doubt influenced by the similarity between the genitive plural of the pronominal and adjectival endings and the locative plural, as shown by the agreeing adjective with the ending *-ihъ* in the example. There is a possibility that this may indeed represent a transitory case desinence in some dialects, a hypothesis supported by the marginal plural form in the Rhodopian Momčilovci dialect, shown in example **(10)**, a view held, for example, by Karl Meyer (1920, 29).

Since the adnominal genitive was mostly replaced by the dative, the manuscripts' examples in which the genitive is replaced by a case other than the dative are revealing, since they may indicate that the particular genitive form was lost from the spoken language and could not therefore be produced by the scribe. Duridanov (1956, 119) notes as early as in OCS four instances where the genitive has been allegedly replaced by the nominative or

accusative.[43] Of these, the two most compelling show some important tendencies, which would be characteristic of later texts as well:

(13) Codex Suprasliensis

otidošę	*otъ*	*rigeon-a*	*naricajem-yji*
left	PREP	region-GEN	call.PRS.PASS.PART-NOM/ACC

mal-yji	*trъg-ъ*
small-NOM/ACC	marketplace-NOM/ACC

'they left the region called the small marketplace'

Instead of agreeing with *rigeona* in the genitive, the participle *naricajemyji* is given in the nominative/accusative. For Duridanov, this is a clear indication of changes in the spoken language. This may be true, but instead of the loss of the particular genitive form, whose correct use is attested throughout the manuscript, this instance may instead be indicative of syntactic change. The word *naricajemyji* may be interpreted as belonging to the following NP, which, again, is intended to be parenthetical. Since there are several similar examples in the MBS manuscripts with case forms whose disappearance is not expected because of their retention, even in the modern dialects, this type of mistake is most likely not due to the loss of that particular case form.

Duridanov's second example is, nevertheless, much more likely to reflect certain tendencies in the spoken languages. In the *Codex Marianus*, he points to the phrase *lestь bogatьstvije* 'lust for wealth', where the neuter singular noun *bogatьstvije* appears in nominative/accusative instead of in the correct genitive. This mistake, one of the very few in the otherwise meticulously systematic OCS manuscript, may be dismissed simply as scribal error. Nevertheless, in Duridanov's (1956, 126–128) Middle Balkan Slavic corpus, the erroneous use of case with the prepositions *otъ* and *do*, shows two types of nouns appearing most often: the nominative/accusative singular neuters with the ending -(*n*)*ije*, as in *bogatьstvije*¸ and *ā*-stem accusative singular feminines.

As will be discussed in **2.3**, the sound changes operating during the MBS period affected the (*j*)*ā*-stem nouns and adjectives more than any other declension, and therefore, it is credible that this observed vacillation is a sign of the rearrangement of the case system in the spoken language. In light of similar examples in which the (*j*)*ā*-stem accusative singular appears

[43] The term nominative/accusative is used to denote case forms that are homonymous and therefore cannot be assigned the status of being either a nominative or an accusative.

instead of the expected genitive, Duridanov (1956, 136–138) asks, whether there was a transitory possessive structure that functioned through the juxtaposition of the head and the new oblique case.[44] He concludes that this indeed happened, but only in a limited number of dialects. An alternative explanation is that the soft *jā*-stem genitive ending -*ę*, which would have spread through analogy to the hard *ā*-stem nouns too, would have been denoted with the back nasal <ѫ>, which in OCS stands for *ǫ*. Then the use of the accusative ending instead of the genitive would be the result of the so-called confusion of the nasal vowels, in which the two nasal vowels are believed to merged partially (the phenomenon is discussed in more detail in **2.3**). Sobolev (1991, 29) argues against both of these proposals. He claims that while the use of the back nasal is due to the confusion of the nasal vowels, the phenomenon is merely textual without any connection to sound changes in the spoken languages.

However, what Sobolev fails to explain is why the perfectly useful hard-stem genitive singular ending -*y* was replaced. I believe that there is a possible explanation, independent from the interpretation of the confusion of the nasals, why in the MBS manuscripts the *ā*-stem accusative singular form sometimes appears instead of the genitive: After the merger of *y* and *i* (see **2.3**), the *ā*-stem genitive singular form became homonymous with the *jā*-stem dative. Because of the homonymy, the resulting -i_2 ending would be interpreted in several dialects as a dative. Although the dative was increasingly used as the adnominal case, dialectal data shows that it never acquired the genitive's functions as a prepositional case, nor did it inherit the use of the genitive as the direct object. Therefore, the -i_2 ending, interpreted as the dative, could not be used anymore with the prepositions requiring the genitive or as the case of the direct object, and the accusative was used instead to fill the gap. Whether the *ā*-stem accusative singular was used adnominally too, as Duridanov suggests on the basis of the MBS manuscripts, is far from sure. In any case, this transitory use of the accusative could not have survived long, since it was soon replaced by the dative and the prepositional constructions in this function.

Genitive replaced by prepositions

Another typical quest in this field of study has been the search for constructions with *na* to express the adnominal possessor, although it is widely believed that it was the dative as a verbal argument that the

[44] See **4.2.5** for juxtaposition as an adnominal strategy in modern Balkan Slavic.

construction first replaced. Steinke (1968, 74–77) is very well aware of the risk of misinterpretations and chronological problems, since his manuscripts dated from the 13th century at the latest; thus, in introducing some examples of a possible *na* structure replacing an adnominal genitive, he admits outright that there are alternative readings. However, there is one instance where with great certitude he states that "*[u]nter diesen Umständen muß man dieses Beispiel als vollwertigen Ersatz des adnominalen Kasus betrachten, der durch den Einfluß der Volkssprache in das D[enk]m[al] eindrang*."

(14) 1. Cor. 12:8[45]

ov-omu	*dajetь*	*sę*	***duh-omь***	***slov-o***	***na***	***prěmǫdrost-i***
this-DAT.SG	give.3SG	refl.ACC	spirit-INSTR.SG	word-NOM.SG	on	wisdom-LOC.SG

in-omu	*že*	*slov-o*	*razum-ь*
other-DAT.SG	moreover	word-NOM.SG	sensibility-NOM.SG/ACC.SG/GEN.PL

'The word of wisdom is given by the Spirit to one, the word of sensibility to the other…'

The Greek original in the passage reads *lógos sophías*, using the genitive. However, instead of being a clear-cut example of the *na*-construction replacing the adnominal genitive, as Steinke believes, this example is in fact rather complex. Steinke disregards the fact that Modern Bulgarian, too, employs a locative construction in expressing the field of expertise: *toj e dobăr văv volejbola* 'he's good at volleyball' (in BCMS: *on je dobar u odbojci* with the locative case). While older translations of the Bible seem to follow the Greek genitive, as demonstrated by the *King James* "For to one is given by the Spirit the word of wisdom; to another the word of knowledge…," modern translations acknowledge the abstract nature of the compound formed by these words. The Finnish translation from 1992 says: "*Yhden ja saman Hengen voimasta toinen saa kyvyn jakaa viisautta, toinen kyvyn jakaa tietoa…*" "By the power of one and the same spirit, one receives the ability to share wisdom and the other, the ability to share knowledge."

Nevertheless, if we assume that the passage in example **(14)** is indicative of changes in the syntax of the vernacular, another question raised by Steinke's reading is that it entails the assumption that either the *na*-

[45] The manuscript uses only the front jer. Also the case in the latter word in *slovo razumь* must be accounted for. The form can stand either for the nominative/accusative singular or the genitive plural. One, not perhaps very likely explanation could be that the nominative/accusative was intended by the scribe, and the structure would represent the use of the adnominal nominative or accusative in the vernacular, discussed above.

construction grammaticalized from constructions with the locative (or from both the locative and the accusative) or that the development *dative* [recipient] > *na* + *accusative* [recipient] > *na* + *accusative* [adnominal possessor] had already taken place and that the non-accusative singular case in *prěmǫdrosti* is a mistake (however, see **2.2.3** *Dative with prepositions* for the northern dialects of Bulgarian). The predominant view among scholars today, as stated before, is that the adnominal use of the *na*-construction grammaticalized most likely from the expression of recipient, which relied on the goal interpretation of *na* + the accusative, not locative as in Steinke's example. I do not necessarily fully agree with this view, since, as I propose in **2.4**, there may have been another pathway for the *na*-construction into the adnominal position. This grammaticalization path did not require the recipient function as an intermediary step—and therefore both the constructions with *na* + the accusative and *na* + the locative may have been involved in the process.

Should we then accept that constructions with *na* + the locative contributed to the adnominal use of the *na*-construction, it seems to be, however, the opinion of many scholars that its origins do not lie in the abstract use of the locative case and the preposition *na*, but instead in the constructions with a concrete locative meaning (for these views, see Minčeva 1964, 115–116). Yet in MBS manuscripts, there are examples of benefactive constructions, unknown for OCS, that employ *na* + the locative, shown in **2.2.3** *Other uses of the dative as a verbal complement*. These may point to a set of sources for the *na*-construction that is much more varied than what is generally thought.

In addition to the replacement of the adnominal genitive by the preposition *na*, there is also a theory which proposes that the preposition *otъ* (here in the OCS form), indicating source, would have preceded *na* in this function. In the modern varieties the construction with *na* is generalized, although dialectally the construction with *otъ* may also be predominant (Conev 1984 [1919], 477; see also **4.2.3**). This view of the previous role of *otъ* is often repeated, but usually without any concrete examples. Duridanov (1956, 135–136) reiterates this view, pointing to the increased use of the preposition in the *Wallachian Letters*. However, Bernštejn (1948, 308), whom Duridanov cites, merely observes that, in the light of the letters, the construction with *otъ* must have been older in this function than *na*. However, Duridanov goes on to say that the construction with *otъ* never gained the status of the replacement for the adnominal genitive in all of the

dialects and that its dialectal attestations may be of Romance influence. The role of the Balkan Romance languages as the source also seems to be an oft-repeated truism. Here, it must kept in mind that the adnominal possessive constructions with the corresponding preposition *de* (< Lat. *dē*) in Romanian, for instance, are rather limited, although there is some indication that the use of the preposition may have been more frequent in the past (see **3.2.2**). In addition, a parallel construction with the preposition *apó* in Northern Greek dialects must be taken into account. Yet Dionysios Mertyris (2014, 262) states that the constructions in Macedonian and Aromanian may in fact have reinforced the possessive use of *apó* in the Northern Greek dialects. The expression of adnominal possession with prepositions is discussed further from a general Balkan perspective in **4.2.3**.

2.2.3 Dative

In grammatical descriptions, dative is the name given to a case that, according to Barry Blake (2001, 143–144), habitually encodes the following three functions:

1) the indirect object of two-place verbs low on the transitivity scale,[46] such as 'to help', 'to like', and 'to seek';
2) the indirect object of three-place verbs, such as 'to give', 'to show'; and
3) purpose (she went for fish) or beneficiary (she went for [on behalf of] her mother).

Additional, cross-linguistically less frequent, yet typical functions include

4) possessor,
5) destination (goal),
6) indirect object of detransitivized constructions, for instance, in the antipassive,
7) the direct object of certain aspects or tenses, and
8) the indirect subject of certain verbs or of all verbs in certain aspects.

In the absence of exhaustive typological research, it is difficult to judge how well this listing works universally. In the case of OCS, however, it seems that at least the functions 1), 2), 3), 4), 5), and 6) apply to the dative case, with the least overlap with other cases or prepositional constructions being found in function 2), the recipient of ditransitive constructions (see also Haspelmath 1999b, 126 for a semantic map of the dative functions).

Of the modern Balkan Slavic varieties, the dative case survives only in a limited number of dialects, mostly with male personal names and kinship terms. The *ŏ*-stem singular ending is the most often used. In the dialect of Kičevo, spoken in western Macedonia, two types of forms occur:

[46] For a comprehensive list of semantic and pragmatic properties contributing to high transitivity, see Hopper & Thomason 1980, 252. For its critical discussion, see Kittilä 2002, 26–30.

(15) The dialect of Kičevo (Labroska 2008, 162)

sin	*mu*	***Stojan-u***	*pošol*	*vojnik*
son	he.DAT.M.SG	Stojan-DAT.M.SG	started.out.as	soldier

'Stojan's son became a soldier'

(16) The dialect of Kičevo (Labroska 2008, 145)

mu	*dadof*	***Trajče-tu***	*pila*
he.DAT.M.SG	I.gave	Trajče-DAT.M.SG	saw

'I gave Trajče a saw'

As can be seen in the examples, the dative form may express both the recipient and the possessor (these examples are discussed further in **4.2.1**). Example **(15)** shows the *ŏ*-stem dative singular ending, while **(16)** employs a similar strategy in the formation of the dative form, as was shown in **2.2.2** regarding the genitive–accusative in the Macedonian dialects: the *-t-* element is from the *t*-stem declension, while the dative desinence is from the *ŏ*-stem. Again, the greatest variety of case forms is attested in the Rhodopian dialects. In addition to the type presented in example **(15)**, there is a form that continues the *t*-stem dative singular *-ti* (Stojkov 1968, 35). There is also one additional type that employs the pronominal masculine singular ending **-omu*/*-emu*: *Ivan-umu* (ibid.). The definite dative singular article for the masculines is grammaticalized from the dative demonstrative pronoun: *čilḗk-u-tumu* 'to/of the man'.[47]

The *(j)ā*-stem dative singular, found in the Rhodopian dialects, is limited to animates. The first type, according to Stojkov (1968, 36–37), reflects exclusively the soft-stem ending **-i*. However, Miletič (1890, 256–257)

[47] While I acknowledge the burden that comes with the term grammaticalization (for the criticism of some of the uses of the concept, see Joseph 2014; Campbell & Janda 2001), it is used throughout this study to describe certain instances of language change. This is done mostly for lack of a better word to describe the process through which an element becomes a marker of a grammatical function. This may or may not involve the semantic bleaching of the element in all contexts or it becoming more bounded on the cline from an independent word to a bound morpheme. In this study, however, these criteria are mostly fulfilled when the term is used. Although I do not adhere to unidirectionality as a general tenet of language change, the definition presented here still entails that we must be able to define what is grammatical—which, in my opinion, is the most difficult problem about the concept of grammaticalization. An attempt to do this often confronts the risk of circular reasoning: We may have an intuition about what counts as more grammatical, but this intuition is reinforced by repeatedly observing certain diachronic processes in different languages which we then think of as a development toward the more grammatical. However, grammatical is often defined also in terms of obligatoriness. What is obligatory in a language is possible to operationalize and present as a statistical variable, for instance, through acceptability judgments by the speakers (I would like to thank Eric Prendergast for reminding me of this). Although not entirely without problems, this is the working definition of grammaticalization adopted in this study.

finds examples of the (*j*)*ā*-stem dative singulars in Macedonian dialects as well, with both the *-i* and *-e* endings (see also the discussion in **2.2.2** *Genitive in dialects*). In the Rhodopian dialects, there is still one very marginally attested dative singular form, with the ending *-ɛhi*. While *-ɛ-* (*-e-*) part in the ending continues the hard-stem dative desinence, the *-hi* part has a pronominal origin, shown clearly when the word appears with the definite article: *kráv-ɛhi* – *kráv-e-*(*tohi*) '(the) cow'. In the Rhodopian Tihomir dialect, there is another definite dative form used with the (*j*)*ā*-stem singulars: *žén-oite* 'to/of the woman, wife' (Stojkov 1968, 38). According to Stojkov, the ending represents the pronominal dative case ending and the oblique case of the demonstrative pronoun, but since *-te* is the plural form of the oblique definite article, this reading is not unproblematic. In addition to singular forms, some Rhodopian dialects preserve a plural dative form that can be used with all nouns. The generalized ending seems to continue the soft *jŏ*-stem **-emъ*, and the definite form continues the dative plural of the demonstrative pronoun: *sinov-ém-tem* 'to/of the sons'.

Ivan Duridanov (1956, 226) summarizes the main problems in the history of the dative in Balkan Slavic as follows: how did the preposition *na* grammaticalize into a marker comprising the functions of the dative and the genitive? When can the first signs of this development be observed and what interim steps were required? While the grammaticaliziation of *na* 'on' is indeed the most important question, many aspects of its development remain unknown, despite several attempts to explain it. In the modern Balkan Slavic varieties, *na* + the uninflected form and the remaining pronominal datives express, along with the local functions of the pronoun, roughly the same functions as the dative in OCS, in addition expressing adnominal possession, one of the key functions of the OCS genitive. There are, however, certain minor functions that the dative lost. Some of these are the same in those varieties of Slavic that preserved the dative case, such as the loss of the dative absolute,[48] and therefore they need not be dealt with here. Owing to the large number of the dative's functions, there are many smaller questions regarding the "interim steps" that might be answered by the manuscripts of the Middle Balkan Slavic period.

One general observation, however, is important. Although not explicitly pointed out in the previous literature, there seems to be one general

[48] The dative absolute is a use of the dative in non-finite, participle clauses, where both the participial and the subject are in the dative case. It is similar in behavior to the Latin ablative absolute and the Greek genitive absolute.

tendency, which is apparent in the examples given by Duridanov (1956), Rusek (1964), and Steinke (1968), namely, a large number of instances in which the nominative/accusative is used instead of the expected dative contain a neuter singular noun. One potential textual explanation for this has been noted by several scholars: In the Cyrillic manuscripts, *u*, which is the dative singular desinence of most of the neuters, is often denoted by the digraph <oy>. If the second of the graphs is accidentally omitted, what is left is <o>, which stands for the frequent neuter singular nominative/accusative ending *-o*. Two pieces of evidence, however, support the hypothesis that this phenomenon may indeed represent an early loss of case inflection in the neuters: First, there is another observation regarding the neuters that cannot be explained by an accidental omission of a graph: The neuters ending with *-e* are reported to appear strikingly often in the nominative/accusative instead of in the expected case (see also **2.2.2** *Genitive replaced by other cases*). Second, as will be discussed in **2.3**, a common type of vowel reduction in Balkan Slavic renders the unstressed *o* and *u* homonymous. This development may have contributed to the early non-inflection of the neuters, at least in some dialects.

The destiny of the dative is perhaps the most important question in the study of the Balkan Slavic case system. Therefore, considerable space has been devoted to this topic by comparison with the discussions in the other subsections.

Dative as the recipient

The expression of the recipient role (the indirect object) is probably the most common criterion for identifying a case as dative (Newman 1998, 10). As Blake's list suggests, in OCS as well as in the modern Slavic languages, one of the most salient functions of the dative is no doubt the recipient of ditransitive constructions. The usual trivalent verb in OCS is dati/dajati 'to give'. Consequently, the modern varieties of Balkan Slavic employ the remnants of the dative and the na-construction in this role. In a limited context, the OCS recipient of ditransitive constructions is formally separable from the similar semantic role, the addressee of a speaker's communication. A typical recipient, an animate and conscious being handed a concrete object, cannot be expressed by any other means than the bare dative, whereas the addressee sometimes receives the preposition *kъ* with the dative.

The recipient function of the dative has not usually been appreciated in previous research addressing the loss of case inflection. For example, both Steinke (1968, 89) and Meyer (1920, 67) conflate the recipient use with all other functions of the dative as a verbal complement. Duridanov (1956, 149), on the other hand, calls the basic function of the dative “objective,” although his examples show that it is not the indirect object he means, but rather the benefactive or goal function of the dative. Elena Češko (1970, 148) approaches the tasks of the dative by contrasting three “addressee” functions, one of which is the recipient (*adresat predostavlenija*). However, she places the recipient role parallel to various other functions of the dative. Jerzy Rusek (1964, 121) meanwhile establishes a category containing the datives that function as a complement of verbs meaning ‘to give’. This seems to delimit the recipient function the best (see, however, the discussion further on in this section).

The most important observation regarding the recipient function of the dative is that, even though scholars present several examples from MBS sources pointing to constructions whereby, instead of the expected dative, another case or prepositional construction occurs, none of their examples is about marking the recipient. Yet Rusek (1964, 121) suggests that in the *Hludov Triodion* two such instances are found. He takes two examples with the preposition *kъ* used with derivatives of the verb *dati* that he seems to regard as ditransitive:

(17)

svojeǫ	*voleǫ*	*prědavši*	*sę*	*kъ*	*tli*
his/her.INST.SG	will.inst.sg	surrender.act.pst.ptcp.nom.sg	refl	prep	rot.dat.sg

‘With his/her will surrendering himself/herself to decay’

(18)

m(*ol*)*itvy*	*vъzdadite*	*kь*	*h*(*rist*)*u*
prayer.ACC.PL	give.imp.2pl	prep	Christ.DAT.SG

‘Send prayers to Christ’

Example **(17)** can be accounted for without great problems. The reflexive verb *prědati sę* ‘to surrender’ is rather far from a clear-cut ditransitive verb. While able to accommodate a dative complement, the “recipient” of the surrender or the surrendered, the complement is not obligatory to the same degree as with the verb *dati/dajati*. Also in this particular case the suspected recipient is highly atypical: it is an abstract noun, not a person consciously able to receive anything. It is more credible that the use of *kъ* here relies on

the meaning ‘directed toward’ in the construction, not an extension of *kъ* + dative into a marker of the recipient.

Example **(18)** poses more problems. In the context, Christ is an animate being, capable of consciously receiving objects. In OCS, the verb *vъzdati*, however, often appears in a phrase used to translate the Greek verb *eucharistéō* ‘to thank’: *hvalǫ vъzdati*, literally ‘to give praise/thank’. The construction can appear without a recipient. The word *molitva* can also contribute to the choice of *kъ*, since in OCS with the verb *moliti* ‘to pray’, the addressee can be expressed both with the bare dative and with *kъ* + the dative. In any case, while the use of *kъ* here could illustrate of a more general vacillation with the bare dative, it is not an example of a typical ditransitive construction because the theme, the object being handed, is not a concrete thing. All in all, the meaning of the phrase *kъ* + the dative is closer to the addressee of communication than the recipient of giving.

Rusek (1964, 121–122) observes a handful of examples in the *Hludov Triodion* with the verb *dati*/*dajati*, which have either the preposition *vъ* or *na* with the accusative. As he himself points out, these instances directly translate Greek constructions with locative prepositions. In addition, the examples are often understandable only as abstract, biblical use of language. An interpretation of a passage from the *Hludov Triodion* by Ljubomir Miletič (quoted in Mirčev 1957, 39), who claims that it represents an early instance of replacing the dative with *na* and the accusative, stumbles on an interpretive difficulty arising from complex religious symbolism:

(19)

i	*ne*	*daždь*	*na*	*ny*	*krъve*	*pravednyǫ*
and	not	give.IMP.2SG	prep	us.ACC	blood.GEN.SG	righteous.GEN.SG

Mirčev (1957, 39) reminds us of the risks of searching for only co-occurrences without a good interpretation of the text: The translation “do not give us righteous blood,” implied by Miletič, is theologically impossible, and a similar metaphor can be found in Matthew 27:25: “And all the people answered, ‘His blood be on us and on our children!’” (*eph' hēmâs kaí epí tá tékna hēmō̂n*). In many Slavic translations outside the Balkans, the preposition used is *na*. Mirčev himself translates the passage *ne dopuskaj pravedna krăv da tekne vrhu nas* “and do not let the righteous blood to flow on us.”

Another passage, cited by both Meyer (1920, 70) and Duridanov (1956, 232), with a suspected use of *na* instead of the dative, this time with the

locative/dative (the two cases being homonymous in the non-gendered singular pronouns), is from Matthew 14:15 as found in the *Evangeliary of Vratsa* from the 13th century:

(20)

kupęť	*brašno*	*na*	*sebě*
buy.PRS.3PL	flour-ACC.SG	PREP	oneself.DAT/LOC

'[to] buy themselves food'

While perhaps only bordering on the recipient function and being instead a benefactive, the pronoun is expected to appear in bare dative, since both the OCS translations and the Greek original contain the bare dative. Duridanov wonders whether the pronoun in *na sebě* was intended as a locative or whether the preposition was used to "strengthen" the dative, as in some Northern Bulgarian dialects (see *Dative with prepositions*). The identification of the case of the pronoun, however, is more complicated. Benjo Conev (1914, 47–48) counts five instances in the *Evangeliary of Vratsa* manuscript where *ě* is used instead of the expected *e* of the accusative ending of a non-gendered pronoun, thus rendering the form similar to the dative/locative. This is also pointed out by Mirčev (1957, 39–40) in his criticism of Meyer's reading.

While the manuscript itself does not offer any conclusive evidence to clarify this instance, there are certain things that must be taken into consideration. The last syllable of the word *brašno*, which precedes the preposition *na*, is very close to the preposition, an observation that makes Mirčev (ibid.) suspect a scribal error. A further indication that this could be a case of dittography (the insertion of a superfluous duplicated syllable, a typical scribal error; for the *Hludov Triodion*, see Lunt 1965, 308), is that the original probably had the accusative plural *brašъna*, which corresponds to the Greek *brṓmata*, as in the *Codex Marianus*, the *Codex Zographensis*, the *Book of Sava*, and the *Evangeliary of Assemanius*. The alleged preposition is written together with the word *brašno*, as would be expected in any event, since in the manuscript tradition space was used only occasionally to indicate a word boundary. Thus, the preposition becomes the last word in the line.[49] In addition, the surrounding folia contain several mistakes and corrections, indicating scribal negligence. For example, on the previous folio (55v) a deletion was made by scraping, while on the

[49] Also, the line-final *a* in *na* is written with the in-line graph, not the line-final, "italicized" variant used elsewhere. While perhaps of little value, this observation shows that the passage is atypical in other ways as well.

surrounding folia several more deletions were made by wiping off the ink. In addition, a curious mistake can be found two lines below the passage discussed here. The scribe first wrote *esti* 'to eat', but corrected himself by adding *a* after *e*, making the two letters look like a ligatured *i* and *a*, corresponding to the OCS and Balkan Slavic form *jasti*. If the original contained indeed *esti*, it is most likely because the original of the *Evangeliary of Vratsa* was of East Slavic origin, as Conev (1914, 6–7) suggests.

Nevertheless, these observations can point in opposite directions. While the scribe's carelessness supports the theory of dittography, the struggle with an original text containing foreign elements and, more important, the scribe's willingness to make changes that differed from the original manuscript could indicate that the suspected *na*-construction was intentional, indicating that the language spoken by the scribe was leaking into the manuscript. However, I find the latter possibility unlikely: The correction *esti* > *jasti* could be interpreted as a correction toward the spoken Middle Balkan Slavic, but also toward the norm. All in all, the close study of this particular passage is intended to show that this alleged early instance of a *na*-construction in a recipient-like function is anything but certain.[50]

Dative as the addressee: Increase in the use of the preposition *kъ?*

Several researchers have noted a growing tendency to mark the addressees of communication with *kъ* and the dative instead of the bare dative (see, e.g., Steinke 1968, 89). This competition between the dative and the prepositional construction occupies a central role in Meyer's argument, the preposition *kъ* paving the way for the new construction with *na* and the accusative (Meyer 1920, 72–74). While many researchers report an increase in the use of the prepositional construction compared with older texts, they seldom give any data on its distribution or frequency. In the 13th-century *Hludov Triodion*, however, Rusek (1964, 118) counted 107 instances of marking the addressee of the verb *rešti* 'to speak, say' with the bare dative and 122 with the preposition, a ratio of 9:8, the bare dative narrowly dominating. By comparison, in the 14th century *Combined Paterikon*, which represents more normativized language than that found in the *Hludov Triodion*, the

[50] A similar example of a suspected *na* + dative construction in the *Hludov Triodion* is discussed originally in Miletič 1905 and circulated further without much criticism by Mirčev (1957, 39), Rusek (1964, 124), and Duridanov (2004, 225). However, Horace G. Lunt (1965, 306–308) shows convincingly that this interpretation is based on a misreading of the preceding, abbreviated word as *gospodi*, not *gospodina*.

ratio is 5:3, the bare dative being more frequent than in the *Hludov Triodion*.[51]

Since this observation is used as an example of an overall increase in analyticity, a comparison with the OCS use of *kъ* + and the dative expressing addressees is needed. In the *Codex Marianus*, in constructions with the verb *rešti* 'to say' which include an addressee, the ratio between the bare dative and the prepositional construction is 5:1 in favor of the bare dative.[52] However, the gospels differ significantly in their use of this construction. When the distribution of the two structures with the verb *rešti* is examined only in the Gospel of Luke, the ratio is 3:2, significantly less in favor of the bare dative.[53]

This particular variation in OCS is explained largely by the Greek source text. A good picture of the role of the source is given in the semantic analysis of the New Testament Greek prepositions and their translation equivalents in OCS by Olga Thomason (2006). In the addressee function, *kъ* + the dative most often results from the translation of the Greek construction *prós* + accusative.[54] The number of addressees marked with *pròs* + accusative is affected only to a slight degree by the sizes of the gospels and the number of the speech verbs there, Luke, for instance, being well-known for a particularly frequent use of the construction (Thomason 2006, 22).

	Matthew	**Mark**	**Luke**	**John**
Greek	2	14	113	24
OCS	11	13	121	23

Table 2. The addressee use of the Greek prós + accusative and the OCS kъ + dative constructions in the gospels (Thomason 2006, 23, 139).

Although the use of *kъ* + dative often coincides with a preposition in the Greek original, Thomason (2006, 303, 328) notes "several instances" in OCS where both spontaneous prepositional constructions and bare datives occur in this function. The frequency of the *kъ* + dative addressees depends heavily on the original, yet the correspondences are not without exceptions. Therefore, observing these two features without a systematic comparison with the source texts is not enough to make claims about an increase in the use of the prepositional construction. Acknowledging this, Rusek (1964,

[51] N=386; corpus size: 91,893 words.
[52] N=537; corpus size: 58,628 words.
[53] N=186.
[54] Although she combines addressees with recipients in her analysis, it emerges from her examples that "recipients" expressed with *kъ* + dative include only the addressees of speech verbs (Thomason 2006, 302, 328).

118–119) samples 99 constructions with the verb *rešti* in the *Hludov Triodion*, which shows the ratio 5:3 between the constructions, the bare dative dominating. The use of the bare dative corresponds to the Greek dative in nine out of ten cases, the remaining datives being translated from the Greek *pròs* + accusative. The *kъ* + dative translates the Greek prepositional construction in two of three cases, one-third of them representing the spontaneous use of the prepositional construction.

Rusek's observations point to an increase in the spontaneous use of the prepositional construction in the *Hludov Triodion* compared with OCS. However, in modern Slavic languages *kъ* + dative is seldom used to express the addressees, and the Balkan Slavic dialects, together with the Damascenes, do not exhibit any traces of this use. It is possible that the source of the construction in OCS and in later literature was merely a word-for-word translation from Greek, not a feature of any spoken variety of Slavic. This is further supported by the fact that in the *Tale of Troy* from the 14th century, whose source text is an unidentified non-Greek language, *kъ* + dative is used only three times with the verb *rešti* compared with 55 cases of the bare dative.[55] For these reasons, it is not wise to consider an increase in the use of the *kъ* + dative in the earlier MBS manuscripts as evidence of a similar development taking place in the spoken language.

Nevertheless, this phenomenon in the MBS manuscripts could still be viewed as evidence of changes in the spoken languages of the scribes and translators, although not in Meyer's sense. He alleges that the *kъ* + dative also replaced the dative in the spoken language as a part of an increase in the overall analytism of the language. As I have demonstrated above, it is unlikely that an interim construction with *kъ* + dative preceded the *na*-structure in expressing the recipient. Yet in my view, the more frequent spontaneous use of the prepositional construction could be indicative of the phrase having become a formulaic expression, whereas in OCS the construction resulted more often from word-for-word translations. The increase in the formulaic use of language in the manuscripts as well before the orthographic normativization in the 14th century is an equally telling, albeit less specific, indication of the growing distance between the written and the spoken language.

[55] Nevertheless, the *Tale of Troy* poses several problems with regard to its source and the origin of the Slavic translation. It contains several lexical and grammatical features that point to BCMS.

Other uses of the dative as a verbal complement

The majority of the more convincing instances of replacing the dative with an alternative construction are found in verbal complements other than the recipients and addressees. The dative can act as the complement of a wide range of verbs in OCS, roughly divisible into four categories. The first group includes, to use Blake's (2001, 143–144) wording, two-place verbs low on the transitivity scale, such as *pomošti* 'to help', *služiti* 'to serve', *věrovati* 'to believe something/somebody', *ponositi* 'to reproach', and *zaviděti* 'to envy'. Their equivalents often have a dative complement in other Indo-European languages as well. These verbs invariably have the pronominal dative or the *na* + uninflected form in modern Balkan Slavic. The second group includes reflexive verbs, such as *javiti sę* 'to appear to somebody', *radovati sę* 'to rejoice at', *klańati sę* 'to kneel before somebody in veneration', *smějati sę* 'to laugh at someone', and *naděјati sę* 'to hope for something'. The third group includes verbs with the prefix *pri-*. These verbs rely on the goal interpretation of the dative and often have *kъ* + the dative as their complement. The fourth and least closed group consists of usually transitive verbs that accept the dative as an optional argument, relying on its benefactive interpretation or, more generally, its interpretation as an affected participant. These verbs include, for example, *sъtvoriti* 'to do (something) to someone' and *pokazati* 'to show (something) to someone'.

The examples of atypical constructions instead of the expected dative in the previous research can be divided into two types, both having different implications for the analysis. The first type of constructions includes constructions that translate a Greek prepositional phrase, constructions that in OCS would take the bare dative. The second type of constructions deals with uses of prepositional constructions instead of the expected dative; such constructions cannot have been modeled on the Greek original because it does not contain a preposition. As in some previous examples, an explanation of the first type is not straightforward. It could either indicate the translator's poor language skills or a rigid attitude to the original: instead of finding a proper Slavic translation, the translators have resorted to word-for-word translations that reflect neither the written tradition nor the spoken language.

Yet, as with the *kъ* + dative expressing the addressee, the use of the first type of constructions may indirectly reflect growing differences between the written norm and the spoken language, but in my opinion, it does not offer concrete evidence for any particular change in the spoken language. For

many scholars, these constructions represent, however, evidence of corresponding structures in the spoken language, despite the fact that they potentially result from a word-for-word translation. On the other hand, examples of the second type, while being perhaps more easily analyzed, are not free of the effect of literary calques, since there is often the possibility that an expression which has its origin in a word-for-word translation may be used as a formula also when the original structure is not translated word-for-word.

As an example of the first type, Mirčev (1957, 44), who is generally very skeptical of other scholars' observations, cites an instance with *na* in the clarification of Psalm XXI in the *Bolognese Psalter* from the 13th century as an "analytically rendered dative."

(21)

o	*tomъ*	*bo*	*blažǫtъ*	*sę*	*vsi*	*věrovavšei*	*na*	*h[ristos]a*
PREP	that.LOC	then	benefit	REFL	all	believing.NOM.PL	PREP	Christ.ACC.SG

'everyone believing <u>in Christ / Christ</u> benefits from that'

Mirčev unfortunately goes astray with his reading. He does not take into account the difference between the two semantically close constructions with the verb *věrovati*, 'to believe someone/something' and to 'to believe in someone'. The first of these constructions has the bare dative, while the second has *vъ* + accusative in all modern Slavic languages, including Balkan Slavic. Mirčev states that "normally" the verb takes the dative in OCS, but frequently also the preposition *vъ*, and that the latter use represents a calque from Greek.

Mirčev is on the right track, because in the *Novum*, the Greek language does not explicitly distinguish between the 'to believe in' and 'to believe' meanings of the verb *pisteúō*. Both the bare dative and the preposition *eis* 'in' with the accusative are used in the sacred sense, although the latter use is characteristic only of the Gospel of John, leading, not surprisingly to a similar preference for the prepositional construction in the OCS translation. Thomason (2006, 18, 131) counts 35 uses of the prepositional construction in the Gospel of John in this function as opposed to six in all the other Gospels combined. The corresponding figures in OCS are 37 versus three. As the comparison of the Gospels shows, OCS did not yet make this distinction systematically, since the word-for-word translation was acceptable to the translators. However, and this is what Mirčev fatally ignores, the prepositional construction could only be used in the sense of

'believe in someone', and there is no evidence to suggest that the example with *na* given by Mirčev represents the construction 'believe someone', thus replacing the bare dative.

Nevertheless, the use of *na* still remains to be explained. The Greek verb *pisteúō* could be used with the preposition *epí* 'on', as shown in example **(22)**, which corresponds to the preposition *na* 'on' in its locative use.

(22) Romans 4:24

toís	*pisteúousin*	*epí*	*ton*	*egeíranta*	*Iēsoûn*
those.DAT	believe.PRS.PTCP.DAT.PL	PREP	DEF.ACC	raising.AOR.PTCP.ACC.SG	Jesus.ACC

'to those believing in the one who raised up Jesus [from the dead]'

Mirčev does not give the Greek original, which may be unknown, but here again, it is likely that example **(21)** follows the original word-for-word. Furthermore, it is worth noting that both the Slavic and the Greek examples in fact represent a nominalized construction, which perhaps more easily permits an untypical choice of preposition.

Another interesting construction in the *Bolognese Psalter* observed by Mirčev (1957, 44–45) is a recurring phrase with a preposition and the locative *na člověcě* 'for the sake/benefit of man', often with the verb *sъtvoriti* 'to do' and an abstract, religious concept like clemency, love, or salvation, which is "done" for someone. This phrase, which does not appear in OCS, is most likely in all instances a translation of the Greek phrase *eis tón ánthrōpon*. Because the sense is clearly benefactive, the expected form would be the dative, as Mirčev points out, and he gives examples with similar constructions, but using the dative. In addition, Mirčev (1957, 46) and Duridanov (1956, 233) both observe instances with *na* + locative in the *Chronicle of Manasses* with the word *synъ* 'son'.[56]

While far from clear, this particular construction is interesting because of the choice of preposition, both in the original and in the translation: structures of the type *na člověcě* seem to be an innovation that cannot be explained by a word-for-word translation where the more expected Slavic

[56] In addition, Duridanov (1956, 231) makes the important observation that with the verb *javiti* 'to notify, show', the one that is shown something is often expressed with *na* + locative, translating *eis* + accusative. However, he erroneously combines the reflexive verb *javiti sę* 'to appear' with the non-reflexive verb and interprets the *na* + locative here as replacing the dative complement of *javiti sę*. Consider the semantic difference (Luke 1:11 in the *Codex Marianus*) *javi že sę jemu*[DAT] *anđ[e]lъ g[ospodь]nь* 'an angel appeared to him' and (the *Sinai Book of Prayers* 25a.8 in ibid.) *javi nyně mil[o]stь tvoję na*[PREP] *rabě*[LOC] *tvoemь*[LOC] *semь*[LOC] 'show now your mercy upon this servant of yours'. A similar meaning is often expressed in OCS with *sъ* + instrumental which translates *metá* + genitive (Thomason 2006, 307–308).

preposition would still perhaps be *vъ* 'in'. In effect, the Greek construction *eis* + accusative developed into a construction expressing the recipient in Modern Greek in a manner similar to the Slavic preposition *na*. It could be that the translator's choice of preposition is the result of equalizing the non-locative, grammatical functions of the two prepositions *eis* and *na*, which would support Mirčev's claim that this construction would indeed represent the broadening of the functions of the preposition *na* to express "a datival relation." However, the Slavic use of case further complicates things: The locative is not expected on the basis of the Greek original. Nevertheless, in OCS, *na* + accusative marks, in addition to direction, the abstract malefactive function of acting against someone, conceptually opposite the benefactive intention of these examples. Speculatively, one could ask if the choice of locative is to avoid this confusion.

The second type of atypical constructions in which the Greek original does not have a preposition includes Duridanov's (1956, 231–232) observations about the use of two verbs, *vъzęti* and *otęti*, both meaning 'to take'. In post-OCS texts, these verbs sometimes have as their optional argument the construction *na* + locative, meaning the one that something is taken from. However, Duridanov considers this construction as potentially replacing the dative, since the dative is sometimes used in similar constructions in OCS, as the passage from Mark. 7:27 in the *Codex Marianus* shows: *otęti hlěba*[gen.] *čędomъ*[dat.pl.] 'to take away the children's bread'. Yet the use of the dative here does not necessarily mean 'from somebody', expressed with *otъ* + genitive in OCS. It could also stand for the possessor of the bread, which, admittedly, is semantically close. For further proof, unlike *otъ* + genitive the dative cannot be used in these constructions without the direct object in OCS. Notably, *na* + locative is used without the direct object in all of Duridanov's examples.[57]

***

Duridanov's efforts should not be judged as misguided, however. In fact, he does not set out to find instances in which the prepositional constructions have indisputably replaced the dative, but where the functions of the

[57] The reason that the construction *na* + locative in these texts replaces a construction with the preposition *otъ* that is still used in the modern varieties of Balkan Slavic remains to be explained. Duridanov does, however, mention an Old Serbian passage employing the same structure, this time with a direct object. This should be seen as a clear indication that this construction does not need to be connected in any way to the reorganization of the Balkan Slavic case system.

preposition *na* could be close enough to the dative to provide fruitful contexts for the further grammaticalization of the *na*-constructions (Duridanov 1956, 229). Although some of the previous examples are clearly mistaken in this sense as well, this approach has nevertheless provided a few compelling instances, perhaps analyzed most systematically by Elena Češko (1970, 187–194).

Relying on Duridanov (1956, 229–230) and Mirčev (1978, 285), among other scholars, Češko (1970, 189) mentions a group of verbs that, along with the dative, accepts complements with the preposition *na* in a similar function. These are *radovati sę* 'to rejoice in/at something', *blagovoliti* 'to content oneself with', *naděję ti sę* 'to trust in', *upovati*, 'to trust in', *razuměti* 'to understand', *želati* 'to desire', and *žędati* 'to thirst for'. This grouping is somewhat artificial, relying, in Češko's words, on the shared functions of perceiving (*vosprijatie*) and regarding (*otnošenie*). Of these *razuměti* stands out as it accepts in OCS the accusative, genitive, dative, and *po* + locative (on the latter, see the *Codex Marianus*, Luke 9:11), no doubt in somewhat different functions. Both constructions with *na* + accusative presented by Češko are nominalized (see further **2.4**). Then again, the examples given to show the use of *želati* and *žędati* are clearly instances in which the choice of *na* + accusative relies on its strongest known function in OCS, the directional, and clearly do not belong to this grouping.[58]

Nevertheless, the abstract verbs *radovati sę*, *blagovoliti*, *naděję ti sę*, and *upovati* are of extreme interest. With the exception of *upovati*, they mostly take the dative in OCS, sometimes also *o* + locative, but in later texts they vacillate among *na* + accusative, *na* + locative, *o* + locative, and the bare dative (and *blagovoliti* also with *vъ* + locative). As for the synonymous *naděję ti sę* and *upovati* as found in OCS, *upovati* always takes *na* + accusative, while *naděję ti sę* most often takes the dative. As Duridanov (1956, 229–230) observes, these two verbs have perhaps most clearly affected each other, since the use of *na* + accusative increases in later texts. To this group could be added two isolated examples observed by Duridanov (1956, 232). In the *Chronicle of Manasses*, plural forms of the demonstrative pronouns *jь* and *sь* are used in nominalized structures in the locative with *na*: one with the verb *pomagati* 'to help' and the other with *sъmiliti sę* 'to conciliate'.

[58] In fact, all the examples reflect Psalm 42:1: "As the hart pants after the water brooks, so pants my soul after you, O God." The *Septuagint …epì tàs pēgàs… …pròs sé ho theós*, explains the use of *na* + accusative very well.

Dative with prepositions

Examining the use of the dative with the prepositions *kъ* 'toward', *po* 'along, according to', *prěmo* 'opposite of', and *protivǫ/protivъ* 'against' is a particularly easy task compared to the other uses of the dative because the inflected noun is the dependent of the preposition, a situation that does not leave much room for interpretation. However, there are a few problems that require more analysis.

Steinke (1968, 92–93) acknowledges the following three versions of the same verse in the three apostle manuscripts, although he does not try to explain them:

(23) Hebrews 2:4

po *svo ego*[M.GEN.SG] *voli / po* *svoego*[M.GEN.SG] *voli / po* *svoemu*[M.DAT.SG] *voli*

'[while God added his testimony through signs, wonders, various miracles, and gifts of the Holy Spirit distributed] according to his will'

The use of the possessive adjective *svojь* in the genitive is curious. With the possessive adjective, agreement is expected with *volji*[fem.dat.sg.] and therefore the expected form is feminine dative singular *svojejь*. It appears that the first two versions are word-for-word translations from Greek, which employs the personal pronoun in the genitive: *autoû*. In the first two translations, the possessive adjective employs the genitive, which must result from confusing the adjective with the use of the personal pronoun in genitive. Compare the following structure in the OCS manuscript *Codex Marianus*:

(24) Luke 12:47

po	*volji*	*jego*
PREP	will.F.DAT.SG	he.GEN

'according to his will'

Steinke's example (**23**) is nevertheless complicated by the third version, which has a masculine dative singular that seemingly agrees with the case of the noun *voli*[f.dat.sg.], but not with the gender. I think it is clear that the dative is used here for the same reason as the genitive: the form is used as a personal pronoun, not as a possessive adjective, and the choice of case is an example of the adnominal use of the dative, not agreement within the NP governed by the preposition *po*. What is misleading here is the word order, which is copied from the Greek original, and which, at first glance, would suggest the interpretation of the pronoun as an adjective modifier. The

choice of the possessive adjective is, nevertheless, puzzling. It is not likely that the scribes could not use it right, since the possessive adjective is still used in Balkan Slavic, although it is not inflected for case any longer. The possessive adjective *svojь* is used when the possessor is the subject of the clause. Perhaps what was then intended with the choice of the possessive adjective instead of a personal pronoun was to explicitly point to the right correlate, God, the subject of the sentence, which is separated from the possessive construction by a long passage.

Neither Steinke, Češko (1970, 142–194), Rusek (1964, 137–139), nor Duridanov (1956, 142–154), draw much in the way of conclusions about the use of the dative with the prepositions: There are no tendencies, distinct from the overall vacillation in the use of the cases, that would be considered direct reflections of developments in the spoken language. Meyer's (1920, 72–74) belief that the use of *kъ* + dative primed the language for an analytical construction to replace the dative is not based on conclusive evidence of increase in the use of the preposition, as was shown earlier. In addition, the preposition *kъ* was never used to mark the recipients proper. In addition, it seems that the dative prepositions *kъ* and *prěmo* withered away together with the dative in Balkan Slavic and were never used with the new oblique case. Interesting evidence of this is found in the Damascenes: For example, in the *Damascene of Trojan* (Ivanova 1967), a manuscript from the early 17th century, the preposition *kъ* appears only five times, and exclusively with masculine singulars inflected for the dative, the only remaining context for the dative case. Elsewhere, this structure is replaced by *kămto* + *casus generalis*, the precursor of the modern Bulgarian *kăm* and Macedonian *kon*. On the other hand, *po* and *protivǫ* survived, perhaps relying on their use together with other cases.[59]

There is still one particular problem related to the use of the dative with prepositions, which is confusing, to say the least. As a general tendency, the prepositions were used increasingly with the oblique case, that is, with either the accusative or the nominative. However, in the Cserged Prayers the preposition *na* is sometimes used with the dative adnominally or to express the recipient:

[59] The preposition *po* was used in OCS together with the accusative to indicate the senses of 'along, within the limits of'. The destiny of *protivǫ*, on the other hand, varies in the Slavic languages. In BCMS (and East Slavic), the preposition takes the genitive. It is possible that such use of the preposition existed in Balkan Slavic as well, at least transitionally, before the loss of the genitive.

(25) *Cserged Prayers* (Miletič 1987 [1896–1900], 163)

na	*moie*	*Dufs-i*	*Lecsina*
prep	my.F.DAT.SG	soul-F.DAT.SG	blessedness

'blessedness of my soul'

Miletič (1926, 59) notes that in some dialects from Northern Bulgaria, where the Cserged Bulgarians are thought to have originated, there are still remnants of the preposition *na* used with the datival pronouns. He gives the following example: *na nam*[dat.] 'to us'. Whether this construction had a larger distribution in the Balkan Slavic dialects as an intermediate step in the development of the construction with *na* + oblique case is unclear. However, it has to be assumed that the prepositional construction with the dative must represent a stage where the preposition *na* was already widely used with the oblique case to express the recipient and the adnominal possessor. Only after these functions of the preposition *na* were firmly established could the preposition be used with the remaining dative forms, as the *na* + oblique case must have originated in contexts in which the dative case forms could no longer be attained. A tentative suggestion is that the attestations of the construction with *na* + dative on the northern periphery of Balkan Slavic might be indicative of a later diffusion of the *na* + oblique into these dialects from a more distant center of innovation. In my opinion, the new construction probably still faced competition from a large number of preserved dative forms, unlike in the dialects where the prepositional construction was born out of necessity to mark the functions of the lost dative case. Note that this view of the need for a replacement construction is in stark contrast to Karl Meyer's (1920, 31), who claims that the loss of case forms is not possible if the replacing construction does not already exist.

2.2.4 The locative and instrumental cases

The common denominator of the locative and instrumental cases is that both were practically lost in their entirety. They were largely replaced by prepositional constructions employing the oblique case or the uninflected form. The loss of these cases is connected to two functional mergers, shared with the other languages of the Balkans (see **3.3**): the merger of goal and location, relevant for the locative, and the merger of instrument and accompaniment, involving the instrumental.

Locative

In OCS, the locative was used almost exclusively with prepositions. The only exception is a small number of verbs with the prefix *pri-*, with which the bare locative indicated 'being over by or in the vicinity of something'. The locative with prepositions *vъ* 'in', *na* 'on', *o* 'around', *po* 'after', and *pri* 'by' was used chiefly to indicate location. When the same prepositions (with the exception of *pri*) were used with the accusative, the meaning denoted goal. Very few forms of the locative are preserved in the dialects. With the exception of the plural case form analyzed first as the genitive (discussed in **2.2.2** *Accusative*), only one form survives. Stojkov (1968, 39–40) observes a single case form in the Rhodopian dialect of Tihomir:

(26) Rhodopian dialect of Tihomir (Stojkov 1968, 40)

ženata	*paodi*	*na*	*pazar-ɛ-nɛ*
the woman	left	PREP	marketplace-LOC-that.LOC[60]

'the woman left for that marketplace over there'

As example **(26)** illustrates, the case form that continues the locative does not express location alone, but also goal (cf. the Albanian "locative" construction in **3.2.1**). It could thus be said that neither of the two preserved locative forms continues the earlier functions of the locative as such.

Because the prepositions used with the locative in OCS appear in the modern dialects and early vernacular texts of the 17th century with the uninflected form or the oblique case, what has been sought in the MBS manuscripts is the use of the nominative or accusative instead of the expected locative. Steinke (1968, 105–106) observes several instances in his 13th century corpus where the accusative or nominative/accusative appears instead of the expected locative. This prompts him to claim that the locative must have been rare in the spoken language of the time (ibid., 107). Although he does not acknowledge it explicitly, over half of his examples involve neuters.[61] Sobolev (1991, 33) notes a similar tendency in even older manuscripts: the earliest examples of the nominative/accusative instead of the locative involve the neuters, especially those with the ending **-ije*,

[60] The distal demonstrative pronoun attaches to the word similarly to the definite article. In some descriptions, the distal and proximal clitic pronouns in Balkan Slavic are treated as definite articles, based on their form. Yet the main criterion in disambiguating a demonstrative from a definite article is exactly that the article does not express deixis (Topolinska 2006).

[61] As a point of comparison, the number of neuter forms appearing in the locative in the *Codex Marianus* is 383. The number of feminine locative forms is 349 and the number of masculine locative forms 479.

whose locative ending was **ij-i*. Sobolev suspects that the reduction of the unstressed *e* to *i* (see **2.3**) may have contributed to the loss of inflection with the neuters. It is indeed likely that the functional distinction between the locative and accusative was lost early, mostly in favor of the emerging oblique case. Here, it is important to note example **(26)**, which shows that the key phenomenon must have involved a functional merger, not necessarily the loss of inflected forms. On the other hand, the neuters may have offered the first context in which the new default prepositional case appeared in the locative function.

Instrumental

The bare instrumental was used in OCS to indicate instrument and route in the sense of going through, across, or along. In addition, it expressed the agent of a passive construction. In modern Slavic languages, it is also used to some extent as an alternative predicative case, contrasting in this function with the nominative. A handful of prepositions used with the instrumental conveyed various local meanings. These prepositions could in some instances be used with the accusative as well, to indicate movement toward. However, the use of the instrumental with these prepositions could also mean movement toward, differing in this sense from the division of labor between the locative and accusative. Used with the preposition *sъ*, the instrumental case indicated accompaniment. In Balkan Slavic, the replacing construction for the instrumental expressing instrument and accompaniment uses this preposition and the uninflected form or the oblique case. No productive forms of the instrumental survive in the dialects, although there are several petrified instrumental forms used as adverbials (for these, see Miletič 1890, 265–268).

Duridanov (1956, 113) reminds us that the preposition *sъ* appeared occasionally as early as in OCS, used with the instrumental case to express instrument. Since the varieties of Slavic that display the instrument–accompaniment merger all have a history of significant contacts with languages that have a homonymous expression for these functions (Nomachi & Heine 2011), it is likely that these early instances may already show the effect of language contacts with non-Slavic Balkan languages. Sobolev (1991, 31–32) points out that the first instances of the preposition *sъ* with the oblique case appear in manuscripts from the 12th century. He continues by noting that the instrumental case was subject to homonymy as early as in the Proto-Slavic era, and the later sound changes in Balkan Slavic further

affected the instrumental case (on the sound changes, see **2.3**).[62] The destiny of the instrumental case is "simple" in a manner similar to the locative: The replacing prepositional constructions increasingly begun to use the oblique case, and therefore no intermediary stages need to be reconstructed.

2.3 Sound changes and the loss of case inflection

There are two questions related to the role of Middle Balkan Slavic sound changes in the loss of case inflection. First, what were the immediate consequences of sound changes to the Balkan Slavic case system in terms of the syncretism they created? Second, did the increase in syncretism—the new homonymy among the case forms—contribute to the loss of case inflection beyond their immediate consequences? Answering these questions is not a straightforward process, as some of the sound changes thought to play a key role in the resulting syncretisms are themselves in dispute. Most important, the destiny of the two Late Proto-Slavic nasal vowels involves complex problems. In addition, other developments in the inflectional paradigms add to the difficulty of analyzing the effect of the Balkan Slavic sound changes.

As discussed in **1.4**, the effect of MBS sound changes and the resulting homonymy between the case forms have been regarded by many as a factor in the demise of the case inflection. Although Samuil Bernštejn (1948, 357–366, see also **1.4.2**) claims that most of the prominent Bulgarian linguists between the late 19th and mid-20th centuries supported the sound changes as the primary cause, this is not true, at least not for Benjo Conev and Kiril Mirčev. Both scholars do cite the sound changes as one of the reasons for the loss of case inflection, but by no means do they see it as the only reason (Conev 1984 [1919]b, 455–457; Mirčev 1978, 292–294). Nevertheless, since the sound changes are evaluated in most accounts of the loss of case inflection, their consequences need to be discussed here. **Table 3**, which contains the same OCS inflectional paradigms as **Table 1**, illustrates Klaus Steinke's (1968, 20) summary of the immediate effects of the sound changes (with some minor modifications, which I have made). Highlighting indicates the case forms that are assumed to have become homonymous in Balkan Slavic as a result of sound changes following the OCS period:

[62] It is difficult to say how relevant the old homonymy was, since it involved only the *ŏ*-stem plural nouns.

		‘slave’	‘man’	‘son’	‘woman’	‘land’	‘bone’
		***ŏ*-stem**	***jŏ*-stem**	***ŭ*-stem**	***ā*-stem**	***jā*-stem**	***ĭ*-stem**
Singular	**nom.**	*rab-ъ*	*mǫž-ь*	*syn-ъ*	*žen-a*	*zemĺ-a*	*kost-ь*
	acc.	*rab-ъ*	*mǫž-ь*	*syn-ъ*	*žen-**ǫ***	*zemĺ-**ǫ***	*kost-ь*
	gen.	*rab-a*	*mǫž-a*	*syn-u*	*žen-y*	*zemĺ-**ę***	*kost-i*
	dat.	*rab-u*	*mǫž-u*	*syn-ovi*	*žen-ě*	*zemĺ-i*	*kost-i*
	loc.	*rab-ě*	*mǫž-i*	*syn-u*	*žen-ě*	*zemĺ-i*	*kost-i*
	instr.	*rab-omь*	*mǫž-emь*	*syn-ъmь*	*žen-**ojǫ***	*zemĺ-**ejǫ***	*kost-ьjǫ*

		***ŏ*-stem**	***jŏ*-stem**	***ŭ*-stem**	***ā*-stem**	***jā*-stem**	***ĭ*-stem**
Plural	**nom.**	*rab-i*	*mǫž-i*	*syn-ove*	*žen-y*	*zemĺ-**ę***	*kost-i*
	acc.	*rab-y*	*mǫž-ę*	*syn-y*	*žen-y*	*zemĺ-**ę***	*kost-i*
	gen.	*rab-ъ*	*mǫž-ь*	*syn-ovъ*	*žen-ъ*	*zemĺ-ь*	*kost-ьji*
	dat.	*rab-omъ*	*mǫž-emъ*	*syn-omъ*	*žen-amъ*	*zemĺ-amъ*	*kost-ьmъ*
	loc.	*rab-ěhъ*	*mǫž-ihъ*	*syn-ъhъ*	*žen-ahъ*	*zemĺ-ahъ*	*kost-ьhъ*
	instr.	*rab-y*	*mǫž-i*	*syn-ъmi*	*žen-ami*	*zemĺ-ami*	*kost-ьmi*

Table 3. The assumed intraparadigmatic homonymity resulting from the post-OCS sound changes in Balkan Slavic. Texts in bold show the “confusion of the nasal vowels” and contraction; underlining indicates the merger of /i/ and /y/.

The sound changes involved a) the merger of the (unrounded) high central /y/ into a high frontal /i/, b) the contraction of the word-final V_1jV_2 sequence into V_2, and c) what is called the confusion of the Proto-Slavic nasal vowels **ǫ* and **ę* (Bg. *smesvane na nosovkite*, Ger. *Nasalwechsel*). The *y–i* merger is not limited to Balkan Slavic, in fact, the original distinction between the two sounds was lost in almost all Slavic languages. The resulting syncretism in the *ŏ*-stem plural forms between the nominative, accusative, and instrumental was, however, partly resolved in most varieties of Slavic by analogical leveling, generalizing the bisyllabic instrumental plural desinences (or dual as in BCMS; see **2.1**) of the other declensions. This analogy probably also took place in Balkan Slavic before the loss of case inflection, because manuscripts from the 12th century already show the diffusion of the *ŭ*-stem desinence -(ъ)*mi* to the *ŏ*-stem nouns (Mirčev 1978, 168). Thus, the homonymy between the nominative/accusative plural and the instrumental plural clearly did not derail the case system. Interestingly, this particular syncretism is in fact tolerated by modern Slovene and Czech, for example, which did not acquire analogical forms for the instrumental plural, but retained the homonymic endings. At the same time, the resulting nominative–accusative homonymy in the frequent *ŏ*-stem declension may be seen as detrimental to the case inflection, since, as discussed in **2.1**, only Balkan Slavic failed to develop a differential marking for at least a subset of

the accusative plurals, unlike the other Slavic languages with the exception of the marginal plural case form in one Rhodopian dialect, shown in example **(9)**.

Much of what was said above also applies to the word-final contraction resulting in b), accusative–instrumental homonymy between the *ā*-stem singular forms: The accusative–instrumental syncretism is tolerated for *ā*-stem singulars in Slovenian, which preserves all Proto-Slavic case distinctions. However, c), the "confusion of the nasal vowels," involves a radically more complex set of problems. The terms confusion or mixing seem, in effect, contradictory, since nearly all varieties of Balkan Slavic preserve the distinction between the reflexes of the two Proto-Slavic nasal vowels, although they are realized in very different ways, with some peripheral dialects even preserving a nasal element, such as a nasal consonant in the reflexes of the nasal vowels. However, the predominant reflex of **ę* is *e*, whereas **ǫ* is realized mostly as *ə*, as in standard Bulgarian, or as *a*, as in standard Macedonian. What is in fact meant by the "confusion" is often the attested failure of the Balkan Slavic post-OCS texts to use the Cyrillic graphemes <ѫ> back nasal and <ѧ> front nasal, standing in OCS for the nasals *ǫ* and *ę*, in an etymologically correct manner (Ivanova-Mirčeva & Haralampiev 1999, 63). This vacillation in the use of the nasal graphemes is exemplified by the differences in the three apostle manuscripts from the late 12th and early 13th centuries, studied by Klaus Steinke (1968, 36–37). The earliest of these manuscripts, the *Ohrid Apostle* form modern-day Macedonia, displays a new distribution of the graphemes, as <ѧ> is used after palatal (or soft) consonants and <ѫ> in all other positions. In the other two manuscripts, according to Steinke, the use of the nasal graphemes is rather arbitrary, although one of them occasionally uses <e> instead of <ѧ>, as would be expected on the basis of OCS.

A earlier dominant view accounted for the vacillation between the nasal vowel graphemes by assuming a merger of the two sounds. For example, Kiril Mirčev (1978, 110–117) proposes that the two nasal phonemes became oral and finally merged into a schwa-like vowel, similar to the pronunciation of the jers (for the jers, see **2.2**), a process finalized during the 13th and 14th centuries. Mirčev explains the fact that the majority of the modern Balkan Slavic varieties continue distinct reflexes of the two Proto-Slavic nasal vowels by assuming that the front nasal remained "palatalized," that is, the difference between the sounds was preserved by the palatal pronunciation of the preceding consonant. Blaže Koneski (1983, 40) is critical of Mirčev,

noting, that many western (that is, Macedonian) dialects of Balkan Slavic show outcomes such that to assume a shared development involving a merger of the nasal vowels is contrived. Koneski's criticism regarding the extent of the phenomenon is supported also by Boryana Velcheva (2014, 59). In Koneski's view, the textually observed "confusion of the nasal vowels" reflects complex and dialectally divergent developments that do not allow for straightforward conclusions about the spoken reality. In a similar vein, Ivan Haralampiev (Ivanova-Mirčeva & Haralampiev 1999, 63–68) proposes an alternative chain of events that does not presuppose the temporary merger of the two nasal vowels.

Many crucial problems regarding the destiny of the Proto-Slavic nasal vowels in Balkan Slavic remain, alas, beyond the scope of this study. However, as Jussi Nuorluoto (2012a, 454) suggests, the nasal vowels in a part of the varieties of MBS were probably affected by the phenomenon of timbre correlation, the preceding palatal consonant affecting the pronunciation of the following back vowel by giving it a front timbre or coloring. Although this may have been true only for nasal vowels following historically palatal consonants, the confusion of the nasals in those dialects where it is relevant could be explained, at least in part, by the timbre correlation, as it could have brought closer together the pronunciation of the reflexes of *ę and *ǫ. In the case of the *jā*-stem, this could have produced the kind of syncretism presented in **Table 3**. In fact, in the 14th century *Chronicle of Manasses*, in which the scribes sought to use the nasal graphemes in an etymological manner, the confusion of the nasals in the nominal inflection is observed only after the historically soft consonants (Šin 1996, 145–164). Unfortunately, the contexts in which *ę might have been observed in today's dialects following a palatalized consonant were usually found in the inflected endings, many of which were lost or altered by analogy. This lack of context is the result of a process of hardening, or dispalatalization, which affected the previously palatalized consonants over

a long period of time with varying dialectal outcomes (see e.g. Koneski 1983, 46–57).[63]

The question of whether the *jā*-stem accusative and the instrumental singular became homonymous with the genitive singular and with the nominative and accusative plural, as **Table 3** suggests, is complicated by the fact that nouns belonging to the soft *jā*-stem declension (and the *jŏ*-stem declension as well) begun increasingly to use the hard-stem endings, a development observed in the MBS manuscripts (Mirčev 1978, 171). In **2.2.2** I propose a development that possibly deviates from this widely-held view of the dominance of the hard stems; the dialects having the dative form like *čerki* may point in opposite direction, thus casting doubt on the truism of the unidirectionality of the influence between the soft and hard-stems.[64] Yet if we adhere to the idea of the dominance of the hard stems, it is not clear whether the sequence of this process should be thought of, first, as the hardening of the stem-final consonant and, second, as applying the hard-stem endings, or rather simply as an analogical leveling independent of the changes in pronunciation in the stem-final consonant. This is particularly relevant, as there are indications that the hardening did not occur uniformly in all dialects or at the same time for all soft consonants (see e.g. Koneski 1983, 41). What is important is that this change helped, at least to some degree, to maintain the case distinctions in the originally *jā*-stem nouns, even if we assume a merger of the nasals after the soft consonants.

In addition to the sound changes mentioned here, Balkan Slavic is characterized by vowel raising in the unaccented syllables, with the exception of the westernmost varieties (including standard Macedonian). This vowel reduction phenomenon (in the broad sense of the term) involves the following approximate changes, although the phonetic realization of the

[63] However, there are some contexts in which the reflex of *ę is found after *j*. In the western Balkan Slavic varieties, as in standard Macedonian, a prothetic or epenthetic *j* may precede *ę, and, indeed, both display the same reflex as *ǫ: Mac. *jazik* (OCS *językъ*) 'tongue, language' and *zajak* (OCS *zajęcь*) 'hare' (cf. Mac. *pajak* < Proto-Slavic **pajǫkъ* 'spider'). Standard Bulgarian reflexes are *ezik* and *zaek*. There are no indications of the nasal merger after the unpaired soft consonants *ž* and *š*, for example, which may indicate that the merger may have been limited to contexts with certain soft consonants. However, there are dialectal differences here too, since in the Ohrid-Prespa dialects one finds *čədo* (OCS *čędo*) 'child', whereby the merger has occurred after the unpaired soft affricate (Mladenov 1979, 125).

[64] BCMS is used as an example of one of the stem types dominating and therefore suggesting a similar development also in Balkan Slavic (see e.g. Sobolev 1991, 29). However, with the BCMS *ŏ*-stem nouns, the only diagnostic ending is the accusative plural, whose preference may stem from the need to have a distinct accusative form, a process also characteristic of other Slavic languages (see **2.1.2**).

raising varies among the dialects: *o* → *u*, *e* → *i*, and *a* → *ə* (see e.g. Vidoeski 2005, 109). André Vaillant (1958, 23) suggests that this development caused the "muddling" of a number of case endings. Here, the chronology of the changes in the sound system is crucial, since, for example, the *ā*-stem singular nominative would have become homonymous with the accusative (and the instrumental after the contraction) only after the centralization and loss of nasality of the Proto-Slavic **ǫ*. Nevertheless, the vowel reduction is found relatively early in some manuscripts. Mirčev (1978, 144–147) observes the certain signs of the *o* → *u* raising in manuscripts from the 12th century on, the *a* → *ə* reduction from the 13th century, and the *e* → *i* raising from the 15th century on. An interesting question is whether the frequent -*o* instead of the expected dative singular -*u* of the singular neuters (see **2.2.3**), observed in the MBS manuscripts, may have been the result, not only of the omission of the last part of the digraph <oy>, but also of the vowel reduction, rendering the two sounds homonymous.

It seems that the overall effect of the sound changes remained rather limited. However, two pieces of evidence support the crucial role played by the sound changes in the singular forms of the *ā*-stem nouns. First, a number of Balkan Slavic dialects that retain a reflex for the Proto-Slavic **ǫ*, different from the reflex of **a*, also preserve (or preserved until modern times so that they could be attested by dialectologists) a distinct accusative of the *ā*-stem singular nouns, which functions as the oblique case (on these dialects, see **2.2.2**). The dialects are mainly found in the Rhodopes and in the transitional dialects between Balkan Slavic (Bulgarian and Macedonian; see **1.3.1**) and Serbian. Given the widely preserved ancient masculine genitive–accusative desinence -*a*, it seems that the persistence of the accusative singular as the oblique case may, indeed, have depended on the availability of a distinct, inherited form. The destiny of the accusative is discussed further from a wider Balkan perspective in **3.3.4**. Second, as proposed in **2.2.2**, the frequent replacement of the *ā*-stem genitive singular -*y* with -*ǫ* of the accusative may be the result of an analysis of the accusative ending as the marker of the genitive as well, which may indicate that *y*, which merged into *i*, was discarded as the genitive marker after this sound change. The new form, now homonymous with the *jā*-stem dative, continued in several dialects as the marker of the dative, which later also replaced the transitory genitive–accusative as the adnominal case.

2.4 Conclusions: The Middle Balkan Slavic case system

All researchers who have sought to examine the development of the Balkan Slavic case system have used texts that precede the first clearly vernacular texts. While these are the only direct sources we have, it is harder to interpret them than researchers often seem to believe. One key difficulty is that morphosyntactic variation can be interpreted in many ways: There is always the possibility of the scribe's, the author's, or the translator's mistake, rather than an emerging new norm or re-interpretation based on interference from the spoken language. While a single exception from the observed norm is never conclusive, with a limited amount of data it can, nevertheless, be tempting to give such an exception as an example if it occurs in a place where variation would be anticipated because of a change in the spoken language.

What can be said about the Middle Balkan Slavic corpus is that, in almost all instances where vacillation is expected on the basis of changes in the case system of the spoken language, this vacillation is found. Yet in search of the first attestation, scholars neglect to report the frequency of the mistakes. In my view, it is reasonable to believe that some signs of the case loss can be assumed to have been apparent in the 12th century. However, I do not think that much earlier datings for the beginning of the loss of case inflection are wise: The examples are too sporadic and the assumed lost distinctions could still have been amended by analogies, as in other Slavic languages. However, from the 12th century on, there is no reason to assume that the re-interpretation of a case function must have encompassed either the whole paradigm, including all words—regardless of their declination class, gender, or number—or just any context. As was shown earlier, the *ā*-stem singular nouns and the neuters behave very differently from the *ŏ*-stem masculines. Moreover, as the known grammaticalization processes show, the eventual change is often preceded by a phase of competing structures. Although the vacillation in the use of cases in the MBS corpus is, in my view, often credibly attributed to actual changes in the spoken language, the replacement structures, most notably the grammaticalization of the *na*-construction, are too hastily dated, even centuries earlier than it is wise to assume (see below).

In light of the key observations of this chapter, **Table 4** presents an intermediary case system of Balkan Slavic between the 13th and 16th centuries based on the MBS corpus, the first vernacular texts, and evidence from the dialects. Needless to say, this attempt is not intended to be a

reconstruction of any single case system, especially since it must be assumed that historically there was also great dialectal variation between the individual systems. Rather the table endeavors to show the key tendencies observed throughout the varieties of Balkan Slavic. The forms that with great certainty were no longer present in a significant number of dialects at the end of the period are put in brackets. The forms that were not distinct from the nominative are given in grey.

		masculine	**neuter**	**feminine**	**feminine**
		‘son’	‘forehead’	‘woman, wife’	‘winter’
sg.	**nominative**	*sin*	*čel-o*	*žen-á*	*zím-a*
	oblique	*sin-a*	*čel-o*	*žen-ə́*	*zím-a*
	dative–genitive	*sin-u*	*čel-o*	(*žen*-[*í*/*é*])	(*zím*-[*i*/*e*])
pl.	**nominative**	*sinov*[*e*/*i*]	*čel-a*	*žen-í*	*zím-i*
	dative–genitive	(*sinov*-[*om*/*em*])	*čel-a*	(*žen-ám*)	(*zím-am*)

Table 4. The Middle Balkan Slavic case system.

Before concluding this chapter, a few remarks on the development of the *na*-construction, presented in several subsections, are in order. With regard to the MBS manuscripts reflecting the grammaticalization of *na* + oblique case, two distinct schools can be observed. On the first hand, Meyer, Rusek, and Duridanov believe that the texts exhibit an already relatively well-developed analytical structure, homonymic to the dative case, in the spoken language of the 13th century. On the other hand, Mirčev and Lunt, for example, claim that the manuscripts do not offer proof to support the early emergence of the *na*-construction nearly as clearly as the others suggest, although they do not categorically deny this possibility. It is significantly easier to align with the latter group of scholars: in many instances the examples are suspect for various reasons, and, perhaps more important, the more compelling examples are extremely few.

However, from the meta-analysis of the previous research, certain observations emerge. The core function of the dative, the expression of the recipient, does not show any credible signs of the *na*-constructions spreading there; neither does the marking of addressees of communication. The increase in the use of constructions with *na* is apparent only with abstract verbs, exhibiting vacillation in their complements as also appears in OCS. In addition, the Greek source texts often have various constructions with the same verbs, and these choices are reflected with varying degrees of rigor by the translators. Indirect evidence for an increase in the distance between the spoken language of the scribes and their written target language

is provided by the greater variation between different structures and the increase in the use of certain formulaic patterns found in the manuscripts. Yet the vacillations or the formulaic expressions do not support or discredit any particular hypothesis about the reorganization of the Balkan Slavic case system. Nevertheless, while not yet emerging as a replacement for the dative, the preposition *na* with its increased use in abstract senses in the MBS corpus is indicative of its semantic bleaching, a necessary condition for its more grammatical role.

It is most often hypothesized that the syncretism of the marking of the adnominal possessor and the recipient in Balkan Slavic was attained through the constructions with *na* first replacing the dative. The following step involved the development of the dative as the default adnominal case, a process whose beginnings are already witnessed in OCS. While this certainly is the main storyline, it is, however, important to note that a significant number of examples with *na*-constructions replace the expected dative in nominalized structures (see **2.2.3** *Other uses of the dative as verbal complement*). Biblical Greek uses nominalization copiously through participles, but also through compounds that involve infinitives and deverbal nouns. These are usually translated into Slavic with corresponding non-finite structures. Cross-linguistically, deverbalization characteristically involves the use of the default adnominal case, the genitive, since in languages inflected for case, the patients when rendered adnominal often cannot appear in the accusative, the case of the direct object. The increase in the use of *na* marking the patient of abstract deverbal nouns in adnominal constructions is, therefore, potentially significant: Nominalization seems to provide a shortcut for the construction to the adnominal position, without first supposing the replacement or even a significant weakening of the dative in its core roles in finite constructions. In addition, if we assume that the locative was replaced early on by the oblique case, there are no limitations on the two ways, namely, the replacement of the dative and the marking of the patients of the nominalized structures, both of which sped up the grammaticalization of the *na*-structure.

***

In this chapter the internal processes of the development of the Balkan Slavic case system were discussed. I have demonstrated that, although the sound changes were not the only explanation, they did have an impact on the expression of several key cases in frequent declensions. However, there are

two issues that characterize the Balkan Slavic case loss, but which cannot credibly be attributed to internal causes: First, there are developments that are not explained by the sound changes or pre-existing syncretism, as these did not have the same effect on the other Slavic languages. These include the emergence of one case or one analytical structure to replace the ancient genitive and dative, the full merger of the goal and location functions—even in constructions that preserve the ancient locative—and the merger of the instrumental and accompaniment functions. Second, the inability of Balkan Slavic to develop new case forms through analogy to replace the lost distinctions is indicative of a typological change: The inflectional strategy in case marking did not survive, but was readily replaced by analogical constructions employing an uninflected form. I believe that in order to explain the two types of phenomena plausibly, the role of language contacts must be taken into account. To do this, in the following chapter I will address the development and outlines of the case systems in the other Balkan languages belonging to the contact linguistic phenomenon known as the Balkan linguistic area and present some general characteristics of the effect of language contact.

3 The Balkan case systems

The purpose of this chapter is to establish an understanding of the similarities and differences among individual case systems in the linguistic contact phenomenon called the Balkan linguistic area. This will be done by providing an areal point of comparison for the diachronic development of the Balkan Slavic case system. After an introduction to the contact linguistic phenomena relevant to the convergence of the Balkan languages and the loss of case inflection in Balkan Slavic (**Section 3.1**), I will map the synchronic case systems and their development in Albanian, Balkan Romance, and Greek (**Section 3.2**). My intention is to follow the diachronic development within each language in a search for possible paths of borrowing or influence among the languages. A characteristic problem throughout this chapter is that the cases are not easily comparable cross-linguistically, despite occasional coinciding terminology, such as identical names for cases in different languages. This is because each grammatical case reflects language-specific phonological forms, syntactic properties, and functions. Contrasting the cases thus requires a functional approach, which at the same time calls for a wider analysis of the expression of grammatical relations. **Section 3.3** therefore will address some key issues of argument marking that reach beyond mere case inflection. As a conclusion, **Section 3.4** presents concisely the similarities among the individual Balkan case systems and evaluates the contact-induced elements in their development.

3.1 Language contact and areal linguistics

Throughout this volume, the discussion of the theoretical dimensions of language contact is divided into two parts. The notions of both a linguistic area and of contact-induced simplification are discussed in subsections **3.1.1** and **3.1.2**, whereas **Section 4.1** extends the debate to the concept of areality within linguistic typology.

3.1.1 Language contact and convergence: The Balkans as a linguistic area

The term Balkan linguistic area[65] (also called linguistic union, linguistic league, or *Sprachbund*) is used in this volume to denote a group of linguistic varieties spoken mainly in the Balkans. These languages are hypothesized as displaying a number of shared features thought to have been diffused through language contact. The type of contact that occurred can best be characterized as intense, intimate, and sustained, as Friedman and Joseph (2014, 16) describe it, or, using Lindstedt's (2000, 242) term, as mutual reinforcement. What is meant by mutual reinforcement is that the conditions of the contact contributed to the types of features diffused, and further, that it is difficult to point out the source of a particular shared feature, a situation that could best be characterized as convergence among the varieties involved.

Several general accounts on the Balkan linguistic area are available, although modern handbooks are still few (for a concise survey of the literature, see Friedman 2011; see also Cyhun 1981; Solta 1980). The languages traditionally associated with the core of the Balkan linguistic area are Albanian, Bulgarian, Macedonian, Greek, Romanian, Aromanian, and Megleno-Romanian. In addition, there are several other linguistic varieties that have been shown to display features common to the region. These include Bosnian, Croatian, Montenegrin, and Serbian (BCMS), especially those varieties spoken in southern Serbia and Kosovo, together with Romani and Balkan Turkish.

A general definition of a linguistic area as a term coincides more or less with the above definition of the Balkan linguistic area, not least because of the history of the concept (see below). However, this definition is not particularly comprehensive and consequently leaves many problematic questions unanswered. Below, I will address some of the more challenging issues related to the concept of a linguistic area in general and that of the Balkans in particular. However, the introduction of areality as a new variable in modern approaches to linguistic typology has presented some challenges to stable concepts such as the Balkan linguistic area. These challenges—as well as their possibilities—are addressed in **Chapter 4**,

[65] Not to be confused with language area (Sprachraum), a term that refers to a continuum of genealogically-related linguistic varieties. "Linguistic area" is a translation of *Sprachbund* by H.V. Velten in 1943 (Campbell 2006, 3).

which, as a whole, explores novel ways that could benefit areal linguistics and the study of the Balkan linguistic area.

The Balkan linguistic area: A history of the notion

Nikolai Trubetzkoy postulated in Russian the concept of *jazykovoj sojuz*, that is, a *Sprachbund* or linguistic area or union in 1923 (Campbell 2006, 3). Although there are useful accounts of the origin of the term (see, e.g., Campbell 2006; Tosco 2008), some of its aspects are highlighted here. With the groundbreaking discovery of the genealogical relatedness of languages in the early 19th century and the emergence of the field of comparative linguistics, similarities observed between languages could be used to introduce hypotheses about the relatedness of languages and the hypotheses could then be verified using the comparative method. Growing linguistic data, however, brought new questions about similarities among languages that could not be explained by a common origin. A new term was needed within the heuristic concept of a “language group” (*Sprachgruppe*), first and foremost to distinguish between a language family (*Sprachfamilie*), which was based on genealogical inheritance, and a “linguistic union” (*Sprachbund*), which was defined above all by the absence of a genealogical component.

Although not much was known about areas that displayed shared linguistic features that could not be explained genealogically, Trubetzkoy’s influential proposal (1928, 18) in the first international congress of linguistics made a rather strong claim about the languages forming a *Sprachbund*: The languages in such a union greatly resemble each other with regard to syntax and principles of morphological structure, and they share a great number of cultural words, as well as sometimes superficial similarities in the sound system. Trubetzkoy also observed that a *Sprachbund*, unlike a language family, does not display systematic sound correspondences that match the phonological form of morphological elements, nor is there a shared basic vocabulary among the languages involved. Trubetzkoy’s definition coincides in a very timely way with Kristian Sandfeld’s call for a special field of study, *balkanfilologi* (1926, 8, 10, my translation):

> Now when all these languages [Greek, Albanian, Bulgarian, Serbian, Romanian, and Turkish] are so different in origin and thus each of them has its own, especially phonologically and morphologically sharply marked, structure, how is it justifiable to treat them as one unit and make them into a starting point of a comprehensive study? […] When one proceeds from one language to another—excluding Turkish—in many cases it is actually only the inflected forms and vocabulary that differ, while the manner of expression stays the same, so that one is of the lively impression that these languages are, as mentioned earlier, animated by the same spirit.

As the quote from Sandfeld shows, there is a fundamental problem with the original definition of *Sprachbund* or linguistic area. Trubetzkoy's *Sprachbund* is essentially the Balkan linguistic area or how it was then perceived, because by the year 1928, the only linguistic area that had been studied in any detail was the Balkan.[66] Therefore, seeking to identify linguistic areas in Trubetzkoy's sense carries a risk similar to the Eurocentric bias of linguistics: the definition is based on a well-known set of data that is also assumed to be representative of the situation in other parts of the world. Although the definition is perhaps intended as a diagnostic tool, it may well fall short when more data become available. Also, there is a certain vagueness surrounding the definition, since the proposition only gives similarities, adding that the languages "greatly resemble" each other. However, despite the shortcomings of the original definition of a linguistic area and the risk of circular reasoning therein, the term "linguistic area" is used extensively beyond the Balkans, proving that some aspects of the original idea remain useful. The expansion of the field of areal linguistics has meant that several clarifications and additional criteria have been added, regarding both linguistic areas in general and the Balkan linguistic area in particular.

Numbers and types of shared features in linguistic areas

One of the key problems in the case of the Balkan linguistic area as described by scholars has been that individual features encompassing all Balkan languages seem to be rather few. This leads to several problems. In comparative linguistics, defining a language family is relatively straightforward, the most important criteria being sound correspondences in words among the hypothesized daughter languages and a relatively small

[66] As early as 1829, the Slovenian philologist Jernjej Kopitar observed of Bulgarian (Balkan Slavic), Wallachian (Balkan Romance), and Albanian that among these languages "nur eine Sprachform herrscht, aber mit dreyerley Sprachmaterie" (K. 1929, 86). For what is perhaps an even earlier mention, see Friedman 2011, 2.

shared vocabulary, which can suffice to prove a genealogical relationship, since the probability of a similar phenomenon existing by chance or borrowing is negligible. Additional, but not conclusive, evidence can be found in shared typological features. By contrast, a linguistic area is characterized by the lack of cognate words unless the languages are related; hence, for the most part a linguistic area relies on similarities in grammar.[67] The original definition of linguistic area, however, lacks a description of the exact nature of these similarities, and thus inherently the defining characteristics are on a weaker empirical basis to begin with and clearly require more qualitative criteria.

Modern scholars of the Balkan linguistic area portray the shared features in much the same way as Trubetzkoy did. The table of contents in a comprehensive study by Petja Asenova (2002) lists the language areas where Balkanisms are found, with morphophonology occupying the most space followed by vocabulary, and, as “predicted” by Trubetzkoy, with the chapter on phonology being by far the shortest. Nevertheless, in most accounts the individual shared features of the linguistic area are perhaps surprisingly few. Olga Mišeska Tomić (2006, 26–27) gives prominent morphophonological features of the Balkan linguistic area as treated by seven individual authors. These authors agree on only three features, namely, the post-positive article, the loss of the infinitive, and a future tense with an auxiliary verb meaning ‘to want’.[68] Several listings include or imply at least two further central features that are relevant for the study of the loss of case inflection in Balkan Slavic and its intermediary stages: the dative–genitive merger, discussed in **4.2**, and a general tendency toward expressing grammatical relations through analytical means (see **3.1.2** and **3.3**).

The problem which scholars of the Balkan linguistic area face is that any list of features that tries to account for the linguistic area as a whole remains surprisingly meager, despite a scholar’s often very clear experience of similarity among the languages. Some scholars from outside the field of Balkan linguistics have acknowledged the same problem. Balthasar Bickel and Johanna Nichols (2006, 3–4) observe that the list of classic Balkanisms indeed does not describe the shared grammar very well (for more discussion on this view, see the outset of **Section 4**). The key question then is what

[67] See **4.1.2** for further discussion on the problematic role of typological similarities in determining genealogical relatedness.
[68] Among these authors, the greatest number of features—twelve—is given by Jouko Lindstedt (2000). However, Lindstedt’s research is directed to finding an epicenter of the linguistic area and thus includes features not shared by all languages under study.

does the "shared grammar" consist of if not of clearly distinguishable, all-encompassing features.

Partly owing to the above problems, the term "convergence" is frequently used to characterize the Balkan linguistic area. However, Bernd Heine and Tania Kuteva (2006, 286–287) are critical of this term. The overall impression is that their objection arises because they are promoting a model for contact-induced language change, which is supposedly responsible for a great majority of instances and which assumes a very clear dichotomy between the source and the receiving languages. The term convergence does not imply that kind of contact situation, of course. Yet the concrete reason these scholars give for opposing the use of the term is that language contact does not invariably lead to structural equivalence, a stance with which I could not agree more. However, this is not what is generally meant by convergence. Despite its origin as a word describing movement, convergence is used to illustrate the observed result of language contact, not its exact mechanism in the sense that all interaction between the involved languages must lead to similarity. Therefore, a process that leads to a language contact effect that does not increase similarity between the languages does not amount to convergence.[69]

The fuzziness involved in the notion of linguistic areas leads Lyle Campbell (2006, 459) to end his account of areal linguistics with a pessimistic opinion of the term's viability. For Campbell, there is no such thing as a linguistic area, and the observed phenomena, such as the Balkan linguistic area, accounts at most for "an accumulation of individual cases of 'localized diffusion'." He adds that "[a] linguistic area, to the extent that it

[69] The stance of these scholars is symptomatic of their theoretical framework (see Heine & Kuteva 2005; 2006), which deals more with the assumed mechanics of language contact than with its results. Their concept of contact-induced grammaticalization, intended as a model for a typical mechanism responsible for grammatical language contact phenomena, is often rigid and counter-intuitive. For example, in the case of borrowing a grammatical structure, the speakers of the receiving language are assumed to be able to replicate the historical grammaticalization process of the source language. In my view, this seems highly unlikely, since it would require the speakers to possess information about synchronically opaque historical processes in the source language. Nor does the model account for sociolinguistic factors involved in language contact, discussed in **3.1.2**. In addition, it emerges from their model that the source language always grammaticalizes the structure the farthest. Thus, it would follow that the level of grammaticalization of a given feature should serve as a tool with which to determine the source of the feature. If this were true, it would be ideal for the Balkan language area, where the source of the shared feature often remains unclear. However, the model fails in the Balkans; consider, for example, the various pathways of the grammaticalization of the dative–genitive case (see **3.2.1**, **3.2.2**, and **3.2.3**). The result of the dative–genitive merger is best described precisely as convergence, and the only way to account for its emergence is an analysis that combines historical, dialectal, and sociolinguistic data.

may have a legitimate existence at all, is merely the sum of borrowings in individual languages in contact situations." I agree with one of Campbell's basic points of criticism: there is a risk in using rigid, geographical, *a priori* assumptions. Nor I do find Campbell's description of the accumulation of localized diffusion faulty. The problem with Campbell's criticism is that it leads to a dead end or, in effect, to the starting point of the whole debate: What do we then call those areas that seem to have a high occurrence of instances of "localized diffusion"? I see no reason to insist that the borders of an assumed linguistic area need to be clear cut. Yet it is absolutely necessary to take into account areality when studying the diachrony or variation of a linguistic feature if the geographical area or the languages involved are known to display a great deal of contact-induced convergence. It is therefore a matter of preference whether we call that reasonable presupposition a linguistic area or something else.

Brian Joseph (2010, 628–629) , echoing Eric Hamp's (1989, 47) concept of a "spectrum of differential bindings," gives what is perhaps the best explanation for the discrepancy between the limited number of features shared by all of the core languages in the Balkan linguistic area and the impression that the languages have a shared grammar. There is both localized and broadly realized convergence:

> What we really have then is *clusters* of convergent languages, where the convergence is on various features in various locales. A sprachbund—the Balkan sprachbund in this case—thus is really to be defined as a cluster of such clusters […].

This view, in fact, is rather in line with Campbell's "accumulation of individual cases of 'localized diffusion'," although Joseph is less shy about giving the phenomenon a name. Joseph further describes the broadly realized convergence as essentially local diffusion, which, given enough time and the right sort of contact at the relevant "edges" of the locales, can cause a "diffusionary chain reaction." Perhaps then what characterizes the Balkan linguistic area best is that it formed, and to some extent still forms, an area that enabled the spread of innovations among linguistic varieties more freely than in some other parts of the world.

I argue that the main problem of classic Balkanisms is that first, they are individual features, and second, they are defined very broadly to encompass all the languages involved. A major drawback to such an approach is that it fails to account for the minor similarities of grammar that are found in the vicinity of the major features. The more localized features, pointed out by

Joseph, may still be closely related to a major feature. For this reason, an attempt is made in **3.3.4** and **4.2** to include some related minor phenomena in analyzing the acknowledged major features. This approach is taken partly in the spirit of Howard Aronson's (2007) suggestion to give functional criteria a more central role in the study of the Balkan linguistic area. Also, despite their similarities, the Balkan linguistic area languages are fundamentally different in some of their features, such as the basic word order of the NP. As will be argued in **3.3.3** and **4.2.2**, the result is that full structural parallelism cannot always be the expected outcome of language contact unless the contact also affects the fundamental differences.

In **4.2.6** it is noted that purely syntactic criteria may indicate a wider areal diffusion of a feature than criteria that take into account finer differentiation on the level of semantics and morphosyntax. Finer differentiation produces a patterning that corresponds more closely to the Balkan linguistic area. Yet, for this very reason, Aronson's view that functional phenomena should be given preference over morphosyntactic similarities must be regarded with some caution. The effect of different criteria raises another important question, namely, whether there is some special qualitative flavor to the linguistic areas. In other words, are features diffused within a linguistic area somehow different from other contact-induced phenomena, perhaps giving added legitimacy to the notion of a linguistic area? This discussion will be continued below. A related set of questions involves the typological frequency and the co-occurrence of features. A typical argument against the contact-induced origin of features of the Balkan linguistic area involves the observation that a particular feature is not cross-linguistically rare. Therefore, the following question must be asked: if a feature or the co-occurrence of two shared features thought to have resulted from language contact is typologically common, then is that feature less useful as evidence in support of a language contact hypothesis? These questions will be elaborated on below in **Subsections 3.3.4**, **4.1.2**, and **4.2.5**.

Historical and sociolinguistic factors leading to convergence

Joel Sherzer's volume *An areal-typological study of American Indian languages north of Mexico* (1976) sparked an interesting debate in areal linguistics. Sherzer claims to show areal convergence on the basis of his data from North American languages. His approach called on critics to define the field of areal linguistics and the study of linguistic areas as requiring a diachronic component in the analysis. The criticism with which

Sherzer's work was met was twofold: First, his approach did not exclude the chance of a genealogical component as a source for similarities. Second, the linguistic area hypothesis was not based on a credible historical contact situation. (For Sherzer's critics, see Hamp 1977; Campbell 1985.) In the same vein, demands for a historical component in the study of the Balkan linguistic area have been made, for example, by Brian Joseph (2008, discussed in further detail in **4.1.1**).

Because of these developments, the role of diachronic and historical sociolinguistic approaches in areal linguistics is today a significant one. For example, Sarah Thomason (2008, 49) gives five steps (abridged here) to be taken before a claim of contact-induced change can be considered firmly established: 1) The languages should be looked at as a whole; if structural interference of some kind has occurred, it is highly unlikely to be an isolated instance. 2) Identify the source language and show that the contact was sufficiently intense. 3) Identify shared structural features in the proposed source language and in the receiving language. 4) Prove that the feature did not exist in the receiving language prior to the proposed contact. In the event there is no historical evidence, examine the languages related to the recipient language. 5) Prove that the proposed feature was present in the source language before it came into contact with the recipient language. As can be seen, three out of five prerequisites directly involve diachronic or historical sociolinguistic tasks.

In the study of the Balkan linguistic area, the main theories about the historical developments that led to convergence among the languages are similar to those presented in **1.4.1** for the loss of case inflection in Balkan Slavic. Previously, substratal theories were among the most popular. The key problem with these approaches, as discussed earlier, is that lack of knowledge of the structure of the Proto-Balkan languages prevents any of the hypotheses from being substantiated. Naturally, for the same reason such theories cannot definitely be ruled out either. However, I believe that the need to attribute significant changes resulting in convergence to a considerable language shift is based on a limited and outdated understanding of the dynamics of language contact. Even if the convergence among Albanian, Balkan Romance, and Greek were to have resulted from a shared substrate, one would have to assume a different mechanism for the changes in the language of the later arrivals, such as the Slavs, who entered the Balkans in the wake of the invasion of the Avars in the mid-6th century (Hupchick & Cox 2001, Map 8). This raises the question of why would one

need a substrate explanation for the remaining languages if the diffusion of the features is possible without any such explanation.

Of course, one of the reasons for seeking an external source for the Balkanisms is that very few of the shared features can be pinpointed with certainty as originating from a particular language of the Balkan linguistic area. However, theories that highlight the role of one of the languages in the formation of the linguistic area are numerous (for some of these, see **4.1.3**). In his assessment of these theories, Jouko Lindstedt (2000, 235–238) concludes that, although some languages may seem more credible than others as the source of an individual feature, a single language cannot be found. The reason is not, however, that we lack historical information. On the contrary, Greek and Balkan Slavic, for instance, are sufficiently documented in the relevant period. According to Lindstedt, the reason is rather that the sociolinguistic contact situation has caused changes that would not have occurred in the languages by internal drift. Therefore in many cases no source language exists in the traditional sense.

Lindstedt (2000, 239) lists the sociolinguistic conditions, characteristic of the Balkans after the Avar invasion in the 6th century, as follows:

1) Speakers of different languages live close together, often in the same villages.
2) There is no single dominant lingua franca.
3) Speakers of each language have sufficient access to the other languages they need.
4) Native languages are important symbols of group identity.

These conditions resulted in extensive, mutual multilingualism, but, because the languages served as an important source for group identity, no major language shifts took place. Furthermore, despite the intense contact, certain aspects of the Balkan languages, such as the retention of relatively complex inflectional verb systems, suggest that no extensive creolization took place (for the effects of creolization, see **3.1.2**). According to Lindstedt (2000, 241), the stable bilingualism led to a preference for explicit syntactic marking: “structural conflicts between the languages are solved analytically, by syntactic means, because cross-language identification between analytic structures is easier than between inflectional categories.”

Both Sarah Thomason (2001, 125) and Brian Joseph (2010, 625) agree on the decisive role of stable and institutionalized multilingualism in the formation of linguistic areas and in particular, in the Balkan linguistic area. Joseph (2010, 625–628) adds that imperfect command of L2 (the second language) may have played some role, as the L2 speakers would have used structures influenced by their L1. The L1 speakers also probably tried to

streamline their own use of the language to coincide more closely with that of the others, favoring structures that had analogies in the other language. This mutual accommodation led to increasingly convergent structures. Joseph further points out that, although loanwords may spread in relatively shallow language contact, the use of certain loanwords closely tied to discourse is indicative of continued and intensive face-to-face interactions in the Balkans.

An important observation that speaks for the role of L2 speakers is that the number of Balkanisms found in the individual languages seems to be tied to the status of the language on a prestige scale. According to Lindstedt (2000, 242–243), the fact that Greek displays the fewest number of shared Balkan features is indicative of its role as the language with the greatest prestige in the Balkans, which meant that it displayed the least amount of mutual bilingualism. On the other hand, speakers of Balkan Slavic, Albanian, and Aromanian, whose languages were more or less in the middle of the prestige scale, learned Greek, but also practiced a fair amount of mutual multilingualism. Romani, discussed in the following subsection, is distinct in the sense that, although its speakers were often multilingual, it had the fewest number of L2 speakers because of its status as the least prestigious of the Balkan languages.

3.1.2 Language contact and simplification: Resistant Romani

It has been argued that certain kinds of language contacts often bring about changes that can be characterized as simplification, such as loss of inflectional categories. Efforts to categorize the different types of contact and their effects are not new; terms like adstratum, substratum, and superstratum are often used to refer both to the sociolinguistic contact situation and to its result. Yet strong claims have increasingly been put forward to suggest that a significant number of second-language speakers may reduce the morphological complexity of a language. Some of the mechanisms leading to such loss of inflectional categories were introduced above, in **3.1.1**. Here, this hypothesis is discussed regarding the loss of case inflection in the Balkans, illustrated through the Balkan Romani case system, which retains significantly more case distinctions than Albanian, Balkan Romance, Balkan Slavic, or Greek. I will claim that the conservatism of Balkan Romani is explained by its special sociolinguistic status, which differs from that of the other languages in the Balkan linguistic area.

Romani: Retention of inflection

Romani is an Indo-Aryan language sharing the common Proto-Indo-European (PIE) origin with the Balkan languages, excluding Turkish. Although often mentioned in connection with the Balkan linguistic area, Romani is largely omitted from the present study with the exception of this subsection. The reason is based on the simple observation that the Romani case system is not comparable to those of the other Balkan languages in almost any degree. Nevertheless, Romani, the native language of many of the Roma people, does offer an interesting perspective on the loss of case inflection in the Balkans and on the overall debate about linguistic complexity and language contact. A common denominator of Romani varieties, widely dispersed around the world and often lacking mutual intelligibility, is their high susceptibility to influence from surrounding languages. This tendency to language change induced by contact with other languages is also evident in the Romani varieties spoken in the Balkans, yet without any apparent simplifying effect on their inherited case inflection.

Like the closely related Hindi-Urdu, Romani displays two historical layers of case markers: the old layer consists of the nominative and oblique cases (also called the accusative in some descriptions) and the morphologically case-like vocative. A second layer comprises four or five agglutinative case suffixes, using as their stems the oblique that derives from the Old Indo-Aryan genitive (Friedman 1991, 86; Matras 2002, 79). These affixes include the dative, locative, ablative, instrumental, and genitive, although the combining properties of the genitive differ significantly from other cases in the second layer.[70] Also, the cases in the second layer attach only to the noun, while the modifying adjectives and the definite articles display the oblique form.

[70] While many grammars straightforwardly list the "genitive" suffix as a case (see, e.g., Lee 2005, 124), the situation is more complex. First of all, the "genitive" has several features of a possessive adjectival derivational suffix: it agrees in number and gender with the head, and it is a productive way to derive new adjectives; see Hancock 1995, 72–73. In addition, it participates in *Suffixaufnahme*, i.e., it combines with other cases, which are always positioned after the genitive. However, Friedman (1991) argues for interpreting the suffix as a case. Most important, he demonstrates that, at least in some dialects, the genitive has important syntactic functions beyond marking the possessor (ibid., 95). The synchronic difficulties in the classification of the genitive, however, reflect diachronically an interesting grammaticalization process toward a genitive case, not stemming from directional or locational construction; see Matras 2002, 89. A similar process of grammaticalization of the genitive has also been proposed for the Uralic languages (Hakulinen 1979, 100). A comprehensive, cross-linguistically informed overview of the Romani genitive is found in Koptjevskaja-Tamm 2000.

Despite major differences in comparison with the other Balkan case systems, Romani shares a number of Balkan features, such as the definite article and the loss of the infinitive,[71] features that are absent from other Indo-Aryan languages. The Romani varieties differ greatly from one another, owing to the migration of their speakers from the Balkan Peninsula, where the language had arrived probably in the 11th century, reaching various parts of Europe by the 14th century (Matras 2002, 17–18). Yet the definite article and the loss of the infinitive encompass all Romani varieties and therefore must have taken place before the dispersion of Romani, between the 11th and 13th centuries. In addition to these Balkan features, which are indicative of intense language contact, several layers of loan words bear witness to the proximity of speakers of Romani to Greek, Balkan Romance, and Slavic (ibid., 20–22).

Yet the intensive language contacts did not lead to changes or case loss in the Romani case system. One possible explanation might be found in the relatively late arrival of Romani in the Balkans. Yet the comparatively briefer presence of Romani in the Balkan Peninsula vis-à-vis other Balkan languages serves at most as a partial explanation for its preservation of the Indo-Aryan case system. Contact-induced increase in analytism does not necessarily require particularly long periods of contact, as demonstrated, for example, by the destiny of the Balkan Slavic case inflection and the changes in Romani itself. However, the fact that Romani, owing to its specific sociolinguistic status, as mentioned in **3.1.1**, always had a very small number of L2 speakers might prove to be a much more fruitful explanation, as will be argued below.

Simplification in language contact: The number of L2 speakers

Since the dawn of the millennium, the study of linguistic complexity has received unprecedented attention in linguistics (for dedicated volumes, see, e.g., Miestamo, Sinnemäki & Karlsson [eds.] 2007; Dahl 2004; Hawkins 2003). Traditionally, many linguists have held that all languages are equally complex, a view that implicitly contains the assumption that the complexity of a language is maintained through a mechanism of complexity balancing or trade-offs among linguistic sub-systems. This outlook is expressed, for instance, by David Crystal (1987, 6), who wrote: “All languages have a

[71] According to Friedman (2000, 97), some dialects of Romani outside the Balkans are in the process of losing two Balkan features by developing a new infinitive and by giving up the definite article. For example, the Romani spoken in Finland has lost the definite article and displays infinitive complementation (Granqvist 2011, 146–147).

complex grammar: there may be relative simplicity in one respect (e.g., no word-endings), but there seems always to be relative complexity in another (e.g., word-position)."

Yet these views have remained mostly unsubstantiated (see e.g. Sinnemäki 2008, 68). Moreover, Kaius Sinnemäki (2011, 16–18) argues that the equicomplexity hypothesis is unfalsifiable, since it would require, first, devising a common metric for all aspects of a language's grammar and second, giving appropriate weight to, let us say, morphological and syntactic complexity. For the present as well as for the foreseeable future, these problems remain in principle unsolvable. Yet the equicomplexity hypothesis must be seen against the backdrop of earlier views that equated the assumed complexity and sophistication of certain, typically major Western languages with the "complexity," that is, the cultural, social, and technological progressiveness, of their speakers. Therefore, the notion of relative simplicity in characterizing an aspect of language is sometimes perceived, unnecessarily, to imply that the speakers too lack "complexity" or sophistication in one way or another.[72]

Although measuring linguistic complexity remains problematic, challenging such truisms as "all languages are equally complex" has led to an interesting debate about the effect of language contacts. Language contact as a source for simplification has been put forward and debated, for instance, in creole studies, in which John H. McWhorter's provocatively entitled paper "The world's simplest grammars are creole grammars" in *Linguistic Typology* (2001) sparked a heated discussion in the same issue of the journal, with 222 pages of commentaries. While creoles, the nativized mixed languages, are no doubt the most extreme form of language contact, similar claims for contact-induced simplification in other sociolinguistic situations have been presented by such scholars as Wouter Kusters (2003), Gary Lupyan and Rick Dale (2010), Christian Bentz and Bodo Winter (2013), and Péter Maitz and Attila Németh (2014).

Lupyan and Dale (2010) propose that, similar to biological organisms, languages too face certain evolutionary pressure to which they need to adapt. This analogy does not, however, go further than to suggest that the "ecological niche" to which the languages must adapt is characterized by the number of speakers, geographical spread, and amount of language contact to which they are subjected. These researcher claim, based on large data sets, that languages spoken by large groups of people in a large geographical area

[72] Cf. also 1.4.1.

and, more specifically, with many adult language learners, tend to have less inflectional morphology. In an attempt to test further Lupyan and Dale's Linguistic Niche hypothesis, Bentz and Winter (2013) specifically studied the role of L2 speakers. Their conclusion is that languages with a considerable number of L2 speakers generally display relatively small nominal case systems or no nominal case at all. Similar results have been obtained, based on a qualitative analysis, by Kusters (2003) regarding the amount of inflectional verbal morphology. In a recent study, Maitz and Németh (2014) find support for the L2 hypothesis, but they claim that the retention of complexity also depends on the speakers' attitude to the linguistic norm. A summary of similar findings regarding the L2 speakers' role in simplification is given by Peter Trudgill (2011, 15–20).

These views too have met with criticism. Seán Roberts and James Winters (2012, 94–96) have voiced doubts regarding aspects of Lupyan and Dale's (2010) methodology. Roberts and Winters warn, for instance, that large data sets often provide ostensible explanations as a result of accidental correlations. For example, they find a highly significant positive correlation between the geographical distribution of a species of acacia tree and the occurrence of tonal languages in the same region. Yet claiming causality between the two would clearly be absurd. Sarah Thomason (2008), on the other hand, challenges the role of L2 speakers as a straightforward source for reduction in complexity. With compelling examples, she demonstrates how L2 influence may in fact increase complexity by introducing new grammatical distinctions or even inflectional forms. Perhaps it is also relevant that Thomason claims to adhere to the equicomplexity hypothesis, bouncing the burden of proof back to its opponents. Most of her examples concern language change induced by language shift, which is likewise significant, since language shift is only one possible sociolinguistic context and, as argued before, is perhaps less relevant for the Balkan linguistic area.

All in all, the discussion regarding the effect of language contact on language complexity and, specifically, the role of L2 speakers in reducing inflectional morphology is highly significant for the question of loss of case inflection in Balkan Slavic. Based on the material presented in **3.1.1** (especially Lindstedt 2000), Balkan Slavic, because of its status in the middle of the prestige scale, had more L2 speakers than either Romani or Greek. In accordance with the hypothesis about the role of L2 speakers, both Romani and Greek have retained their case inflections, unlike Balkan Slavic. However, Albanian seems to present a problem for this hypothesis, because

it contains more case distinctions than Balkan Romance, Balkan Slavic, or Greek, yet is deemed to belong roughly to the same prestige category as Balkan Slavic and Aromanian, which display significantly impoverished case systems (see **3.2.2**). Although this topic is beyond the scope of the present study, further analysis of the historical sociolinguistic status of Albanian is clearly needed in order to shed more light on this hypothesis.

3.2 The diachrony of the Balkan case systems

The following three subsections introduce the basic characteristics of the individual case systems of Albanian, Balkan Romance, and Greek for purposes of comparison with Balkan Slavic. This is done, among other, to assess, whether a unified account of a Balkan case system is possible. While some synchronic aspects of the case systems are discussed in more detail, the main emphasis is, however, on their diachronic development.

3.2.1 Albanian

In this subsection, I introduce the basic characteristics of the Albanian case system and its development. I will address the question of the so-called locative case, attested dialectally, and discuss the historical sources of the modern cases. As a starting point, the modern standard Albanian case system of nouns is summarized in the following table:

		indefinite			definite		
		m.	m.	f.	m.	m.	f.
		'mountain'	**'friend'**	**'road'**	**'mountain'**	**'friend'**	**'road'**
singular	nominative	*mal*	*shok*	*rrugë*	*mali*	*shoku*	*rruga*
	accusative	*mal*	*shok*	*rrugë*	*malin*	*shokun*	*rrugën*
	dat.–gen./abl.	*mali*	*shoku*	*rruge*	*malit*	*shokut*	*rrugës*
plural	nom./acc.	*male*	*shokë*	*rrugë*	*malet*	*shokët*	*rrugët*
	dat.–gen.	*maleve*	*shokëve*	*rrugëve*	*maleve*	*shokëve*	*rrugëve*
	ablative	*malesh*	*shokësh*	*rrugësh*	*maleve*	*shokëve*	*rrugëve*

Table 5. Declension of the words 'mountain', 'male friend', and 'road' in standard Albanian.[73]

Table 5 shows the declension of three Albanian nouns, one feminine and two masculine, belonging to the total of three main declension paradigms in the singular. There is only one declension paradigm in the plural, but the formation of the plural stem varies. The historical Indo-European neuters

[73] The decision to present a single dative–genitive case instead of two separate cases, as in the Albanian descriptive tradition, is discussed in **3.3.1**.

have become increasingly rare in modern Albanian, yet new verbal and abstract nouns of neuter gender are derived from participles and adjectives respectively (Hetzer 1978, 146).

The Albanian case system and the case agreement in general within the Albanian NP is characterized by extensive syncretism and homonymy and further complicated by a linking article, which is used mainly with adjective and genitive modifiers (see, e.g. **3.3.1**). The typically postposed adjective modifiers agree only in gender and number, not in case, while the linking articles and the more rarely preposed adjectives also express the case. Personal pronouns more often distinguish the ablative, which, along with nouns, is distinct from the dative–genitive case only in the indefinite plural. Also, the accusative, which differs from the nominative only with definite singular nouns, has more forms distinct from the nominative within the pronominal system, with first- and second-person personal pronouns, however, displaying some homonymy with the dative–genitive.

The fused marker of case and definiteness becomes somewhat less opaque when certain changes that have taken place relatively recently, that is, since the 15th century, are taken into account. Still visible in the definite nominative and accusative plural, the definite article was historically attached to the case ending in the dative–genitive and ablative forms: *maleve* < *mal-e-ve-t* (Demiraj 1993, 145).[74] A distinct ablative of definite plurals was replaced by the dative–genitive only during the last few centuries (*bjeshkë-sh-it* 'the mountain pastures [abl.]') (ibid.).

In addition to the case forms presented in **Table 5,** there is a special definite form of noun used with a preposition in some spatial constructions in some varieties of Albanian. For example, in a linguistic variety found in the village of Leshnjë and belonging to the Tosk dialect, one of the two major dialects of Albanian along with Geg, there is a special desinence *-t* in variation with the accusative, which is used with some prepositions to express spatial relations.

(27) Tosk dialect of Leshnjë (Jully & Sobolev 2002, 46)

un	*shkoj*	*në*	*mal /*	*mal-t*
I	go	PREP	mountain.ACC	mountain-?

'I go to the mountain'

[74] The reverse has also been argued: Some authors claim that the definite article was originally attached to the nominal stem, but Shaban Demiraj (1993, 144) considers this view erroneous; he is seconded by Vladimir Orel (2000, 246–247), who derives the modern forms credibly from the Early Proto-Albanian noun-demonstrative sequences.

The use of this form, despite the name "locative" often given to it, demonstrates the Balkan tendency to merge the marking of goal and location, discussed in **3.3**. Likewise, in the Leshnjë dialect, in addition to goal meaning in example **(27)** the same form is used in locative expressions; for example, *ai jeton në malt* 'he lives on the mountain' (Jully & Sobolev 2002, 46). A noun in the accusative following the preposition *në* cannot appear with the definite accusative case suffix *-n*, although it is interpreted as definite. Thus, it seems likely that the birth of the structure is related to the need to mark overtly the definiteness of the NP. However, with a modifier, the definite accusative form can be used: *në malin e madh* 'on/to the big mountain'.

The status of the desinence *-t* as a distinct case is unclear. Shaban Demiraj (1993, 147–148) argues that it must be treated as a special and relatively novel use of the accusative, not, for instance, as a vestige of some lost historical case. Demiraj supports Holger Pedersen's (1894, 310) view that the form resulted historically from adding the definite element *-t* to the indefinite accusative form. Demiraj (1993, 148) asks whether the definite element *-t* could have been taken from the oblique cases by analogy, since the "locative," like the definite singular dative–genitive and ablative cases, receives the linking article *të* instead of the expected definite accusative *e*. The definite element resulting from the non-nominative cases by analogy is supported by the fact shown earlier that the *-t* element used to be an even more frequent marker of definiteness, attached to all plural forms. However, in my view the linking article *të* is likelier to have been inherited from the indefinite accusative, being indicative of the late birth of the structure. In other words, the linking article is preserved from the earlier accusative construction, but having become opaque in the sense of definiteness, the linking article did not change its form, although the NP became definite.

The Proto-Indo-European (PIE) case system is reflected only in the indefinite declension of Albanian and in an extremely reduced form.[75] The distinction between the nominative and the accusative was lost early, largely as early as the Proto-Albanian period (Orel 2000, 233–238). The dative–genitive singular of both masculine declension types (old PIE **o*-stem nouns) continues the PIE locative ending **-ei* (ibid., 238–239). In fact, the modern variation between *-i*/*-u* is only secondary and in complementary

[75] There are some attempts to demonstrate that the modern indefinite case inflection is a secondary development and that it does not reflect the PIE cases. These views, considered marginal today, are summarized in Demiraj 1993, 111, 113.

distribution: *-u* appears after vowels and velars. Vladimir Orel (2000, 238–239) holds that the feminine (PIE **ā*-stem) dative–genitive singular *-e* reflects the sequence of the PIE stem vowel and the ending of the dative and locative **-āi*. The plural dative–genitive *-(v)e*, on the other hand, seems to continue the Proto-Indo-European genitive plural **-ōm* (ibid., 237).

The ablative forms an interesting part of the Albanian case system. The -*sh* element of the plural originates from the PIE locative plural **-su* (it has the same source as the Slavic locative plural; e.g., OCS *vъ selě**xъ*** 'in the villages') (ibid., 237–238). Absent as a distinct case in other Balkan languages, the ablative has a combination of functions that is also somewhat peculiar. Although rarely used without a preposition, it has several adverbial uses indicating source, both literally and figuratively, and it can also mark the agent of a passive construction (Demiraj 1993, 107). Most of its forms are indistinguishable from the dative–genitive, and the prepositions that govern the ablative, in particular *prej* 'from' and *ndaj* 'toward', can also take the dative–genitive when the cases are distinct. Most intriguing is that the ablative is used in compounding as the case of the attributive element. When the forms are homophonous, this use is distinct from the possessive structures employing the dative–genitive, because the linking article is not used: *hapëse konserva**sh*** 'can opener'; *shpall lufte* 'declaration of war'.

The definite case forms, according to Orel (2000, 247), pre-date the linking article, because in Old Albanian (the earliest attested form of Albanian), the definite forms are fully developed, whereas the linking article is not yet obligatory. The definite declension originates from a demonstrative pronoun, like the linking article, and the similarities between the definite declension, linking articles, and demonstrative pronoun are still evident. Only *-n* of the definite singular accusative needs to be explained, since the nasal element, inherited from the PIE accusative, is not visible elsewhere. It appears that the accusative nasal element originates from the case ending of the noun, preserved by the demonstrative pronoun, which itself was elided (Orel 2000, 247; Demiraj 1993, 132).

The origin of the Albanian case system, with its four-part distinction of nominative, accusative, genitive–dative, and ablative, is presented above. However, the sources of the Albanian case forms do not reveal much about the exact developments that have led to the current system, and while the modern Albanian case system is highly reminiscent of other Balkan languages, the chronology of its development remains largely unknown. In the study of case syncretism briefly discussed in **1.3.2**, the homonymity

between two historically distinct case forms is addressed by assessing the potential semantic overlap between the cases in question. Ideally, such an analysis would be based on diachronic material whereby the changes in the use of the cases could be assessed in their contexts. Here, we have to settle for some general observations about the destiny of the Proto-Indo-European case inflection within individual languages.

Against this background, the sources of the Albanian cases are not atypical. There is much syncretism between the locative and dative in the Indo-European languages, but more important, according to Robert Beekes (2011, 187), the locative and dative might, in fact, continue one single case, separated by a mutually exclusive context, the dative being used with animates and the locative with inanimates. The PIE ablative had only one distinct form, the one with the **o*-stem singular nouns; otherwise, the genitive would be used (ibid.). The Albanian ablative does not continue that case form but the ablative originated from the locative instead, and it is unclear whether, similar to its form, its function, too, is a re-innovation. The key question from the Balkan point of view is the history of the genitive–dative case, this merger being a typical feature of the Balkan case systems. The history of the case is twofold, since its form originates from the genitive in the plural and from the locative or dative–locative in the singular. Without any traces of differentiation between the two distinct cases, genitive and dative, this merger must have come into existence long before the Old Albanian period.

Yet in this respect, the ablative, which is lacking from other Balkan languages and whose origins seem to be in the PIE locative, is especially interesting because it is used in compounding. The modern Albanian expression of adnominal possession proper must be viewed as a relatively recent strategy, because, in addition to the possessor marked with the genitive–dative, it involves the linking article, grammaticalized only during the era of written records. It is a well-known fact that compounding strategies can preserve traces of otherwise obsolete ways of marking possession, such as old genitive forms in Scandinavian (see e.g. Bauer 2009, 406). Therefore, it could be asked whether the ablative used in compounding in fact reflects the original strategy of marking possession. Tentatively, I propose that the distinction between the PIE genitive and dative was lost early, but the new ablative, originating from the PIE locative, took over the functions of marking possession, relying on its partitive and source-

indicating meaning, only later to be replaced by the genitive–dative, homonymous with the ablative of singular nouns.

3.2.2 Balkan Romance

This subsection discusses the Balkan Romance case system and its development. Unlike Albanian, there is written evidence to inform us about the history of the Balkan Romance case system, namely, sources in Classical Latin and non-standard Latin.[76] Nevertheless, a significant part of the development of Balkan Romance remains obscure, because the non-standard Latin texts do not amount to a "literary language" of Proto-Romance, not even to the limited extent that Old Church Slavonic does in relation to Late Proto-Slavic (for non-standard Latin as evidence of the syntax of the spoken language, see Halla-aho 2009, 26–42). Therefore, much of what is known or assumed about the development of Balkan Romance before the first Romanian texts in the 16th century is based on reconstruction and other indirect evidence.[77]

The standard Romanian case system is summarized in the following table:[78]

		indefinite			definite		
		f.	m.	m.	f.	m.	m.
		'girl'	'boy'	'brother'	'girl'	'boy'	'brother'
sg.	nom.–acc.	*fată*	*băiat*	*frate*	*fata*	*băiatul*	*fratele*
	dat.–gen.	*fete*	*băiat*	*frate*	*fetei*	*băiatului*	*fratelui*
pl.	nom.–acc.	*fete*	*băieți*	*frați*	*fetele*	*băieții*	*frații*
	dat.–gen.	*fete*	*băieți*	*frați*	*fetelor*	*băieților*	*fraților*

Table 6. Declension of the words 'girl', 'boy', and 'brother' in standard Romanian.

Balkan Romance nouns distinguish two inflected cases, the nominative–accusative and the dative–genitive (the vocative is not discussed here; see **1.3.2**). With personal pronouns, there is an additional distinction between the nominative and accusative (the number of Balkan Romance cases is discussed in **3.3.3**). The examples in **Table 6** display some

[76] The term Vulgar Latin is avoided here. On the problems of the term, see Väänänen 1981, 3–6 and Adams 2013, 3–27.
[77] For the first reliably dated Romanian text, namely, the letter of Neacșu of Câmpulung from 1521, see, e.g., Rosetti 1968, 468–469.
[78] This table does not attempt to represent all declension paradigms of Romanian. For a concise, but complete presentation of Romanian morphophonological variation, see Dumitrescu 2008.

morphophonological alternations, a change in stem vowel (*fată–fete*) and a palatalization of the stem-final stop, a feature of plural formation in Balkan Romance together with a change in the stem-final vowel (*băiat–băieți* [bəˈjat]–[bəˈjetsʲ]).

The individual Balkan Romance languages, namely, (Daco-)Romanian, Aromanian, Megleno-Romanian, and Istro-Romanian, differ somewhat in their case systems.[79,80] The case systems are similar to standard Romanian in their main principles, presented above, but some varieties may show fewer distinctions. Megleno-Romanian and some varieties of Aromanian and Istro-Romanian do not inflect nouns for case, although the indefinite article, for example, may distinguish between the nominative–accusative or genitive–dative cases (Markoviḱ 2007, 51; Mišeska Tomić 2006, 152–153; Kovačec 1962, 76). As in Albanian, in Balkan Romance too the typical NP word order is 1. noun 2. adjective modifier; likewise, the adjectives show more limited case agreement. Furthermore, when the adjective appears preposed, the definite article attaches to it instead of to the noun. Similar to Albanian, there is a linking article that agrees with the head of the NP. In Romanian, however, its use is more limited than in Albanian, because the linking article is employed almost exclusively with a possessor that does not immediately follow the head, or the head is indefinite or modified by preposed elements (the linking article is discussed in more detail in **3.3.3**).

Classical Latin had seven cases—nominative, vocative, accusative, genitive, dative, and ablative—together with the very marginal locative. Most nouns in Balkan Romance reflect either the Latin nominative or

---

[79] Often, when the Romanian language is discussed in the context of other Eastern Romance languages, the name Daco-Romanian is used to refer to Romanian. Ancient Dacia coincided geographically with the modern states of Romania and Moldova. Note that Romanian is spoken in Moldova, too, and is also the official language according to the constitution of the Republic of Moldova (DIRM 1991). The name "Moldovan" for the language is connected with political antagonisms of the Soviet era and is rarely used any longer.

[80] Istro-Romanian, a seriously endangered Eastern Romance language spoken in Istria, Croatia, must be considered as forming part of Balkan Romance, despite its current geographical distance from the other languages. In this sense, its status is similar to the varieties of Romani spoken outside the Balkans; see **3.1.2**. Its relation to the other Eastern Romance languages is contested. The main question involves its status either as an independent branch of Proto-Romanian (the proto-language of Daco-Romanian, Aromanian, Megleno-Romanian, and Istro-Romanian) or as a daughter language of the later "Proto-Daco-Romanian." The estimates of its time of secession from the other Eastern Romance languages range from the 11th to the 12th centuries. (Zegrean 2012, 175) Istro-Romanian is generally considered to be the least "balkanized" of the Balkan Romance languages (ibid.), i.e., displaying the smallest number of common Balkan features, no doubt because of its long separation from the heartlands of the linguistic area.

accusative: Rom. *şarpe* < Lat. *serpēns* [nom.] (not *serpentem* [acc.]) 'snake' Rom. *lege* < Lat. *lēgem* [acc.] (not *lēx* [nom.]) 'law' (Dimitrescu et al. 1978, 202–203). When no change in the stem occurred, these two forms were often rendered homonymous by sound changes (Väänänen 1981, 66–68). The only form inflected for case that directly continues a Latin nominal case other than the nominative or accusative is the Romanian indefinite singular genitive–dative of feminine nouns (in **Table 6**, *fete* < Lat. *fētae* [gen./dat.]). This particular case form already displayed genitive–dative syncretism in the Latin first declension (the same syncretism also occurred in the fifth declension).

The fact that Daco-Romanian in particular resorts predominantly to the bare genitive–dative case for nouns in expressing possession and for the recipient is interesting because of the remarkably uniform modern Western Romance use of the reflexes of the Latin prepositions *dē* (for possessors) and *ad* (for both recipients and possessors) in these functions (for the analytical marking of possessors and recipients in Balkan Romance, see **4.2.3**). The analytical expression of these functions with the prepositions *dē* and *ad* is already visible in Latin (Väänänen 1981, 113–114), but, perhaps surprisingly, older Romanian texts also seem to resort to the corresponding modern prepositions *de* and *a*, used with the accusative, in more contexts than in contemporary Romanian (Dimitrescu et al. 1978, 351–352; Rosetti 1968, 554–555). The historical grammars quoted here do not provide information about the frequency of these prepositions, but especially the use of *a* + accusative in expressing recipients is noteworthy, as this strategy is entirely absent in modern Romanian (in Aromanian, however, *a* is used together with the genitive–dative in this function; see **4.2.3**). This seems to point to a development, discussed further in **3.4**, that is contrary to the common Balkan tendency to increase analytism at the expense of inflection.

Nevertheless, the use of the genitive–dative instead of prepositions in these constructions is in line with the Balkan tendency to merge the inflected case marking for adnominal possessors and recipients, often called the genitive–dative merger, reflected in modern Greek and Albanian, and historically also in Balkan Slavic (for a more thorough analysis of the merger, see **4.2**). Yet a widely-held view of the development of Proto-Romance into the individual Romance languages supposes a transitory three-case system in which one of the cases is a genitive–dative along with the nominative and accusative (Zamboni 2000, 110; Adams 2013, 319). In addition to case inflection, this syncretism is observed in the use of the

preposition *ad* in non-standard Latin especially to mark human possessors together with recipients (Väänänen 1981, 114–115).[81] Balkan Romance could thus be seen as retaining only the syncretism of the genitive and dative of Proto-Romance.

Indeed, the Balkan Romance genitive–dative syncretism can be seen at most as an indirect continuation of the Proto-Romance case system, since most of the forms inflected for case fuse the case marking with the definite article, whose postposed placement is also unique to Balkan Romance, whereas other Romance languages employ a preposed article. One of the very few other attested Romance varieties that preserved some case inflection, namely, Old French, built its two-case system from the Latin nominative (*cas sujet*) and accusative (*cas régime*, replacing all non-nominative cases) (Brunot & Bruneau 1949, 184–185). In Balkan Romance, the definite forms of the nouns preserve most case distinctions for the same reason as does Albanian: The forms originate from a demonstrative pronoun. The definite singular genitive–dative desinences *-lui*[82] and *-ei* continue non-standard Latin dative forms *illūi* (also the source for Fr./It. *lui*) and *illēi* of *ille*, created by analogy with the dative interrogative or relative pronoun *cui* (Dardel 1964, 9). The definite plural genitive–dative desinence *-lor*, on the other hand, stems from the genitive plural *illṓrum* (gen. pl., source for Fr. *leur* and It. *loro*) (Klausenburger 2000, 111–112). The same case marking elements have spread to the numeral ‘one’, used as an indefinite article (Rom. *unui*, *unei*, *unor* < Lat. *ūnus*), and to most pronouns (e.g., Rom. *acestui*, *acestei*, *acestor* ‘this’ < Lat. *eccum istum*).

As we have seen, the Balkan Romance case system, in the varieties that still preserve it for the nouns, is unique among the modern Romance languages, differing also from the Old French division between the *cas sujet* and the *cas régime*. It preserves a two-part system that fuses the Latin distinction between nominative and accusative and between genitive and dative, respectively. The source of the case markers is the demonstrative pronoun *ille*, with the exception of the indefinite feminine singular genitive–dative, which seems to continue the Latin genitive or dative and which are syncretic in the first declension.

Notably, the sources for the genitive–dative forms are the same in Albanian as in Balkan Romance: dative in the singular and genitive in the

[81] Notice, however, that the evidence for this is very limited (Timo Korkiakangas, personal communication).
[82] With given names, *lui* is preposed.

plural. However, unlike Albanian, for Balkan Romance there are credible hypotheses about why the genitive was preferred as the source for the plural marker. Wilhelm Meyer-Lübke (1894, 90) suggested that *illōrum* was better suited to become a full pronoun in Proto-Romance instead of *illīs* (dat. pl.), because of its being "heftier" in phonetic structure. It is perhaps worth noting as well that by losing the word-final *s*, *illīs* would have eventually become homophonous with the nominative plural *illī*, the source for Romanian nominative–accusative plural *ei*.

The assumption that the source of Balkan Romance dative–genitive case lies in the dative–genitive syncretism of the Proto-Romance three-case system contains the risk of circular reasoning. There is evidence of some dative–genitive syncretism in both non-standard Latin (see e.g. Dardel 1964, 13–15) and in certain modern languages, for example, French and Italian, where the pronouns, derived from the dative in the singular and the genitive in the plural, are sometimes used to mark both possession and indirect objects. Yet the most important body of evidence for a stable Proto-Romance three-case system comes from Balkan Romance itself. The evidence provided by the modern destiny of personal pronouns derived from *ille* is also problematic and for two reasons: First, pronouns often express a different number of case distinctions than do nouns, and the distinctions may be organized differently (see, e.g., **3.3.2** and **3.3.3**). Second, and specifically with regard to the further development of *ille*, personal pronouns make reference to humans, who, among other animates, are often treated differently from non-animates in grammar (see the discussion in **3.3.4**). Cases like the genitive and the dative, which mark possessors and recipients, both being grammatical functions with typically human referents, are especially sensitive. In addition, Robert de Dardel (1964, 18), who is cited, for example, by James Adams (2013, 319) as confirming the existence of the three-case system, in fact questioned whether the genitive–dative really existed in Ibero-Romance or in Sardinia, Sicily, or southern Italy.

My intention is not to cast doubt on the concept of the Proto-Romance three-case system beyond the few points of criticism above. However, given that the main goal here is to examine the role of Balkan Romance as a potential source for the development of the Balkan convergence of case systems, these problems prove to be critical. It is clear that Proto-Romance possessed several attested phenomena, some of which seem to have been passed on to Balkan Romance. These phenomena include the genitive–dative desinence -*e*, probably inherited from the syncretic Latin case, the

forms of demonstrative pronouns that developed into postposed definite articles inflected for case, and the limited use of prepositions derived from *ad* and *dē* in replacing the Latin dative and genitive. The fact that many varieties of Balkan Romance, especially Daco-Romanian, preserve the two-case distinction does not, however, require that the genitive–dative must have been a stable feature of Proto-Romance. Indeed, despite Classical Latin already having many overlapping functions and forms of the genitive and dative, what is remarkable is the resilience of the genitive–dative of Balkan Romance and its resistance to replacing the case with analytical marking in the face of an entirely contrary development in other Romance languages. This, in my view, gives reason to suspect an important role for language contact in the selection of the available grammatical means to mark these functions and in their subsequent retention, although, admittedly, not necessarily in the birth of the phenomenon itself.

3.2.3 Greek

The description of the development of the Greek case system is, no doubt, by far the easiest among the Balkan languages. This is because of the virtually uninterrupted literary activities in the same geographical area for more than three thousand years. However, owing to the archaizing nature of the literary traditions, changes in the spoken varieties of Greek show up belatedly and indirectly in the texts. Several features of the case system of Biblical Greek have already been discussed in **Chapter 2**, because Greek served as the source language for the early Slavic literature. This subsection presents the main characteristics of the development of the Greek case system. The modern standard Greek case system is summarized in the following table:[83]

[83] Again, this table summarizes only the basic characteristics of the case morphology, not all inflectional paradigms. For a morphological analysis of the development of Modern Greek noun declensions, see Ruge 1969. Note, too, that the transliteration of Modern Greek differs here from that used for Biblical Greek or Classical Greek: Although the stress is marked, those graphemes that are homophonic in the modern language are not distinguished. It must be said that only a few of the commonly used Romanization systems for English are satisfactory. The system used in this and the following chapter is mainly phonetic, rendering the graphemes <γ,δ,θ,χ> as *gh*, *dh*, *th*, and *kh*, when they are realized as fricatives, but, for example, when <γ> is pronounced as the approximate [j], it is written as *y*. Also, the ancient diphthong <ου> is written as *u*, like other monophthongized vowels written in Greek with a digraph. The names of the Greek authors cited in this volume are written as they themselves have chosen to latinize them.

		masculine	feminine	neuter
		'the teacher'	'the woman'	'the book'
sg.	nominative	*o dháskalos*	*i yinéka*	*to vivlío*
	accusative	*to*(*n*) *dháskalo*	*ti*(*n*) *yinéka*	*to vivlío*
	genitive	*tu dhaskálu*	*tis yinékas*	*tu vivlíu*
pl.	nominative	*i dháskali*	*i yinékes*	*ta vivlía*
	accusative	*tus dhaskálus*	*tis yinékes*	*ta vivlía*
	genitive	*ton dhaskálon*	*ton yinekón*	*ton vivlíon*

Table 7. Declension of the words 'teacher', 'woman', and 'book', and their corresponding definite articles in Standard Modern Greek.

Standard Modern Greek nouns distinguish three inflected cases, nominative, accusative, and genitive. The vocative, which is limited to masculines as a separate form (*dháskale*[voc] 'teacher'), is not discussed here (see **1.3.2**). Only masculine nouns belonging to the declensional type corresponding to the Ancient Greek second declension (PIE **o*-stem nouns) have a differential accusative form. The rest of the masculines combine the accusative with the genitive, whereas with the feminines and neuters, the respective accusative forms are homonymous with the nominatives. As modifiers, Greek adjectives appear predominantly, but not exclusively, preposed and are inflected in a manner similar to the nouns. The pronouns too distinguish the nominative, accusative, and genitive, but display less syncretism among the cases than the nouns. The definite article (shown in **Table 7**) corresponds to the clitic form of the third-person personal pronoun and, together with the indefinite article, often has a separate accusative form along with masculines and feminines, even when the noun does not. However, personal pronouns, especially in the plural, sometimes display syncretism between the accusative and the genitive. This syncretism is different from that of the nouns in the standard language because it results from the accusative forms of the pronouns overtaking the functions of the genitive, whereas that of the nouns is due to the loss of the word-final *n* (Mertyris 2009, 480).

Syncretism resulting from the accusative having assumed new roles is also characteristic of a major Modern Greek dialectal division into southern and northern dialects; this is in addition to the treatment of high and mid-vowels (in the northern dialects, when not stressed, the standard language /i/ and /u/ are often deleted and /e/ and /o/ reduced to /i/ and /u/; Horrocks 2010, 404). In the southern dialects, including Standard Modern Greek, the expression of the recipient was taken over from the dative by the genitive. In contrast, northern dialects resort to the accusative when expressing the

recipient (Horrocks 2010, 284). The northern dialects employ the genitive plural only in a limited fashion, and the expression of adnominal possession is usually overtaken by a prepositional construction with *apó* + accusative.[84]

Historically, the loss of case distinctions in Greek is relatively well attested. In addition to the remaining cases, the oldest texts from perhaps as early as the 15th century BCE, the Mycenaean tablets written in Linear B script, reveal some use of a separate instrumental case. The genitive continues both the Proto-Indo-European genitive and ablative, and the dative of Classical Greek continues both the locative and the dative. The remaining use of the instrumental case was later taken over by the dative. (Horrocks 2010, 10.) The subsequent loss of the dative is one of the key questions surrounding the development of the Greek language after the Classical period. As with Albanian and Balkan Romance, a single case emerged in Greek, namely, the old genitive, combining expression of both the adnominal possessor and the recipient. Many of the functions of the dative were also passed on to prepositional constructions. The recipient, for example, can be marked, in addition to the bare genitive, with *se* + accusative (ibid., 284–285; for the use of prepositions in Biblical Greek, see Thomason 2006, 12–54; see also **4.2.3**).[85]

The use of the dative is thought to have been in retreat as early as the beginning of the first millennium. It is believed, however, that its use in some varieties might have continued until as late as the 10th century. (Horrocks 2010, 185.) The interchangeability of the dative and the genitive is observed first with personal pronouns. On the basis of the genitive personal pronouns used in papyrological Greek, Joanne Stolk (2013) finds support for a largely accepted pathway that might explain the semantic extension of the genitive's functions. In the prenominal position, the genitive personal pronoun may appear in the same position as a pronoun that acts as a benefactive or malefactive:

(28) Papyrological Greek, 2nd century CE (Humbert 1930, 171)

agórasón	*mou*	*tò*	*méros*	*toû*	*eleô̂nos*
buy.IMP.3P	I.GEN	the	part.ACC.SG	the	olive_grove.GEN.SG

'buy my part / me the part of the olive grove'

Finally, Stolk observes a gradual change in the use of the genitive personal pronoun in an unambiguously dative-like role: Almost non-existent

[84] Dionysios Mertyris, personal communication. For the loss of the genitive plural as a reason for the increase in the use of the analytical structure, see Humbert 1930, 39–41.
[85] This strategy is discussed in more detail in **4.2.5**.

in the Ptolemaic period (3rd–1st centuries BCE), the genitive personal pronoun is used to some extent in the Roman period (1st–4th centuries CE), and by the Byzantine period (5th–9th centuries CE), it forms a major part of the use of the pronoun.

According to Jean Humbert (1930, 39–46), a potential reason for the weakening and eventual replacement of the dative is the result of phonological changes that rendered the dative potentially homophonous with the accusative, affecting the Ancient Greek second declension, which acquired increasing numbers of words after the Classical period. This claim does not need to be seen as contrary to any explanation involving syntactic or semantic arguments, as it does not directly account for the expansion of the semantic roles of the genitive. In fact, like Stolk (2013), Humbert (1930, 171) too acknowledged the similar prenominal use of the personal pronouns as a potential context in which the interpretation of the pronoun is uncertain. In a similar vein, Ekkehart König and Martin Haspelmath (1997, 584–586) note the ambiguity of this construction, but argue that when the genitive was first re-interpreted as a verbal argument instead of an adnominal possessor, it acted as a dative external possessor and thereafter became grammaticalized to mark indirect objects as well (external possessor structures are further discussed in **4.2.5**).

It is generally argued that before the clear preference along the north-south dialectal division for either the accusative or the genitive as the replacement for the dative, full NPs with either the bare accusative or genitive would have been used to express indirect objects (Horrocks 2010, 284). Humbert (1930, 200) considers the adoption of the genitive to replace the dative a genuine innovation, whereas the use of the accusative resulted by analogy with those dative expressions that had become homophonous with the accusative. Dionysios Mertyris's (2009) attempt to explain the accusative source of the plural personal pronouns is based on the same hypothesis regarding the dissolution of the dative. He traces the expansion of the accusative forms to the free variation in assumed Early Medieval Greek (from the 4th century on) between the accusative and genitive to mark the indirect object and the subsequent failure to compose clitic genitive forms in the plural. Accepted as a marker of indirect objects, the accusative forms had access to the marking of possessors as well. Mertyris (2009, 488) reminds us that the latter development also took place in other Balkan languages regarding the emergence of the genitive–dative case.

The grammaticalization of the merged marking of the recipients and the adnominal possessors in Greek is interesting from a Balkan point of view, since with the nouns the source of the new case is exclusively the earlier genitive in Greek. In other Balkan languages, where at least some forms of the merged case originate from the dative, a more typical grammaticalization pathway from the dative to the adnominal possessor may be assumed to have contributed to the ambiguity between the two earlier cases (on this grammaticalization pathway, see Heine & Kuteva 2002, 103–104). However, a word of warning regarding the proposed paths of grammaticalization is necessary here as in **3.2.2**: Pronouns are not necessarily representative of the development of the entire case system, since they may display case systems different from those of the nouns and retain more ancient distinctions.

3.3 Balkan argument marking beyond case inflection

This section discusses some key topics involving case marking, but that are not necessarily limited to inflectional morphology. This is crucial for the whole chapter, since, as the previous section **3.2** demonstrated, the pronouns and the marking of definiteness are closely tied to the expression of grammatical case, and therefore a comparison between case systems cannot be limited to case inflection alone. Another reason why such broadening of the themes is needed stems from a basic observation regarding the Balkan linguistic area languages: There has been a gradual increase in analytical marking to replace earlier constructions with NPs inflected for case (for example in Asenova 2002, 76, "tendency toward analytism"). In addition to the grammatical relationships previously expressed with bare case forms, the increase in analytism involves the comparison of adjectives. Besides the themes in this section, an important phenomenon related to the Balkan expression of grammatical relations beyond case inflection, namely, the genitive–dative merger or, as preferred here, the recipient–possessor merger, is discussed from a wider typological point of view in **Section 4.2.**

Although the tendency toward analytism is perhaps best characterized as an increase in the use of prepositional constructions at the expense of case inflection, this does not mean that the preposition inventories somehow became enriched. In fact, there seems to have been a tendency to reduce the number of distinctions made with prepositions. For example, Old Church Slavonic made a distinction between three source-indicating prepositions *otъ*, *iz*, and *sъ*, the last two parallel to the interior–exterior difference made

with the goal-location prepositions *vъ* 'in' and *na* 'on'. Unlike all other Slavic languages, Bulgarian and Macedonian lost both source prepositions marking the interior–exterior distinction, resorting to only one preposition, which continues *otъ*. Modern Greek too drastically reduced the number of prepositions. The interior–exterior difference was lost entirely, when the preposition *se* (< *eis* 'into') was generalized as the only goal-location preposition and *apó* as the only source-indicating preposition (Horrocks 2010, 285).

Another feature of the Balkan nominal syntax, one that is also connected with the expression of locative relations, is the lost distinction between goal and location. This feature does not necessarily involve an increase in the use of prepositions, but rather the loss of a distinction usually made with the differential use of cases. In all Balkan languages, the unmarked expression of goal (*I'm going **to Skopje***) and location (*I am **in Skopje***) is identical and realized with the same preposition and a nominative–accusative (see also examples **(26)** and **(27)** and their discussions). Of course, the same development has taken place in many other European languages, and therefore, an urgent need for a contact explanation is not called for, as Kenneth Naylor (1981, 343), for example, claims. Yet languages that retain some case distinctions seem generally to maintain the distinction, as do German, Icelandic, Basque, and the Fenno-Ugric languages of Europe, for instance. There is further evidence that the loss of the goal-location distinction might be prone to spread through language contact, since some Balkan Turkish varieties have merged their marking, generalizing the locative to both meanings (Friedman 2006, 35–38).[86]

This section continues the discussion begun in **3.2** through four subsections dealing with the category of case in definite and indefinite articles and pronouns, together with the argument marking beyond case inflection.

3.3.1 The article as a carrier of case marking

With the exception of Modern Greek masculine singular and plural and, very marginally, some Balkan Slavic dialects, if there is an accusative case marking in the Balkan languages, it is a post-PIE innovation, most often

[86] Jørn Ivar Qvonje (1979, 135) reminds us that the development of marking goal and location in fact involves several steps: For example, in earliest Latin, the goal and location could sometimes be expressed with the bare accusative and the ablative respectively. In the second stage, dedicated prepositions emerged, and the subsequent developments involved both the use of a single preposition as well as the same case.

acquired through the grammaticalization of a demonstrative pronoun as a definite article (the indefinite article, i.e., the numeral "one," given in parentheses):[87]

	nom. indef.	**acc. indef.**	**nom. def.**	**acc. def.**
Modern Greek	(*énas*) *líkos*	(*énan*) *líko*	*o líkos*	*ton líko*
Albanian	(*një*) *ujk*	(*një*) *ujk*	*ujk-u*	*ujk-un*
Romanian	(*un*) *lup*	(*un*) *lup*	*lup-ul*	*lup-ul*
Macedonian	(*eden*) *volk*	(*eden*) *volk*	*volk-ot*	*volk-ot*

Table 8. The articles as carriers of case markings with the word 'wolf'.

As can be seen in **Table 8**, Romanian, as well as other Balkan Romance languages and Balkan Slavic (excluding standard Bulgarian and some dialects), does not differentiate between the accusative and nominative through case marking. However, the definite article in genitive–accusative cases with animate masculines in a non-subject position is still attested in the Balkan Slavic 17th-century Damascene literature, which makes use of a close equivalent to the spoken language of the time (see **Chapter 2**):

(29) 17th century Bulgarian, the *Tihonravov Damascene* (Dëmina 1985a, 315).

tămničar-a-togo
jailer-M.GEN/ACC.SG-DEF.M.GEN/ACC.SG.
'[him,] the jailer'

In Albanian, in addition to nominal case marking and articles, both definite and indefinite NPs with a modifier, whether adjectival or genitival, receive a linking article that agrees with the head in number, gender, and case. The linking articles *i*, *e*, *të*, and *së* display a great deal of syncretism. Nevertheless, a rough generalization is that with indefinite singulars, the nominative is different from other cases, and, in the definite paradigm, only the masculine singular differentiates among all cases, while others combine the nominative and the accusative. Here, again, only with definite masculines is there a separate accusative form different from both the nominative and the genitive–dative:

[87] In fact, the feminine -*os* declension also has a separate accusative plural, similar to the masculine -*os* declension.

(30) Albanian (Zymberi 1991, 51).

mësuesi	*i*	*Zanës*
teacher.DEF.M.NOM.SG	LINK.DEF.M.NOM.SG	Z.GEN/DAT.SG

i-a	*jep*	*librin*	*e*
to.him.M.DAT.SG-it.M.ACC.SG	gives	book.DEF.M.ACC.SG	LINK.DEF.M.ACC.SG

Teutës	*vëllait*	*të*	*Dritës*
T.GEN/DAT.SG	brother.M.GEN/DAT.SG	LINK.DEF.M.GEN/DAT.	D.GEN/DAT.SG

'The teacher of Zana gives the book of Teuta to the brother of Drita'

Similar to the expression of accusatives, the definite and indefinite articles also contribute significantly to the differential marking of the cases with the genitive–dative. The Greek definite and indefinite genitive singular articles help to distinguish between the accusative and the genitive, otherwise homonymous in several inflectional paradigms. With the exception of indefinite feminine singular nouns, the Romanian definite article that fuses the marking of the genitive–dative is the only element that displays any case distinctions.

3.3.2 Pronouns and the clitic doubling of objects

It is a well-known fact that pronouns generally preserve more case distinctions than nouns (see e.g. Iggesen 2005, 613). In Balkan languages too, personal pronouns have fuller inflectional paradigms than nouns. In both standard Bulgarian and Macedonian, pronouns form a three-case system: nominative, accusative, and dative. Greek and Balkan Romance make the same number of distinctions. Although similar in other respects, Albanian has specific ablative forms for first- and second-person pronouns.[88] Because the demonstrative pronouns acted as the source for the grammaticalization of the definite articles, the definite NPs are also more likely to be marked by an accusative different from the nominative in the Balkan languages with the exception of modern Balkan Slavic.

The Balkan pronouns also contribute to argument marking through clitic doubling, whereby clitic forms of personal pronouns are used to double objects. The doubled constructions in the Balkan languages involve both direct and indirect objects. The initial observations about clitic doubling in the Balkan languages merely involved a structural notion: All Balkan languages resort, at least occasionally, to "pleonasm," involving short, clitic pronouns. Kristian Sandfeld (1926, 110) described the doubling of objects in

[88] Whether the "third" case is called a genitive or a dative in the descriptive traditions of the individual languages is of minor importance, since their functions are nearly identical across the languages; see the discussion in **3.3.3**.

Romanian "nowhere near being a rule," and "far from regular" in Greek. According to Sandfeld, the doubling is "fairly regular" in Albanian and Megleno-Romanian and realized most consequently in Macedonian, while being completely unknown in eastern Bulgarian.

Since the early observations, clitic doubling of direct and indirect objects has received more attention (see e.g. Prendergast 2012; Kallulli & Tasmowski [eds.] 2008; Lopašov 1978). For example, Asenova (2002, 76) lists object doubling as one of the features of the "analytism of the Balkan nominal system." Nevertheless, its inclusion in the list of potential Balkanisms is not unanimously accepted. In Gunnar de Boel's (2008, 102) account of the development of the clitic doubling until the time of Medieval Greek, the author concludes: "…I don't think, moreover, that this is a typical Balkan phenomenon, as we meet it also in languages such as spoken French, English, or Swahili."

As de Boel (2008, 102) observes, clitic doubling in the Balkan languages is not typologically unusual: many European languages, most notably Romance languages such as French, Spanish, and Venetian, resort to it. Zlatka Guenchéva (1994, 25), on the other hand, reminds us that clitic doubling is possible only if there are two sets of personal pronouns, the accented and the non-accented. Therefore, the further grammaticalization of clitic doubling must first be preceded by the structural innovation of the clitic pronouns. In both cases, Slavic serves as the litmus test for the Balkan linguistic area: While clitic doubling could be typologically common, it is significantly limited or non-existent in Slavic languages outside the Balkans (see e.g. Franks & King 2000, 250), although all South and West Slavic languages possess the structural means for its production, namely, the clitic set of personal pronouns. Clitic doubling may, of course, be an inherent development in Albanian, Greek, and especially Balkan Romance, because of its frequency in other Romance languages. Nevertheless, in Balkan Slavic, its level of grammaticalization correlates with the level of grammaticalization of clitic doubling in the adjacent varieties of other Balkan languages, as will be shown further on.

In Albanian and the Macedonian standard language, the doubling of indirect objects is obligatory. Yet in colloquial Macedonian, the indirect object is sometimes left undoubled if it is the focus of the sentence (for more specific criteria, see Prendergast 2012, 154–155; Petroska 2008, 130). The degree of grammaticalization of direct object doubling in the Balkan languages ranges from the optional marking of pragmatic topics to the

obligatory marking of all definite direct objects. The doubling of all definite NPs takes place in western Macedonian dialects, including the Macedonian standard language, yet also indefinite direct objects too, especially when specific, are sometimes doubled (for the conditioning criteria, see Friedman 2008, 41–43; for the status of the Macedonian clitics, see Franks 2009, 218):

(31) Macedonian

(*ja*)	*vidov*	*edna*	*žena*
it.F.ACC.SG.CL	I.saw	one	woman

'I saw a (certain) woman'

Victor Friedman (2008, 58–60) concludes that the center of innovation of the Balkan clitic doubling, the one showing the most grammaticalized use, is in Western Macedonia, where Central Geg Albanian, western Macedonian, and Northern Aromanian have been in intense contact for centuries. Developments in the dialects south of this area have been inhibited by the influence of Greek, which is more conservative in this respect. In the Eastern Bulgarian dialects, object doubling is used only to mark topicalization, while further north in Romanian, its use becomes more restricted, involving conditions like humanness and partitivity.

From the point of view of case marking, the clitic doubling of objects pertains to the expression of case relation. Thus, as with the definite NPs, it could be asked whether the case disambiguating function is a driving force behind the grammaticalization of clitic doubling. Although marking the case may have been a factor in the grammaticalization of the object doubling, in the contemporary Balkan languages case disambiguation is only one of the functions of the definite articles and the doubled clitic pronouns. Yet there is the possibility of a link between the loss of case inflection and the emergence of the grammaticalized expression of definiteness, as will be argued in **4.3**.

3.3.3 The Albanian and Balkan Romance "genitive"

The use of linking articles in possessive structures, mentioned briefly in **3.2.1**, **3.2.2**, and **3.3.1**, is considered in both the Albanian and Romanian grammar traditions to represent a formal difference between the dative and genitive cases. This question is dealt in this section, because its deeper analysis cannot be carried out based only on inflectional morphology. In fact, the kinds of analyses in the prescriptive traditions presuppose the expansion of the definition of grammatical case beyond inflectional

morphology or word-level phenomena, a technical, theoretical, and terminological choice that I avoid in the present study, as mentioned in **1.3**. I do not, however, object to the separate cases on these grounds. Here, I attempt to demonstrate why it is problematic to view these two structures as representative of something like two separate grammatical cases, even in an expanded sense of the concept.

In Albanian, in addition to the adnominal possessive structures, the linking article appears before adjective modifiers:

(32) Albanian

prindërit	*e*	*dashur*	*e*	*Ilir-it*
parents.DEF.NOM.M.PL	LINK.NOM.M.PL	dear	LINK.NOM.M.PL	Ilir-GEN/DAT.M.SG.DEF

'Ilir's dear parents'

In the descriptive tradition of Romanian, the Romanian linking article is called a genitival article. If the possessor does not immediately follow the head or the head is indefinite or modified by preposed elements, then the genitival adjective is added:

(33) Romanian

părinții	*dragi*	*ai*	*Dacian-ei*
parents.NOM/ACC.M.PL.DEF	dear.M.PL	LINK.NOM/ACC.M.PL	Daciana-DAT/GEN.F.SG

'Daciana's dear parents'

However, the linking article is not limited to the structures of adnominal possession. Also structures involving ordinal numbers are affected:

(34) Romanian

părintele	*al*	*unsprezecelea*
parent.NOM/ACC.M.DEF	LINK.NOM/ACC.M.SG	eleventh

'the eleventh parent'

Andrew Spencer (2007, 247–248) has concluded that Albanian does not have a separate genitive case, but that the linking article contributes to a "possessum-agreement" construction: "In the possessum-agreement construction what is actually happening is that a morphosyntactic construction that canonically is used for attributive modification has been seconded to express possession […]." Indeed, the linking article does not contribute to the case marking of the possessor: it merely links the modifier to its head, with which it also agrees in case. This is evident, since the linking article is used in both Albanian and Romanian along with other modifiers besides nouns. In addition, there are several situations in which

the linking article is omitted in adnominal contexts, whereas the morphological case form remains unaffected. An explanation for the grammaticalization of an element linking the head to its modifier is most likely found in the word order of the NP. Unlike Balkan Slavic and to a greater extent than in Modern Greek, both Albanian and Romanian display a strong right-branching preference with the adjective modifiers. The linking article maintains the connection between the right-branching modifiers and their head.[89] Further evidence of the effect of word order is that in Greek, if the adjective modifier follows the head, then the definite article must be doubled before the adjective. The need for this additional coordination with the adjective modifiers is increased by the fact, mentioned in **3.2.1** and **3.2.2**, that both Romanian and Albanian adjectives show significantly less agreement than Greek and Slavic, especially when postposed.

The genitive–dative personal pronouns also function adnominally (see **4.2.4** for examples). The Romanian use of the first- and second-person pronouns is the best argument for a separate case form: The pronouns are often said to differentiate between a "dative" and a "genitive." Nevertheless, this so-called genitive is, in fact, the possessive pronoun, which agrees with its head. However, these possessive pronouns display case-like properties in prepositional constructions: Certain prepositions require either the "dative" pronoun or the "genitive," that is, the possessive pronoun. Also, the "genitive" prepositions require the linking article under the same conditions as in adnominal possessive constructions.

The historical sources of the Romanian prepositions requiring either the "genitive" or the "dative" are quite transparent: The prepositions in constructions employing the "genitive," that is, the possessive pronouns and the linking article, are grammaticalized from adnominally-used nouns. On the other hand, the prepositions taking the "dative" are mostly grammaticalized from verbal adjectives. In addition, many of the "genitive" prepositions still display some nominal qualities, such as grammatical gender: The linking article, when needed in these constructions, agrees with the gender of the preposition. It could thus be argued that this is the first step toward the grammaticalization of a new genitive case and that the prepositional constructions represent the first domain where such a distinction becomes possible. Nevertheless, not all first- and second-person pronouns display the "dative," because the clitic pronouns can be used with

[89] For a discussion on the double marking of definiteness in Greek and Aromanian, see Campos & Stavrou 2004.

both types of prepositions.[90] However, even if the limited set of pronouns shows emerging case differentiation, it is not unusual for the pronouns to distinguish a different number of cases than the nouns, as also occurs with the accusative and nominative in Balkan Romance and Balkan Slavic.

It remains an open question whether the obvious similarities between the Albanian and Romanian linking articles are the result of contacts between these languages. Shaban Demiraj (1993, 130) reminds us that the linking article in Albanian is likely to have emerged first, but he does not see a direct link with the Romanian counterpart. Demiraj (1993, 130) argues further that the linking article must have emerged from the need to differentiate between the "dative" and the "genitive." However, Greek is the first language in which the linking use of the definite article has been attested, and, consequently, the role of Greek cannot be excluded (Lindstedt 2000, 237). As argued here, there are no grounds for the kind of determinism, expressed by Demiraj, regarding the rise of the linking article. The two "cases" are always in a complementary distribution: There is no verb that would take as its complement either the "genitive" or the "dative," because all verbal complements always appear without the linking article.[91] Conversely, there is no adnominal use of the "dative" without the linking article unless the article were to be omitted anyway. A more likely reason for the emergence of the linking article is the need to coordinate with the right-branching modifiers within an NP. The discussion regarding the differences between nominal and verbal syntax is continued in **Subsection 4.2.2**.

3.3.4 Differential object marking in the Balkan languages

Most general accounts of the Balkan linguistic area do not mention differential object marking (DOM) as part of this area's shared features. It is likely that object marking—that beyond clitic doubling—has not been included in the analysis of the Balkan linguistic area for reasons similar to those given by de Boel (2008, cited in **3.3.2**): Splits in object marking based on animacy, human reference, definiteness, and the like are by no means typologically rare and therefore do not require a contact explanation. In other words, these similarities among the individual Balkan linguistic varieties can be explained by the universal effect of the so-called referential

[90] I would like to thank Andrei Dumitrescu for pointing this out to me.

[91] In Romanian, however, there is an existential construction with the verb 'to be' which involves the linking article and the genitive–dative as a predicative: *casa este a vecinului* 'the house is that of the neighbor' (Daniliuc & Daniliuc 2000, 51).

scales, which include hierarchies, such as levels of definiteness or animacy, believed to stem from the invariable qualities of human cognition.

There are three different subphenomena that could be evaluated from this point of view: 1) the inflectional case systems, 2) clitic doubling of direct objects, discussed above in **3.3.2**, and 3) prepositional marking of direct objects. However, some phenomena involving only pronouns are left out of the analysis. This subsection endeavors to determine to what extent similarities in DOM in the Balkan languages can be attributed to language contact and whether the concept of DOM has any advantages for the study of morphosyntactic Balkanisms of the nominal system in addition to, for example, the genitive–dative and goal-location mergers. First, the theoretical background of referentiality as a criterion for DOM will be presented.

The referentiality of the NP can mean its inherent reference, such as animacy or humanness, its relatedness to discourse, including definiteness and topicality, or a word-class phenomenon such as pronouns. In the study of Slavic languages, perhaps the earliest analyses of the role of referentiality in case marking can be traced to Roman Jakobson's *Beitrag zur allgemeinen Kasuslehre: Gesamtbedeutung der russischen Kasus* (1936). More recent studies which have addressed the development of the Slavic case systems from the point of view of referentiality, with a concentration on the notion of animacy, include Emily Klenin's (1987) work on the development of the genitive–accusative in OCS and Old East Slavonic and Laura A. Janda's monograph *Back from the Brink* (1996), which discusses, within the framework of cognitive linguistics, notions such as figure and ground and their role in the development of the category of animacy.

Observations about individual languages, such as the role of animacy in Slavic languages, have led to more general postulates on the significance of referentiality in argument marking. For example, Bernard Comrie (1981, 123) gives this formulation: "The following patterns in particular are found: (a) mark a P(atient) high in animacy, i.e. the accusative case is restricted to Ps that are high in animacy; (b) mark P high in definiteness, i.e. the accusative case is restricted to definite Ps […]." Put differently, this claim means that a differential marking is likely to be assigned for atypical arguments, that is, definite and animate or human P arguments.

However, Balthasar Bickel, Alena Witzlack-Makarevich, and Taras Zakharko ([Submitted draft]) claim that, in fact, the effect of referential scales on case alignment has never been subject to systematic and large-scale quantitative analysis. In their study, by increasing the size of the

sample used in earlier research (Bickel & Witzlack-Makarevich 2008), they set out to test the claim that there is a universal correlation between the odds of overt case marking and referential scale ranks—a negative correlation for subjects, a positive correlation for objects. Their proposal, formulated as an implicational universal, is that, if a language has a split in case marking, then this split will fit one of the universal referential scales. The researchers collected a database of 435 languages, which were then analyzed for case splits, with the notion extended to analytical case marking as well. Next, the splits were fitted to scales, including first, second, and third persons; singular vs. dual vs. plural; pronoun vs. lexical noun; definite/topical vs. indefinite/non-topical; and human vs. (non-human) animate vs. inanimate. Features such as specific vs. non-specific were likewise included. (Ibid., 3–9.)

The researchers' first observation (ibid. 17–18) was that the occurrence of the splits is heavily skewed: Of the 149 P splits observed, 63 percent are concentrated in the five top language families of the 39 displaying the splits, namely, Indo-European, Pama-Nyungan, Sino-Tibetan, Dravidian, and Turkic, which fit two proposed linguistic macro-areas, Eurasia and Sahul. In their conclusions, the authors claim that the impression of a universal referential scale effect in literature is indeed based on the ubiquity of such effects in Eurasia and Sahul: When the possibilities of areal diffusion and diachronic bias of languages belonging to the same language family are excluded, there is no evidence for the universality of the phenomenon. Finally, the authors (ibid., 27) state: "Given these findings, what becomes an urgent task now is research into the ways in which splits spread in language contact," and furthermore, "We submit that any deeper understanding of referential scale effects in individual languages needs to explore how it arose diachronically and what role was played in this by area diffusion." This formulation can also be seen as an encouragement to search for contact-induced explanations regarding the phenomenon in the Balkan linguistic area.

The accusative in the Balkan languages

Based on what was shown in Chapter 2 and Sections **3.2.1**, **3.2.2**, and **3.2.3**, Albanian, Greek, Balkan Romance, and historically Balkan Slavic as well can be said to display some level of similarity in the marking of direct objects with a distinct accusative case. The Balkan tendencies can be generalized relatively well in the following way: These languages

distinguish between two cases, nominative–accusative and dative–genitive, but the likelihood of a noun having a separate accusative case different from the nominative increases when the noun is singular, animate, and/or definite and/or displays the masculine grammatical gender. With definite nouns this is mainly because of the definite articles (see 3.3.1), since bare nouns only display the accusative rarely, as the Proto-Indo-European accusative largely merged with the nominative already at an early stage.[92] The main exception is Greek, in which the distinction is preserved, because the nominative element in masculine singular declension paradigms has been maintained:

PIE	**wl̥kʷ-os* 'wolf-NOM'	**wl̥kʷ-om* 'wolf-ACC'
Proto-Albanian†	*(*w*)*ulk-a*	*(*w*)*ulk-a*
Old Church Slavonic	*vlъk-ъ*	*vlъk-ъ*
Proto-Romance††	**lup-u*	**lup-u*
Modern Greek	*lík-os*	*lík-o*

Table 9. Cognates of an *o-stem word 'wolf' displaying nom.–acc. syncretism already at an early stage, with the exception of Greek, which still preserves the distinction. †Following Orel's (2000, 233–235) reconstruction of Proto-Albanian. ††Dimitrescu et al. 1978, 202.

Sound changes also eliminated Balkan Slavic accusatives in most environments, although feminine *ā*-stem accusative singular forms persisted somewhat longer (see **2.2.2**).[93] Some of the Bulgarian Rhodopian dialects, unlike standard Bulgarian or any other Balkan language, still marginally preserve the accusative singular, reflecting the PIE **ā*-stem case ending *-*m* (PIE *-*ām* > OCS -*ǫ* > -*ɔ*) when the ending falls under stress:

(35) Bulgarian: Rhodopian dialect of Ahăr čelebi (Miletič 1890, 238).

da	*zagúbiš*	*snág-a-ta*	*ti*	*i*	*duš-ɔ́-ta*	*ti*
that	you.lose	force-nom/acc.f.sg-def	your	and	soul-acc.f.sg-def	your

'that you lose your stamina and your soul'[94]

When unstressed, both the nominative and the accusative endings are realized as [ə], as in *snagata* ['snagətə], but for *dušɔta* [du'ʃɔtə], the standard Bulgarian would be *dušata* [du'ʃatə], showing the merger of the accusative with the nominative, including when stressed. In all Slavic

[92] The resulting case can be based on either of the two cases, including the accusative, as in Rom. *carne* 'meat' < Lat. *carnis* (acc.), not *carō* (nom.).

[93] Exceptions are the Northern Macedonian dialects, transitional Bulgarian border dialects, and the Serbian Torlak dialects where **ǫ* renders *u*. These dialects preserve the feminine accusative singular, which is different from the nominative.

[94] Glosses and underlining for stress are by the author. The phonetic reduction of an unstressed /a/ may not always coincide exactly with [ə]; nevertheless, unstressed reflexes of Proto-Slavic **a* and **ǫ* have merged in these dialects.

languages, most masculine singular accusatives, homonymic with the nominative already in Late Proto-Slavic, take the genitive singular ending -*a* of **o*-stem nouns when denoting an animate being. This development started first with the direct objects and later replaced all use of accusative for animate masculine singular NPs. A characteristic of all Slavic languages, this marking persisted until late in Balkan Slavic, although today it is used only dialectally (see **2.2.2**).

It is worth pointing out that in other Slavic languages as well sound changes and other developments often combined the plural accusative and nominative. However, BCMS and Slovenian acquired a distinctive accusative masculine plural by analogy, while North Slavic languages, with the exception of Czech, extended the use of the genitive for human males or, more broadly, for all animate beings in the plural. Therefore, in the Slavic context, Balkan Slavic stands out for not having "reinvented" an accusative marking distinct from the nominative for any group of plurals. Only Greek has a distinctive accusative case in the plural, which is limited to the historical Ancient Greek second declension and the definite article, whereas Albanian and Balkan Romance do not make any such distinction.

In summary, it could be said that, at the level of case markings, there is no shared Balkan invention that could have reconstituted the distinct accusative case. The Balkan Slavic DOM with the genitive singular is a common feature of all Slavic languages, although it has been lost in many modern Balkan Slavic varieties. While the invention of the definite article has increased the chance of there being an accusative case, this must be regarded as merely a by-product of the grammaticalization of the marking of definiteness. The other remnants of the accusative in Balkan languages can be attributed to internal tendencies: The case distinction is retained only if sound changes have not eliminated it. In fact, unlike other Slavic languages, Balkan Slavic has not resorted to any analogical means to preserve the distinction in the plural, with a single exception shown in example **(10)**. The effect of referential scales in the marking of accusatives seems, on the whole, spurious: The fact that the accusative is preserved more often with definite or singular referents must be attributed either to the demonstrative pronouns as the source for the case marker or else to pure chance.

Prepositional DOM in Macedonian of the Ohrid–Struga region

As described above, Balkan Romance lost the distinction between the nominative and the accusative of nouns early on, probably already during

the Common Romance era (Dimitrescu et al. 1978, 51). As in Modern Balkan Slavic, the accusative category is preserved mainly with personal pronouns. However, Balkan Romance developed an analytical marking for direct objects with human referents: Both nouns and personal pronouns denoting specific humans take the preposition *pe* (or *pi*) with the accusative/nominative, as in this Romanian example:

(36) Standard Romanian (Daniliuc & Daniliuc 2000, 282)

L-am	*vizitat*	*pe*	*bunicul*	*nostru*
him.CL-we.have	visited	PREP	grandfather.DEF.M.NOM/ACC.SG	our

‘We visited our grandfather’

Curiously, the Macedonian Slavic dialect of the Ohrid–Struga region in Western Macedonia has developed a similar analytical marking for nouns denoting humans, which imitates closely the local variety of Aromanian:

(37) Ohrid–Struga dialect of Macedonian (Markoviḱ 2007, 91)

Je	*vidof*	*čera*	*na*	*Biljana*
her.CL	I.saw	yesterday	PREP	Biljana

kəj	*go*	*bacvit*	*na*	*Borče-ta*
how	him.CL	kisses	PREP	Borče-GEN/ACC

‘I saw yesterday Biljana kissing Borče’

The similar marking of both direct and indirect objects[95] with *na* is most likely accidental and probably due to the parallelism of the prepositions *pi* and *na* in locative constructions:

(38) Aromanian (above), Macedonian (below) (Markoviḱ 2007, 92)

skafa	*esti*	*pi*	*masă*
glass.DEF	is	PREP	table

čašata	*e*	*na*	*masata*
glass.DEF	is	PREP	table.DEF

‘The glass is on the table’

In addition to this feature, limited to only a few Balkan linguistic varieties, but sometimes acknowledged as a “minor Balkanism,” both of these particular Balkan Slavic and Balkan Romance dialects share a great number of other features (Markoviḱ 2007, 69–74). What is particularly arresting in the case of the Macedonian dialect is that in example **(37)**, three different DOM features are simultaneously visible. The genitive–accusative ending in *Borče-ta* reflects the shared Slavic invention, the animate direct

[95] For the indirect objects, see example **(49)**.

object marked by the genitive, although now limited only to personal names and certain words for male relatives in the Balkan Slavic dialects (see **2.2.2**).[96] The object doubling, however, is a later innovation with a large distribution in Balkan Slavic.

Effect of referential scales on argument marking

The findings of Bickel, Witzlack-Makarevich, and Zakharko ([Submitted draft]) are compelling enough that we should try to accommodate them in our understanding of contact-induced change. Their results are slightly more ambiguous in the case of features relevant for the Balkan Slavic DOM, namely, animacy, human (and male) reference, definiteness, topicality, and specificity, but there is still no conclusive support for the universal effect of these factors. What their research implies for the study of the Balkan argument marking is that, if the effect of referential scales on argument marking is not universal, then a split as an outcome of language change along one of the referential scales is not dictated only by an innate human tendency to create such patterns. Furthermore, that the phenomenon is still significant within language families is indicative of the persistence and diachronic stability of the categories.

The most critical observation made by Bickel, Witzlack-Makarevich, and Zakharko is the areality of the phenomenon. If the effect of distinctions such as referential scales can spread within a macro area, then this presupposes that such mechanisms exist in micro-contexts as well (see the discussion of Joseph 2010 in **3.1.1**). The example from the Macedonian dialect of the Ohrid–Struga region is an indisputable example of such a mechanism in action. The structural model for the borrowed construction in Macedonian exactly parallels Aromanian. It is also important to note that the functional category of humanness as the trigger for DOM did not previously exist in precisely the same sense in the language: While animacy used to be a criterion for DOM, it developed into an accusative marking only for male personal names and with certain words for male relatives. Thus, the semantic criterion was also borrowed.

The clearest example of DOM in the Balkan linguistic area, namely, clitic doubling, which, admittedly, is only touched upon in **3.3.2**, must be considered a two-part phenomenon, similar to the prepositional DOM: On

[96] However, the case ending in *Borče-ta* is by no means a direct reflex of any Proto-Slavic case. The *t*-element in *ta* is derived from *t*-stem neuters, as in OCS *otročę* (nom.) – *otročęte* (gen.) ‘child’, while *a* reflects the *o*-stem genitive, acquired through analogy.

the one hand, there is the form, the clitic personal pronoun, which is used to double the direct object; on the other hand, there are the criteria for its use. The structural precondition for the clitic doubling, the clitic personal pronouns, emerged in Balkan Romance, Greek, and Balkan Slavic probably independently, as similar processes of cliticization took place in other Romance languages and Common Slavic. However, although possessing similar structural means, other Slavic languages do not resort to clitic doubling (see **3.3.2**), which indicates that the clitic doubling in Balkan Slavic is most likely the result of language contact. There are various criteria for clitic doubling, but all pertain to the expression of information structure and all are characterized by different levels of grammaticalization rather than entirely different motivations: The various criteria triggering the doubling include, for example, specificity, topicality, and aboutness (Friedman 2008; specifically for aboutness in Bulgarian, see Leafgren 2002, 151–157). The epicenter for the phenomenon can be located in Western Macedonia, with clitic doubling becoming less grammaticalized the further away one travels from the region (Friedman 2008).

3.4 Conclusions: Is there a Balkan case system?

There have been earlier attempts to give a unifying, systematic description of a “Balkan case system.” In her account, Tatjana Civ’jan (1965, 164–166) concludes that there are essentially two main types of case systems, both characterized by the binary division into *casus directus*, expressing the subject and the direct object, and *casus obliquus*, expressing the indirect object and adnominal attribution, typically adnominal possession. The first case system is called the “caseless system,” indicating that the difference is made analytically without case inflection (note the expanded sense of case; see **3.3.3**). This system comprises Balkan Slavic. The accusative forms in Balkan Slavic dialects, missing from the description, according to Civ’jan, must be regarded as allomorphs of the *casus directus*. The second case system, the “two-case system,” includes Albanian, Greek, and Balkan Romance. Needless to say, this kind of extreme reductionism is not particularly helpful for our task, which is the evaluation of the role of language contact in the formation of the Balkan case systems and eventually the loss of case inflection in Balkan Slavic. It chooses to ignore the Albanian ablative, the Greek dialectal accusative–dative merger, and, most important, the distinct accusative, which in places is very much alive.

Wilfried Stölting (1970, 177–220) based his analysis on Civ'jan, making, however, the important point that the Middle Balkan Slavic intermediary inflectional case system, discussed in **2.3.5**, is reminiscent of that of the other Balkan languages. Most important, the merger of the Slavic genitive and dative preceded the emergence of the analytical construction with the preposition *na*. However, it is characteristic of Stölting to somewhat carelessly attribute several phenomena with expressed certitude to one or another source language. For example, he claims that the analytical Slavic expression of recipient and adnominal possessor with *na* and adnominal possession with *ot*, must *unmistakably* (*unverkennbar*) originate from the use of early Romanian cognates of the Latin prepositions *ad* and *dē*, discussed in **3.2.2** (ibid., 193). Obviously, this view ignores entirely the analogous Greek structures (see **3.2.3** and **4.2.3**). Illustrating the same tendency, Stölting (1970, 191) claims, without further justification, that the Romanian retention of the indefinite feminine singular distinction between the nominative–accusative and the genitive–dative is likely to have resulted from an Albanian substrate.

It is obvious that a comprehensive, but compact description of the "Balkan case system" is not possible without detrimental generalizations. Furthermore, attributing any of the characteristics of the case systems to one particular source is difficult. One of the reasons has been pointed out by Jouko Lindstedt (2002, 305), who writes: "[i]nnovations are more easily shown to be contact-induced than retentions are." This is clearly true as well for Balkan case systems. Intuitively, there is little credibility, let us say, in the idea that the preservation of the distinction between accusative and nominative masculine nouns in Greek is the result of the influence of the differential accusative for animate masculines in Slavic. Although partly governed by similar semantic criteria, since the separate accusative occurs with masculine singulars, the Greek situation is easily explained by regular sound changes, while in Slavic the genitive–accusative is an innovation encompassing related languages outside the Balkans. This is not to say that retention of certain distinctions cannot result from language contact, even if proving it may be difficult. What makes Balkan Romance different from all Western Romance languages is precisely the preservation of a case

distinction whose pattern is very similar to other Balkan languages, making language contact a credible explanation for its survival.[97]

In fact, the only characteristic of the inflectional case systems in the Balkans, including the Middle Balkan Slavic case system, to show remarkable uniformity is the genitive–dative case. However, to determine whether this feature is representative of retention or innovation is complicated. Its grammaticalization involves the process of equaling the expression of indirect objects, prototypically recipients, with adnominal, "genitival" modifiers, typically possessors. Following the steps taken by Thomason (2010), presented in **3.1.1**, to determine whether a feature originates from language contact, we obtain the following picture: 1) There is a great deal of structural interference among the languages involved. 2) Although identification of the source language seems very difficult, the contact between languages has been intense, owing to long-standing mutual multilingualism. 3) The shared structural features in the proposed contacting languages involve a similar marking for both recipients and adnominal possessors.

Steps 4), the status of the receiving language before the contact, and 5), the status of the source language, prove to be the most difficult, since they involve identifying the status of a feature prior to the contact. There are various sources for the case forms. Only Balkan Slavic and Greek nouns display a single source, Balkan Slavic the dative, Greek the genitive. Albanian forms originated in the singular from the dative and locative, in the plural from the genitive; the Romanian forms originating in the singular continue the dative or the genitive–dative, and in the plural, they continue the genitive. It is impossible to determine whether any of these structures was more established in any of the languages before the assumed contact, with the exception of Balkan Slavic, which must have developed syncretism to the full only after the beginning of the OCS period, that is, in the 10th century at the earliest. The history of Albanian is, of course, the most difficult to evaluate. Yet if we take into account the cautious note, made in **3.2.1**, that the ablative may have been used to mark adnominal possessors,

[97] There is another possible explanation: Since Balkan Romance has developed partly in isolation from other Romance languages, practically as a language island, the potential innovations leading to the total loss of case inflection may not have had a chance to spread there. However, this explanation does not have to be seen as an alternative to the contact explanation; it may have contributed to the retention of case inflection together with language contacts with non-Romance languages. On the other hand, as was shown in **3.2.2**, many properties of the Romanian case inflection involve innovations, which means that the term retention is only partially applicable.

then we may speculate that there was at least some competition between two structures in marking this function.

However, there are two additional points: The Greek genitive is argued to have been in competition with the accusative in replacing the dative during the first millennium, and the use of analytical marking to express the recipient and the adnominal possessor in Balkan Romance has been claimed to have decreased after the first attestations in the 16th century. Judging from these observations, the firm and established role of the genitive–dative must have been the result of a long process. I believe that during the period perhaps roughly from the 6th to the 16th centuries, the mechanisms of language contact, operating through mutual bilingualism, helped to cement this feature in the Balkan languages.

The greatest exception to the genitive–dative merger, namely, the accusative replacing the dative instead of the genitive in northern Greek dialects, is somewhat problematic for a contact explanation of the genitive–dative merger. The dialects that, from a purely geographical point of view, are adjacent or that overlap with the areas where Albanian, Balkan Slavic, and Balkan Romance are spoken lack this key feature, which is thought to have resulted from language contact. However, if we assume that the accusative replaced the dative very early in these varieties, then any possible effect of the language contact may have taken place too late, since there were no competing structures whose selection could have been influenced by an extension of the semantic properties of a case in another language.

The picture that emerges from the analysis of the Balkan case systems is less reminiscent of an all-encompassing convergence toward a unified case system. Rather, what is characteristic of the argument marking in the Balkan languages is the convergence of individual features, sometimes with noticeably local occurrences, such as the prepositional DOM. Another typical property of the Balkan argument marking is illustrated by clitic doubling where the key factor for the successful diffusion of the feature must be based on the explicit nature of the structure: The elements in the structure are identifiable cross-linguistically, and all the Balkan languages possess the structural means to mark it. It is important to point out that, although the genitive–dative merger involves case inflection, which is deemed less explicit and therefore potentially less prone to spread, the invention of the genitive–dative merely involves the idea that the same form used to express one function can be used to express another one as well.

Finally, the question of the loss of case inflection in Balkan Slavic must be addressed. As concluded earlier, there are no mechanisms of contact-induced language change that would force an entire case system to fit a uniform template. The similarities between the individual languages are due to the diffusion of separate features, which may or may not be adopted, the genitive–dative case being a good example. This means that there is no feature that reads: "lose your case inflection." As argued in **2.3.4**, although some case distinctions were lost because of sound changes, similar changes did not lead to the loss of case inflection in other Slavic languages. What happened was that the increase in homonymy was not compensated for by "reparatory" analogical change as in other South Slavic languages, but the speakers accepted the resulting ambiguity or resorted to analytical constructions. However, the relatively rapid decline of case inflection cannot be attributed only to uncompensated sound changes or to new, contact-induced innovations. Although not without its problems, I view the hypothesis of the effect of a large number of L2 speakers as being credible. I would even go so far as to argue that this type of indiscriminate effect of language contact may amount to the single most important factor for the loss of case inflection in Balkan Slavic.

4 The loss of case inflection in Balkan Slavic from a typological perspective

In this chapter, certain shared characteristics of the Balkan case systems, defined in **Chapter 3**, are evaluated from a typological perspective, that is, from a cross-language point of view. The need for a comparative typological approach arises from the very nature of areal linguistics as the study of linguistic similarity within an area by comparison with the languages outside that area; the shared features under scrutiny become areally significant only when their absence outside the area is established. This prerequisite provokes certain questions from a cross-linguistic viewpoint that Balthasar Bickel and Johanna Nichols (2006, 3–4) put in the following way, first generally and then specifically for the Balkan linguistic area:

> Suppose the linguist sorts through 200 variables and finds that five of them appear to be area-defining. Is this a significant result, or could one expect to find five out of 200 shared variables for any random set of languages and any random set of variables? […] Our impression is that the classic Balkan features include a few variables of sufficiently low [cross-language] frequency to be of diagnostic value […] [T]he classic Balkanisms do not do a very complete job of defining the shared grammar that makes for the notable intertranslatability of Balkan languages.

The implied criticism in the opening lines of this passage can be interpreted in two ways. On the one hand, similarities between languages can result either by chance, by some connection between them, or by a linguistic universal. These variables, that is, linguistic features, are generally defined in a way that yields a finite set of alternative values: for example, a language may or may not have a definite article. The frequency of definite articles in the languages of the world provides certain odds for the occurrence of an article in any given language, most likely indicating that there is a fairly good chance that any given language has a definite article and that the likelihood that there is a definite article in two adjacent languages is still relatively strong. Therefore, a set of features shared between geographically-connected languages does not necessarily amount to a linguistic area: To list five random, synchronic features, for instance, is clearly not enough, since these could easily emerge at random. On the other hand, Bickel and Nichols's further observations about the Balkan linguistic area are relevant in the sense that the occurrence of typological rarities are of better

diagnostic value. Superficially, Bickel and Nichols's phrasing seems to imply that typological banalities are less important. This is not the case, as the authors themselves go on to point out (see **4.1.2**). This chapter sets out to demonstrate that there are also other diagnostic criteria for a contact phenomenon than mere typological rarity of a particular abstract feature. Both of the case studies presented in **Sections 4.2** and **4.3** are intended as partial answers to the problems illustrated above, but the focus is different. The strategy of this chapter is summarized as follow:

Section 4.1 discusses the relationship between linguistic typology and areal linguistics. The sometimes lukewarm attitude of typologists to the way areal linguistics is conducted appears to be occasionally reciprocated also by contact linguists in their views of linguistic typology. I argue, however, that the major points of criticism are acknowledged and discussed also in modern approaches to typology. Yet, while the methods are still developed to overcome difficulties pointed out by critics on both sides of the fence, the two fields can benefit from each other, or perhaps even more: their key findings are of essential value also for the other.

In **Section 4.2**, I assess the genitive-dative merger from the point of view of its uniqueness when compared with similar mergers or overlap of functions in other, mostly European languages. The data for the comparison is limited, but I intend to show that although some similar developments are characteristic of other European languages, too, the level of grammaticalization and the geographical distribution in the Balkans is enough to consider the merger of dative and genitive an areal feature. I further argue that also from a purely synchronic point of view this feature remains distinctive enough despite of what appears as its typological banality. Yet, reaching such a conclusion needs more fine grained data than what is often available for typological generalizations.

Section 4.3 seeks new ways to understand the co-occurrence of two Balkan features, the overall loss or reduction of case inflection and the grammaticalization of a definite article. Using quantitative typological data, I explore a connection between the two phenomena in the languages of the world. However, since the possible connection does not imply causation, and the mutual diachrony of these two phenomena remains mostly opaque, I attempt to find the areas of grammar where these two features could potentially interact and therefore influence each other. An important question arises from the analysis: if these two phenomena are universally

interconnected, does it then imply that they are of lesser value as defining features of the Balkan area?

This chapter does not have a separate section for conclusions, but the combined results are discussed in **Chapter 5.**

4.1 Linguistic typology and areal linguistics

This section addresses questions regarding the compatibility of linguistic typology and areal linguistics and endeavors to present the role played by areality in modern approaches to typology. The nature of typological evidence is also discussed from an areal linguistic point of view.

4.1.1 Areal linguistics and linguistic typology

Typological approaches to contact linguistics are not always well received by scholars of established linguistic areas. Victor Friedman, having in mind Standard Average European (SAE) as proposed by Haspelmath (1998), among others, blames typologists for what he calls "Eurology" (Friedman 2011, 3). According to him, SAE is based on an ideology, placing the heart of the linguistic area in the territory of the Holy Roman Empire, which corresponds geographically to the nucleus of the European Union, political entities neither of whose existence is connected to any viable socio-linguistic hypothesis for a language contact phenomenon. Although Haspelmath (2001, 1507) does, in fact, tentatively present such a hypothesis, for Friedman, a key problem remains: According to him, areal and genealogical linguistics are essentially diachronic fields, whereas typology is synchronic. In the same vein, Brian D. Joseph (2008) demonstrates at greater length the kind of misinterpretations to which ahistorical, feature-based synchronic analyses can lead in the study of linguistic areas. He proposes, for the sake of argument, a number of morphophonological phenomena that, from a synchronic point of view, might seem to be the result of contact-induced convergence among the Balkan languages. Yet he refutes them one by one, showing that with historical knowledge of the languages involved, a contact hypothesis becomes implausible.

Friedman (2011, 3–4) also sees another minefield, "numerology." His criticism is twofold, and it involves both the way features are selected for analysis and the varieties of languages chosen to represent the bigger picture: Merely listing shared features among languages and checking off whether or not a feature appears does not do justice to the very nature of language contact. The potentially detrimental effect of "numerology" is well

demonstrated by Friedman's analysis of object doubling in the Balkan languages (Friedman 2008). Using dialectal data, he shows how the level of grammaticalization forms an epicenter, the feature becoming less grammaticalized toward the periphery, an observation that would obviously be lost in oversimplified yes/no questions. However, taking only standardized varieties of languages as points of comparison might pose an even bigger risk. Dialectal data increase the granularity of the evidence, revealing important dynamics of phenomena like the Balkan object doubling, which would show much more indecisive distribution with data from standard languages only. Moreover, there are indications that relying on standardized varieties not only decreases the granularity and overall quality of the data, but even might skew the analysis in a particular way. Guido Seiler (2011) lists seven features of the proposed SAE, including relativization and pro-dropping, and shows how dialects of German, one of the nucleus languages of SAE, portray a situation that is considerably less "European" than the one based on the written standard language. Seiler suspects that the common pathways of language standardization are to be blamed for this.

All the criticism presented above is, no doubt, relevant. I do not, however, see areal linguistics and typology as being fundamentally opposite or contradictory approaches. Neither field can be called a school of thought, with theoretical presuppositions that question the other's achievements. The sometimes conflicting views result from the different starting points: Linguistic typology attempts to chart and explain universal similarities, differences, and co-occurrences, as well as the patterning of features in the context of all languages. Therefore, an analysis of a linguistic universal (for a definition, see **4.1.3**) might involve observations about an areally significant distribution of the phenomenon. Areal linguistics, on the other hand, typically studies languages in contact and assesses the similarities between them that cannot be satisfactorily explained by such things as common genealogical inheritance. Yet the opposite emphasis on the premises of areal linguistics can also be found. Colin P. Masica (2001) gives a comprehensive list of 23 potential pitfalls in the study of linguistic areas. Masica (ibid. 207) does insist that the study of pre-defined or assumed linguistic areas must include the characteristics and histories of individual languages, but maintains that this is logically only secondary: "Primarily, areal linguistics should mean *the study of the significantly non-random distributions of linguistic features* *<u>in space</u>*—first of all the facts and if

possible the reasons behind them." Masica (ibid.) reminds us, however, that, in addition to language contacts, non-random distributions can result from genealogical relationship and typological correspondence.

While keeping Friedman and Joseph's criticisms in mind, it is nevertheless easy to agree with Masica's view that it is not important how a hypothesis about areal convergence comes into existence as long as the linguistic phenomenon in question is empirically sound.[98] In fact, as shown in **Sections 1.4** and **3.1**, the early research into the Balkan linguistic area was based on observations about convergence among the individual languages—unaccompanied by clearly stated hypotheses about its emergence. In addition, when the first attempts to account for the phenomenon surfaced, they often involved theories about the effect of a single Proto-Balkan substrate language, many of which have been refuted in later studies. Yet the observations about the structural convergence made by these early scholars of the Balkan linguistic area are not diminished by the lack of a solid historical socio-linguistic hypothesis. Furthermore, it follows from Masica's definition that a comparative typological analysis of the given features becomes a precondition: the significantly non-random distributions of features can be established only through comparison with other languages outside the assumed linguistic area.

4.1.2 Areally-informed quantitative linguistic typology

In recent decades, the study of language contacts has experienced an intrusion by another wave of typologists, who have effectively created a novel idea about the way hypotheses about linguistic areas are formed. In quantitative typology, controlling the effect of language contact in one or more ways when assessing proposed linguistic universals has become routine (see e.g. Dryer 1992; Bickel 2008, 4–6) As a side effect, multivariate analyses of typological data can produce outcomes that show statistically significant areal distribution of features, even when no preconceived areal hypotheses exist (for one such study, see **4.2.2**). Needless to say, in the sense of Friedman and Joseph's viewpoints, these observations about areal diffusion of features do not take into consideration diachrony and historical socio-ethno-linguistic facts. Nevertheless, any such study cannot remain

[98] I do not believe, however, that there is genuine disagreement among these three scholars on any other main questions regarding areal linguistics. They all quote one another frequently and generally in an approving manner. Moreover, Masica expresses concerns similar to Friedman's and Joseph's, some of which are clearly directed to typologists; see, e.g., Masica 2001, 208.

entirely agnostic on areal preliminaries. Several heuristic divisions are used, ranging from two macro-areas to twenty-four smaller areas, often based on geographical regions such as the hemispheres, continents, or subcontinents (see e.g. Dryer 1989; 1992; Bickel & Nichols 2006; Bickel 2008; Sinnemäki 2014).[99]

Yet it is clear that an areally-significant distribution of individual features detected with quantitative typological analyses can create, at the most, hypotheses about linguistic areas, which can then be tested with other methods. Thus, Balthasar Bickel and Johanna Nichols (2006) abandon the typological approach and attempt instead to find confirmation for a hypothesized linguistic area through statistical analysis of a set of variables in one region by comparison with the languages of the world. In other words, what these scholars do is, in essence, areal linguistics, but with methods mostly used in quantitative typology. Bickel and Nichols set out to find support for a theoreticized linguistic area consisting of languages in the Circum-Pacific area (extending somewhat deeper into the mainlands than just the Pacific Rim). The languages of the Circum-Pacific area have been found to display similarity in distribution of features, such as numeral classifiers, head marking, and *n* – *m* personal pronouns (ibid., 3–4). From human genetic and archaeological findings, the Circum-Pacific area is known to have functioned as a contact and migration zone for a long time. Bickel and Nichols account for an obvious problem of the very long time span that any contact explanation has to address by considering the observed features as being generally more persistent areally than within language families because their retention is favored by areal pressure (ibid., 4–6).

Bickel and Nichols observed 100 variables in a genealogically-controlled sample of a maximum of 316 languages, data for all the variables not being available from all languages. The variables were selected only by their availability in databases, not by any preconceived hypotheses about their

[99] It is likely that, in the future, hypotheses about areal convergence will be generated automatically from large databases, independent of any preconceived geographic or other areas. In fact, projects like WALS (World Atlas of Language Structures) already include maps with distributions of linguistic features, offering researchers a visual impression that could be used for the same purpose. Yet these endeavors are hampered, not so much by limitations of computational power available—perhaps relevant for some other fields—as by the quality, granularity, and organization of the data. The greatest challenge, although by no means the only one, is how to add a diachronic level to the data that would include known population migrations, already established language contact phenomena, etc. In any case, the future development, maintenance, and refinement of databases like WALS and AUTOTYP (http://www.spw.uzh.ch/autotyp/) is of crucial importance for any such goal.

role in the areal phenomenon. After controlling for the distributional independence of the variables, the authors were left with 86 variables of which 30 to 40 percent show significant areal distribution, regarded therefore as likely features of the Circum-Pacific linguistic area. (Ibid. 6–9.) Bickel and Nichols call their approach Predictive Areality Theory (PAT). PAT is a historical explanation (in the sense of resulting from any kind of contact-induced change or common inheritance) for non-accidental overlap of typological variables. The authors emphasize that, in this definition, areality is not a property of languages, but is identified as variables or sets of variables, and therefore is not a typological observation. Rather, in terms of PAT, areality is a theoretical predictor variable, predicting observable typological distributions. What is important in light of the criticism presented above in **4.1.1** is that, according to Bickel and Nichols, for PAT to work, it must be grounded in what is known about population history from such fields as archaeology, genetics, ecology, geography, economics, and demography.[100] (Ibid., 3.) Also significant for the discussion in this chapter is the authors' reminder that an area-defining feature does not need to be cross-linguistically rare (ibid., 2).

Language contact often competes with genealogical relatedness as a historical explanation for observed similarities between languages. According to PAT, the effect of known language families is controlled for by sampling methods. Nevertheless, PAT or any similar approach cannot disambiguate between contact-induced similarities and similarities resulting from common inheritance that stems from a deeper historical layer antedating the established language families. Yet recent methodological developments in quantitative approaches to typology have encouraged some researchers to claim that genealogical relatedness could be established with

[100] It must be noted that combining genetic and archaeological data to explain linguistic contact phenomena is not without its problems: Languages, archaeological cultures, and genes do not necessarily spread hand-in-hand. Nevertheless, in this sense, the proposed Pacific Rim or Circum-Pacific linguistic areas lie on relatively solid ground, since the genetic and archaeological arguments are used to account for the spread and retention of linguistic features rather than assuming a one-to-one correspondence among languages, genes, and archaeological cultures. An interesting observation is made by Johanna Nichols and David A. Peterson (1996), who provide evidence for a close correspondence between certain areal linguistic features and one mitochondrial DNA lineage in the Pacific Rim area, features characterizing several language families that cannot be shown to be genetically related.

typological data only. In his criticism of one such controversial claim,[101] Mark Donohue (2012) uses a hypothesis involving the Balkan linguistic area languages to determine how well different typological datasets conform to known genealogical subgroups. Donohue (ibid., 102–103) claims, based on a sample of 36 Indo-European languages spoken in Europe, that mere morphosyntactic features replicate Celtic and Germanic subgroups well, forming visual clusters in a diagram created with the NeighborNet agglomerative algorithm. Romance and Slavic, however, show less clear clusters, because Romanian, Bulgarian, and Macedonian are bundled together with Albanian and Greek and thus are distanced from their genealogical subgroups. However, the analysis of phonological features leads to much more indecisive groupings, which, although often involving geographically adjacent languages, do not conform to the subgroupings or limit themselves to the languages of the Balkan linguistic area (ibid., 103–108).

These observations illustrate both the unsuitedness of typological features in establishing genealogical relatedness and the challenges that the effect of areality poses for the quantitative analyses of linguistic data. Nevertheless, while Donohue is making a methodological point, his approach is also very interesting from the perspective of the Balkan linguistic area, since he finds quantitative support for this area by applying a method that is, in essence, similar to Bickel and Nichols's PAT. In addition, because PAT is based on a random choice of features mainly to avoid favoring already known areal features, PAT could perhaps be further enhanced to test the effect of

[101] Michael Dunn et al. (2007) strongly suggest that there is genealogical relatedness among the Papuan languages, which are too separate in space and time to show any convincing cognate words. The authors base their claim on a statistical comparison of abstract structural features, which they present as a phylogenetic tree of the languages. In their model, the bifurcations, based mainly on typological distance and shared features, represent a hypothesis about the genealogical relatedness of the languages. Anticipating criticism, Dunn et al. tested their model on languages from the Western Oceanic branch of Austronesian languages, claiming that the model, based on typological data, can in fact replicate the phylogenetic tree of Western Oceanic, already established with the comparative method. In their response to Dunn et al., Mark Donohue, Søren Wichmann, and Mihai Albu (2008) refute, as do most historical linguists, the weight of typological data as evidence for genealogical relatedness. These three authors first point out that some of the classifications within the Western Oceanic group and used in controlling the model are still in dispute; second, they show that the tree produced with the method of Dunn et al. in fact does a far better job of replicating the geographical proximity of the languages than their level of genealogical relatedness (ibid. 2008, 225–228). Consequently, Donohue, Wichmann, and Albu argue that what the clustering of typological features in the Papuan languages shows is an areal—not a genealogical—correlation (ibid., 230–231).

qualitatively different datasets. This might improve the understanding of a type of phenomenon that is particularly prone to be diffused areally.

In addition to controlling for the effect of areality, another issue that ranks high among the desiderata of quantitative typology is the greater granularity of data, the lack of which could be described in Friedman's terms as "numerology." This granularity can be achieved by searching beyond what Johanna Nichols (2007, 233) calls "lookup characters," the relatively superficial structural properties available in grammars simply by looking them up. In Balthasar Bickel's words (2007, 245), "[l]inguistic diversity is captured by large sets of fine-grained variables, not by grand type notions." An important tool for obtaining better data is a questionnaire that maps not only the question at hand, but also any phenomena known or hypothesized to be cross-linguistically related to the issue (see e.g. Comrie, Haspelmath & Malchukov 2006 for a questionnaire on ditransitive constructions).

Another question pertaining to the granularity of data is the level of units of observation—in typical accounts, individual languages. As pointed out in **4.1.1**, the inclusion of only standardized varieties in analyses may pose a risk, for instance, owing to the common linguistic thought behind their historical codification. Balthasar Bickel (2011) goes as far as to suggest that we should reconsider the status of languages or dialects as basic data units. Instead, we should consider it an empirical question whether a single value of a feature assigned to a particular language or dialect can, in fact, be representative of the variation within that particular linguistic variety. Bickel shows through two case studies how the variation in referential density, the degree to which speakers use overt referential expressions such as pronouns, and certain case alignments do not necessarily form patterns in a way that makes individual languages stand out, although the variation may be satisfactorily accounted for by some other factors. His first case demonstrates how referential density is subject to a great deal of variation between individual speakers, best explained by sociolinguistic factors affecting the speaker: Growing up in a close-knit society leads to low referential density. His second case shows that individual languages do not necessarily reveal a clear preference toward any particular alignment, whereas such a bias can be clear at the level of language families. Bickel reaches a diachronic conclusion, namely, that the {S,A,..}[102] type, meaning

[102] The formula means that at least S (the single argument of intransitive verbs) and A (the agent argument of transitive verbs) are marked with the same case.

an accusative or neutral alignment, is preferentially retained in those language families where it exists and is preferentially innovated where it does not exist.

As demonstrated here, many concerns expressed in contact linguistics about the methods of linguistic typology are not only acknowledged in recent studies in typology, but also are increasingly considered as deserving a more central role in the creation of new research frames and testing new hypotheses. Moreover, the approaches presented here put Friedman and Joseph's observation about typology as an essentially synchronic field in a new light. If areal and genealogical components are controlled in typological research, then by definition, they bring a diachronic layer to the analyses. This in turn enables areally- and genealogically-informed quantitative typology to create new hypotheses about linguistic phenomena as being prone either to be retained among genealogically-related languages or to be diffused through language contact. These observations provide invaluable comparison for the typology of the effect of language contact, acquired through a meta-analysis of previous case studies and outlined, for instance, by Sarah Thomason (2001, 85–91), Alexandra Aikhenvald (2006, 4–7), and Peter Trudgill (2011, passim). In the following subsection, I will further explore some of the possibilities of how typological approaches and methods might benefit areal linguistics.

4.1.3 Typological universals as explanations

An important concept in the epistemology of linguistic typology is typological universals. These universals remain central to all typological research, regardless of whether individual approaches to typology are characterized by the study of the shared properties between languages or the classification, distribution, and frequency of the variation of these, either from a qualitative or a quantitative point of view. Linguistic universals can be defined as either absolute or probabilistic, also called statistical universals, the latter involving most, but not necessarily all languages. This basic notion implies the following:

> Absolute universals hypothesize that a grammatical property must be present in a language. Probabilistic universals say that a grammatical property is present in languages with some degree of likelihood. There is also a third way of constraining what does and does not occur in languages: by stating what is universally possible, without being necessary or even probable. (Moravcsik 2011, 70.)

Both absolute and probabilistic universals can be further divided into implicational or restricted and non-implicational or unrestricted universals. In Edith Moravcsik's (2011, 70) formulation, the latter involve statements, such as "in all languages, there is feature X," whereas restricted or implicational universals are expressed with conditional statements, such as "if there is feature X, then there is also feature Y."

Moravcsik (2011, 71–74) continues by analyzing the role of linguistic universals as explanations. She takes a language-specific question as a starting point: Why do adpositions precede their noun phrases in English? A structural explanation involves assigning the phenomenon, in this case the adpositions, a more general structural quality: In English, adpositions as non-branching constituents most often precede branching constituents, such as noun phrases. This *language-specific structural generalization* is then explained by a *universal structural generalization*: In most languages, either all or most non-branching constituents precede branching ones or all or most non-branching constituents follow branching ones. The problem with this kind of probabilistic universal explanation is that it is not causal. According to Moravcsik, causal explanations need to trace the temporal, that is, diachronic, process whose input lacks the explanandum and whose output contains it. Again, a *language-specific historical generalization* is first needed: All English prepositions whose source is traceable within the history of the language have grammaticalized from constructions that have merely retained their linear order. The *universal historical generalization*, a probabilistic universal, then reiterates the same observation about the way most adpositions have come about. Finally, the historical process requires a *functional universal explanation*: In all languages, the semantic and phonological reduction of frequently occurring phrases serves ease of production without impairing comprehensibility. A changing linear order in the process does not enhance either production or comprehension.

This characterization of typological universals as explanations serves as the basis of the discussion in **Section 4.3**, which explores the emergence of a probabilistic, restricted universal related to the Balkan case systems. After the statistical analysis of quantitative typological data, the section discusses the implications of a universal tendency behind a phenomenon which is widely thought to have resulted from language contacts. The following section, **Section 4.2**, takes a closer look at one classical Balkanism and endeavors to show that the inclusion of more fine-grained data in the analysis gives a fuller picture and reveals the complexities of the contact-

induced dimensions of the phenomenon. The Balkan recipient–possessor merger is cross-linguistically contrasted with similar phenomena to find out whether the emergence of the phenomenon can be credibly attributed to language contact. The section gives an idea of the reasons behind the elusive "intertranslatability" among the Balkan languages, since many of the interconnected phenomena observed in the section are not included in general accounts of the Balkan linguistic area.

4.2 The Balkan recipient–possessor merger

The identity of the marking of the recipient and adnominal possessor, or "dative–genitive merger," as the phenomenon is most often called, is one of the most frequently acknowledged Balkanisms, included in all modern general accounts of the Balkan linguistic area (see e.g. Asenova 2002, 81–86; Solta 1980, 134–141; Schaller 1975, 205–210). The "dative–genitive merger" in the Balkans has also been addressed in dedicated articles (e.g. Galton 1967; Qvonje 1979; Catasso 2011; Pompeo 2012). The source of the merger remains in dispute, as is typical of the Balkan linguistic area features (for a summary of the proposed theories, see Schaller 1975, 138–141). Also, several other questions remain unanswered or insufficiently clarified. A basic problem is the inclusion of Balkan Slavic in the "dative–genitive merger," since most modern varieties do not display case inflection, but express the functions of the lost genitive and dative with analytical constructions. On the theoretical and methodological levels, the phenomenon raises a complex issue: The typological commonness of such a merger calls for discussion of the criteria by which a feature can be called areal (see **4.1.2**). On the basis of the discussion in **4.1.1**, I argue that this question can be addressed only by means of a more fine-grained analysis of the merger in the Balkans and by comparison with similar mergers in other languages.

This section has two main goals: First, I attempt to map out the scope of the phenomenon, trying to reconcile the manifestations of the merger on different morphosyntactic levels. I argue for a functional term, recipient–possessor merger: The morphological dative and genitive cases form only part of the phenomenon, and the merger of the historical cases is not complete in any of the languages. Furthermore, functionally and typologically, more relevant terminology renders the analysis cross-linguistically comparable, necessary for my second goal, which is to contrast the Balkan phenomenon with similar mergers or overlaps of functions in

European languages. While a quantitative typological analysis of the merger is beyond the scope of this study, a more modest cross-linguistic analysis is carried out to establish its areal limits and further characteristics.

I argue that in addition to the two subphenomena of the recipient–possessor merger—the merger of the morphological cases and the identity of analytical structures with a preposition expressing goal/location—certain phenomena closely related to adnominal possession cannot be left out of a comprehensive account. These include the following: 1) a double marking strategy for adnominal possession and a split marking of possession, bringing the Balkan languages typologically closer to the adjacent Hungarian and Turkish; 2) an areally more restricted expression of adnominal possession with a preposition indicating source; and 3) a pseudo-partitive nominal construction expressed with juxtaposition. All these features pertain to a particular area of nominal syntax, namely, adnominally used nouns. Their areal distributions differ in the Balkans, although the features are typically found in areas that are geographically contiguous. I believe that acknowledging this partial overlap of several closely related areal features is crucial to a better understanding of the common, but fuzzy notion of a "shared grammar" in the Balkan linguistic area (see Bickel & Nichols 2006, 3–4, discussed at the outset of **Chapter 4**).

4.2.1 The merger of the morphological cases

What is most often meant by the merger of the dative and genitive cases in the Balkan languages is illustrated in the following examples:

(39) Modern Greek:

tis	*édosa*	*to*	*vivlío*	***tis***	***Élli-s***
she.GEN.SG	I.gave	DEF.ACC.N.SG	book.ACC.N.SG	DEF.GEN.F.SG	Elli-GEN.F.SG

'I gave the book to Elli'[103]

(40) Modern Greek:

i	*gonís*	***tis***	***Élli-s***
DEF.NOM.M.PL	parents	DEF.GEN.F.SG	Elli-GEN.F.SG

'Elli's parents'

[103] Here, the Greek single-case form is called genitive, whereas in Balkan Slavic it is called dative, according to the prescriptive traditions of the languages. In Albanian and Balkan Romance, the name dative–genitive is used to indicate the single morphological case form: In the respective prescriptive traditions, however, an artificial distinction is maintained between the dative and the genitive, depending on whether the use is clausal or adnominal. For discussion, see **3.3.3**.

(41) Romanian:

i-am	*dat*	*carte-a*	***Dacian-ei***
she.DAT/GEN.SG-AUX	given	book.NOM/ACC.F.SG-DEF.NOM/ACC.F.SG	Daciana-DAT/GEN.F.SG

'I gave the book to Daciana'

(42) Romanian:

părinții	***Dacian-ei***
parents.NOM/ACC.DEF	Daciana-DAT/GEN.F.SG

'Daciana's parents'

(44) Albanian:

i-a	*dhashë*	*librin*	***Ilir-it***
he.DAT.SG-it.ACC.M.SG	I.gave	book.ACC.M.SG	Ilir-DEF.DAT/GEN.M.SG

'I gave the book to Ilir'

(45) Albanian:

prindërit	*e*	***Ilir-it***
parents.NOM.M.PL.DEF	LINK.NOM.M.PL	Ilir- DAT/GEN.M.SG.DEF

'Ilir's parents'

These standard language examples show how a single case form is used to express both the indirect object of ditransitive verbs, known in functional terms as the recipient, and the adnominal possessor, two functions that were historically the properties of separate morphological cases, the dative and the genitive. Similar identity of forms and functions is often observed with the personal pronouns, discussed in more detail in **4.2.4**.

(46) Modern Greek, compare with the recipient in example (39):

i	*gonís*	***tis***
DEF.NOM.M.PL	parents	she.GEN.SG

'her parents'

While modern Balkan Slavic standard languages lack case inflection, several dialects still preserve the dative case for a limited set of nouns in the singular, typically person names and terms denoting humans. As argued in **Chapter 3**, these dialectal archaisms are vestiges of the interim case system of the Balkan type that Balkan Slavic too used to represent: The dative is used in both the recipient and possessor functions, as examples **(47)** and **(48)** show (for historical use of the dative in the marking of the adnominal possessor, see **2.2.2** *Genitive replaced by prepositions*):

(47) Balkan Slavic, Macedonian dialect of Kičevo (Labroska 2008, 145):

mu	*dadof*	***Trajče-tu***	*pila*
he.DAT.M.SG	I.gave	Trajče-DAT.M.SG	saw

'I gave Trajče a saw'

(48) Balkan Slavic, Macedonian dialect of Kičevo (Labroska 2008, 145):

sin	*mu*	***Stojan-u***	*pošol*	*vojnik*
son	he.DAT.M.SG	Stojan-DAT.M.SG	started.out.as	soldier

'Stojan's son became a soldier'

An important exception to the merger of the two case forms are the Northern Greek dialects, including Pontic and Cappadocian Greek, which display, not the merger of dative and genitive, but the merger of dative and accusative (Horrocks 2010, 284).[104] Furthermore, while the standard Albanian, Greek, and Romanian show case syncretism, together with earlier and some modern varieties of Balkan Slavic, at the same time some individual varieties of both Greek and Balkan Romance have lost their case inflection to a greater degree than have the standard languages. Needless to say, these dialects, like modern Balkan Slavic, resort to analytical ways of marking both functions (see **4.2.2** for discussion).

Examples **(39)** through **(48)** illustrate, in most instances, the merger of the two historical cases, which, as discussed in **Chapter 3**, seems to have been largely completed before the second millennium, perhaps with the exception of Balkan Slavic being a latecomer. However, speaking about a case merger becomes problematic, as the new case has several origins in the languages: While in Slavic the case originates from the dative (see **2.3.5**), Romanian shows a two-fold origin from both the dative and the genitive (**3.2.2**), while the Greek case originates from the genitive, with the exception of some pronouns that have their origin in the accusative (**3.2.3**). Albanian, unlike the other languages, does not have an attested earlier form displaying a different or more complete case system, but, by comparison with Proto-Indo-European, the Albanian dative–genitive shows a variety of origins: the locative, genitive, and dative–locative (**3.2.1**).

In addition to the diversity of its origins, the new, merged case does not represent the entire range of functions that the earlier genitive and dative cases used to perform. In Albanian, because of the diverse origins of the case and the lack of knowledge of any intermediate systems, a "merger" of the dative and genitive is not a very good term. In Proto-Slavic, the genitive

[104] Also in standard Greek, the accusative has replaced the dative and genitive with the clitic plural personal pronouns and first- and second-person full personal pronouns.

was used with many prepositions and as the complement of several verbs. None of these functions was overtaken by the dative. Both the Greek dative and genitive were also used with several prepositions, most of which are now used with the accusative only. All Romanian prepositions inherited from the Latin prepositions understandably take the accusative, since most Latin prepositions either took the accusative or the ablative, only rarely the genitive, and never the dative. Yet the genitive functioned as a verbal complement in Latin; most of these constructions now contain either the accusative or a prepositional construction, not the bare dative–genitive case. As will be shown in the next subsection, the different historical developments introduced here are sometimes erroneously held against the dative–genitive case being included in the list of features of the Balkan linguistic area.

4.2.2 Recipient–possessor merger: Between verbal and nominal syntax

Nicholas Catasso (2011) questions whether the dative–genitive merger is a Balkanism, that is, the result of contact-induced convergence between the Balkan languages. While unfortunately riddled with mistakes and unfounded claims,[105] Catasso's argument nevertheless raises three questions relevant here, which need to be addressed. First, according to Catasso (ibid., 87–88), the merger is not complete, that is, the "genitival" functions are not expressed exactly the same way in all languages as the "datival" functions. Second, the merger is not realized in exactly the same way in all languages, nor do the merged cases have a common origin in all languages. Third, such a merger is typologically too common to count as area-defining. In other words, there is no merger, and if there is, it is not similar in all languages, and should it be, it would not count anyway.

[105] Catasso (2011, 78) erroneously claims, for instance, that "Macedonian is generally considered as [*sic*] a dialect of [Bulgarian]." In addition, Catasso (ibid., 87) seems to hold the diverse origins of the merged case (e.g., in Greek, mostly from the genitive and in Romanian also from the dative) against a contact explanation. This is, of course, counter-intuitive as an assumption if we consider that areal features result from convergence between different grammars without matching grammatical categories. All mistakes in the language samples and their transliterations are too numerous to be listed here. Yet to mention a few, Catasso (ibid., 81), who is occupied mostly with the standardized varieties, states that Macedonian has retained more inflectional morphology than Bulgarian, using as an example an inflected personal name *Ivan-a* (acc.). The feature is dialectal in both languages; however, standard Bulgarian, unlike Macedonian, does mark definite masculine singular complement NPs with a suffix if they are not proper nouns (see 2.2.2). Yet it is lacking in the example of standard Bulgarian given by Catasso (ibid., 74), a frequently made mistake, probably originating from Sandfeld (1926, 12).

Despite their provocativeness, Catasso's claims nicely portray some key problems of the merger as a Balkanism and of the definition of areal features in general. In this subsection, I will assess some of the problems regarding the totality of the recipient–possessor merger, both language-internally and between the languages. I argue that a complete merger between the marking of functions, one of which is that of a verbal complement and the other adnominal, is not even a likely outcome of language change because the two functions are subject to very different syntactic conditions. The question of the cross-linguistic commonness of the merger will be addressed in **4.2.5**.

In order to address the question of the "totality" of the merger, some theoretical and terminological clarifications are needed. I consider it necessary to address the phenomenon with semantic rather than morphological terminology. As demonstrated in the previous subsection, the whole functional range associated with the dative and the genitive in the ancestor languages of Balkan Slavic, Balkan Romance, Albanian, and Modern Greek is too broad. In addition, the origins of the Greek and Albanian forms include more than two ancient cases. Therefore, it is safer to say that the merger—or the identity of expression—under scrutiny here involves, at the very least, the expression of the recipient and the adnominal possessor, neither term implying the inclusion of the entire functional scope of the earlier cases. These two functions are cross-linguistically robust and have been thoroughly studied and therefore offer an ideal starting point for an analysis, unlike the names for the cases, which are burdened by vagueness, resulting from the plethora of definitions and language-specific idiosyncrasies. In addition, this definition makes the inclusion of the analytical structures in the scrutiny painless.

In other words, at the least what the new case form, the dative–genitive or the identical analytical marking, inherited was the recipient function of the dative, usually the indirect object of a ditransitive verb, and some of the adnominal case functions of the genitive, most notably the marking of the adnominal possessor.[106] Possession as an umbrella term, however, is problematic. The discussion in **4.2.2** and **4.2.4** demonstrates the need for a more nuanced differentiation for the types of possession in the Balkan languages. In addition, Robert M.W. Dixon (2010b, 265) reminds us, for

[106] Although the paths of the grammaticalization of the merger are not discussed here, it must be remembered that the dative has been thought to have carried some possessive functions as early as in the Indo-European proto-language (see, e.g., Beekes 2011, 185). However, this is not to say that the merger would have therefore been unavoidable: There could have been several other outcomes.

instance, that "[i]t is *not* appropriate to describe nominalization [of verbal arguments] as a type of possession, but rather to note that this shows a further function of a grammatical marker used in possessive constructions." Thus, while possession and its derivatives are used here in a somewhat broader sense than would be acceptable *sensu stricto*, a more refined differentiation is made when it is crucial for the argument.

The inclusion of the analytical marking in the analyses of the merger is not generally viewed as problematic in earlier literature, including by Catasso (2011), even if the terminology used may be based on the names of the morphological cases. The Balkan Slavic standard languages and the majority of the dialects express the functions of recipient and adnominal possessor only by analytical means. While some personal pronouns have a dative form, with the rest of the personal pronouns the dative form may be either homonymous with the accusative or the datival functions must be expressed with an analytical construction.[107] The nouns, which in the standard languages are uninflected, always receive the preposition *na*:

(49) Macedonian:

i	*ja*	*dadov*	*kniga-ta*	***na***	***Slavjanka***
she.DAT.SG	it.ACC.F.SG	I.gave	book.F.SG-DEF	PREP	S.

'I gave the book to Slavjanka'

(50) Macedonian:

roditeli-te	***na***	***Slavjanka***
parents-DEF	PREP	S.

'Slavjanka's parents'

A crucial notion regarding the merger of the recipient and the possessor is that it represents the merger of two different functions, which, although not semantically maximally distant, combine a typical adnominal and a typical verb argument use of case. This fundamental syntactic difference between the two functions is often overlooked in analyses of the merger, sometimes to detrimental effect. For example, the descriptive traditions of Albanian and Romanian distinguish a dative and a genitive, despite the fact that there is only one morphological case. As argued in **3.3.2**, this confusion results from erroneously interpreting a linking article, for instance, in example **(45)**, which agrees with the head, as a case marker, although the article does not contribute to the marking of grammatical roles. The linking article does not

[107] A comprehensive analysis of the pronominal marking of the dative–genitive in Balkan Slavic is found in Cyhun 1981, 25–48.

appear with the nominal modifiers exclusively, but also with other kinds of modifiers, which clearly shows that it belongs to the general nominal syntax of Albanian and Romanian. Catasso (2011, 87–88) seems to hold this difference too, which belongs to the two different syntactic domains, against the unity of the merger in the Balkan languages.

Confusion about the unity of the recipient–possessor merger is not limited only to phenomena pertaining to the particular nature of the nominal syntax. Although on different grounds than those in the Albanian and Romanian grammar traditions, Zbigniew Gołąb (1984, 53) argues for a distinct dative and genitive in Aromanian, which lacks the linking article. For Gołąb, the mutually exclusive contexts where the recipients and possessors occur support the two-case interpretation: "I think that we can and should separate functionally and formally the Dative from the Genitive in Arumanian because in the former function the reduplication [clitic doubling] is obligatory, but in the latter it never occurs." (Ibid.) Gołąb's argument could be disputed with arguments similar to those above: The recipient is a verbal complement unlike the adnominal possessor; therefore the clitic doubling involves only the structures with recipients, and no parallelism in this respect between the marking of the recipient and the possessor can be expected.

As is shown here, the differences between the marking of the recipient and the possessor in the Balkan languages is mostly due to the inherent differences between the verbal and nominal syntax. This applies both to clitic doubling and to the use of the linking article in Albanian and Romanian. I believe that without the historical knowledge of the merged case originating from two distinct cases, such claims would not occur to anyone. Further, I do not believe that an areal feature such as the recipient–possessor merger must, as Catasso assumes, reach the same level of grammaticalization in all affected linguistic varieties in order to be credibly attributed to convergence induced by language contact. This is partly because absolute parallelism is not a probable outcome of a language contact when the borrowed feature is integrated into a part of the language structure that displays major language-specific idiosyncrasies. As is argued in **3.3.3**, the linking article in Albanian and Romanian demonstrates very well how the interaction with existing syntactic properties, such as the preferred word order, affects the outcome of the contact-induced convergence. In fact, deciding too early which areas of grammar to include in the analysis of a functionally-defined phenomenon such as the recipient–

possessor merger carries a risk. A too narrow definition might give not only a flat image of the situation, but also prevent us from seeing the extent of the similarities between the languages, as will be demonstrated in subsections **4.2.3** and **4.2.4**.

4.2.3 Analytical marking of the recipient and possessor

The ways to mark the recipient and the possessor analytically reveal further developments. Although some of these diverge from the functional merger, the manner is often similar in the different languages. The first strategy, namely, the identical analytical marking of the recipient and the adnominal possessor in the Balkans, is not limited to Balkan Slavic, whose situation is shown in examples **(49)** and **(50)**, but it is found in Balkan Romance and in some varieties of Greek too. In the second strategy, only the recipient is expressed in a similar way as in Balkan Slavic with a goal/location-indicating preposition. In the third strategy, adnominal possessors are marked with a preposition indicating source. The first and third strategies seem to be distributed in an areally significant manner.

The first strategy is used in Aromanian, which expresses both the recipients and adnominal possessors through the dative–genitive case and the preposition *a*, shown in examples **(51)** and **(52)**:

(51) Aromanian (Mišeska Tomić 2006, 180)[108]

Petri	*lji*	*deade*	*lilice*	*a*	*featilj(e)i*
P.	she/he.DAT	gave	flower	PREP	girl.DAT/GEN.F.SG.DEF

‘Petri gave a flower to the girl’

(52) Aromanian (Mišeska Tomić 2006, 182)

albul	*porc*	*a*	*ficiorlui*
white.DEF	pig	prep	boy.DAT/GEN.SG.DEF

‘the boy’s white pig’

Some Daco-Romanian dialects also have an identical analytical expression for the recipient and the possessor: A dialect spoken in Transylvania uses the preposition *la* in both functions. The use of the preposition to mark the recipient is visible in example **(55)** in standard Romanian. Similarly, dialects around Bucharest use the prepositionally functioning *lu* (from the dative–

[108] A note is in order regarding Olga Mišeska Tomić’s book *Balkan Sprachbund morpho-syntactic features* (2006). The book was met with sharp criticism, owing to a significant number of mistakes it was said to contain, see e.g., Friedman 2011. Although rarely central to the argumentation, the book is used in this study due to its unique contrastive syntactic approach. I have endeavored to make sure that the examples cited here do not contain mistakes.

genitive pronoun *lui*).[109] In addition to Aromanian, the Sarakatsani dialect of Greek spoken in Bulgaria displays a construction whereby both the adnominal possessor and the recipient are expressed with the preposition *se*, conveying goal and location.[110]

The second strategy is illustrated by standard Greek, for example, when there is a prepositional construction with *se* and the accusative used to mark the recipient along with the bare genitive case, shown in example **(39)**. Unlike the Sarakatsani dialects spoken in Bulgaria, the prepositional construction cannot be used to mark adnominal possessors:

(53) Greek:

édosa	*to*	*vivlío*	*s-tin*	*Élli*
I.gave	DEF.ACC.N.SG	book.ACC.N.SG	PREP-DEF.ACC.F.SG	Elli.ACC.F.SG

‘I gave the book to Elli’

(54) Greek:

**vivlío*	*s-tin*	*Élli*
book.ACC.N.SG	PREP-DEF.ACC.F.SG	Elli.ACC.F.SG

‘Elli’s book’

In standard Romanian, in expressions of the recipient where the dative–genitive cannot be used, usually with quantifiers, the goal–location preposition *la* with accusative is used instead.[111] Unlike Greek, this analytically-expressed recipient can nevertheless be doubled on some occasions. The same construction cannot be used adnominally in expressing possession.

(55) Romanian (Lombard 1973, 239):

am	*dat-o*	*la*	*toată*	*lumea*
AUX.1SG	given-it.ACC.F.SG	PREP	all-NOM/ACC.F.SG	people.DEF.NOM/ACC.F.SG

‘I gave it to all the people’

In Romanian, the adnominal possessor is also sometimes expressed by an analytical construction, but differently from the recipient. Standard Romanian resorts to the preposition *a* (a reflex of the Latin *ad*, expressing goal) instead of the dative–genitive. The prepositional structure with *a* is used to express possession, for instance, with quantifiers:

[109] Andrei Dumitrescu, personal communication.
[110] Dionysios Mertyris, personal communication.
[111] Similar to this standard Romanian construction, all recipients in Megleno-Romanian are expressed analytically with the cognate preposition *la* (Mišeska Tomić 2006, 165).

(56) Romanian:

tatăl	*a*	*trei*	*băieți*
father.DEF	PREP	three	boys-NOM/ACC.M.PL

'the father of three sons'

The third strategy is illustrated by the Greek dialects that express the recipient with the accusative instead of the genitive. These dialects often employ a construction with the source-indicating preposition *apó* plus the accusative to show adnominal possession (Asenova 2002, 93). In these dialects, the use of the genitive, at least in the plural, is often very limited.[112] Megleno-Romanian, on the other hand, employs the preposition *al* with definite human possessors, but uses a source-indicating preposition *di* with all other possessors with the exception of the definite non-human animate possessors, which vacillate between the two markings (Mišeska Tomić 2006, 165–166):

(57) Megleno-Romanian (Mišeska Tomić 2006, 166)

ună	*cartă*	*di*	*un*	*profesor*
a	book	PREP	a	professor

'a book of a professor'

Possessors that do not have articles—and therefore do not, in most instances, distinguish between the nominative–accusative and dative–genitive—are expressed in Romanian in a similar way as in example **(57)** with the preposition *de*, a cognate of the Aromanian and Megleno-Romanian *di*. In Aromanian, all inanimate possessors are expressed with *di*, and so are indefinite animates. (Mišeska Tomić 2006, 150, 183.) However, it has been reported that in at least one Aromanian dialect, definite animate possessors can also be expressed with the preposition *di* (Markoviḱ 2007, 94).

Interestingly, western and southwestern varieties of Balkan Slavic show an expansion of the source-indicating preposition *od* in the marking of adnominal possession. For example, kinship relationships may be expressed with the preposition *od* (Mišeska Tomić 2006, 83):

(58) Macedonian (Mišeska Tomić 2006, 83)

Ja	*vidov*	*majka*	*í*	*od*	*Jana*
I	saw	mother	she.DAT.SG	PREP	Jana

'I saw Jana's mother'

[112] Dionysios Mertyris, personal communication.

With the analytical marking, three diverging strategies emerge: The first consists of an identical marking of both functions, employing a preposition with a goal meaning, which is the situation in most varieties of Balkan Slavic, the Bulgarian Sarakatsani dialect of Greek, and Aromanian. Both the non-Slavic varieties have recent and ongoing intense contacts with Slavic. The second strategy uses an analytical marking for the recipient with a preposition indicating goal, which competes with the bare dative–genitive. This takes place in both Romanian and Greek, clearly without any mutual areal link. In Romanian, however, this strategy is limited to constructions that do not distinguish the dative–genitive case from the nominative. The third strategy consists of marking adnominal possessors with a preposition indicating source, as in southern and southwestern Balkan Slavic dialects, Megleno-Romanian, Aromanian, Romanian, and the northern dialects of Greek, and typical of all modern Romance and most Germanic languages as well.

The first and third of these tendencies show a clear areal distribution: The identity in the analytical marking of the recipient and the possessor is limited to those varieties of Greek and Balkan Romance (with the exception of the two Daco-Romanian dialects) that have experienced and continue to experience relatively recent intensive contacts with Balkan Slavic. The more grammaticalized use of the source-indicating preposition as a marker of possession is concentrated in the geographical area of Macedonia. In addition, what these examples of analytical marking of the recipient and the adnominal possessor show is that a high degree of the loss of case inflection is, quite understandably, a major reason for the exclusive or increased use of prepositional constructions. Albanian, which most consistently preserves the case distinction between the nominative and the dative–genitive, does not resort to analytical marking of recipients and possessors.

4.2.4 Split-conditioned double marking of possession and the juxtapositional pseudo-partitive

It was shown in the previous subsection that there are analytical structures, departing from the dative–genitive merger, that mark adnominal possession. The distribution of the construction with a preposition indicating source is more limited, but still areally significant, implying a contact-induced origin. Similar deviations from the use of the dative–genitive case to mark adnominally used nouns are found in other subfields of adnominally used nouns. Here, we will look at a double marking strategy of possession, a

discussion that must be begun with an overview of the clitic marking of possession and a juxtapositional pseudo-partitive construction.

Frequently, an element used to express the agreement of verbs with their arguments is also employed to indicate the agreement of nouns with their possessors (Anderson 1985, 188). In Balkan languages, clitics often double the direct and indirect objects, as shown in example **(49)** in Macedonian:

i	***ja***	*dadov*	*kniga-ta*	*na*	*Slavjanka*
she.DAT.SG	it.ACC.F.SG	I.gave	book.F.SG-DEF	PREP	S.

'I gave the book to Slavjanka'

With regard to the doubling of the indirect objects, the conditions differ somewhat, but examples **(39)**, **(41)**, **(44)**, and **(47)** all show contexts in which the doubling may occur, although it is not necessarily obligatory in all languages (for more details on clitic doubling, see **3.3.2**). This subsection will demonstrate that the use of the dative–genitive pronominal clitic along with other strategies employing clitics reveals significant uniformities among the Balkan languages, namely, a double-marking strategy of possession and a split conditioning its use. These possessive clitics are also discussed in detail by Roumyana Pancheva (2004), but from a standpoint that is not directly relevant here.[113]

In the Balkan languages, the dative–genitive pronominal clitic can also be used to express adnominal possession in certain contexts. The same element, used in the doubling of the indirect objects, appears as the

[113] Pancheva's paper is a welcome modern account of the dative–genitive clitics in Albanian, Balkan Slavic, BCMS (Bosnian, Croatian, Montenegrin, Serbian), and Modern Greek, because of its attention to the differences between clause-level and NP-internal syntax. Yet there are problems regarding some parts of her analysis: Although mostly preoccupied with formal syntax and not making claims about the areality of the phenomenon, Pancheva concludes that the Greek clitic is best described as valued for the "genitive case," whereas the Balkan Slavic and Romanian clitics display "dative case features" in syntax. However, one of her key arguments to support the genitive interpretation is that, in Greek, "possessor raising," i.e., expressing the possessor on the clause level, is not possible (Pancheva 2004, 188–189). She demonstrates this with an example in Greek, which she deems ungrammatical. Yet she places the Greek phrase parallel to a grammatical phrase in Macedonian: *si mu*[dat.cl.] *gi zel*[predicate] *parite*[possessee] 'you [/seem to] have taken his money'. The phrase can be translated into Greek without a problem, using the external possessor construction: *tu*[gen.cl.] *píres*[predicate] *ta leftá*[possessee], although this construction adds an emphasis on the possessor (Maria Basdekis, personal communication). Pancheva's confusion might stem from the fact that, unlike Balkan Slavic, constructions of external possession (or possessor raising) are often limited to certain kinds of possessors. In Europe, such constructions are usually restricted to those with possessees that are direct objects or unaccusative subjects (Haspelmath 1999, 110), which seems to be the case in Greek as well (Mišeska Tomić 2006, 218). For more discussion on external possession, see **4.2.5**.

pronominal possessor in Greek, in example **(60)**, standard Bulgarian, in example **(59)**, and many Balkan Slavic dialects.

(59) Bulgarian:

kola-ta	***i***
car-DEF	she.DAT.SG

'her car'

(60) Greek:

i	*gonís*	***tis***
DEF.NOM.M.PL	parents	she.GEN.SG

'her parents'

However, in standard Macedonian and in several Balkan Slavic dialects, this strategy of marking the possessor is limited to certain kinship terms, as in example **(15)**. The unmarked pronominal possession in these varieties is expressed with possessive pronouns. However, Olga Mišeska Tomić (2009, 118–119) points out that, instead of adnominal clitics, Macedonian resorts more often than Bulgarian to an external possessor structure, whereby the dative clitic is used on the clause level, a strategy discussed further in **4.2.5**.

Romanian employs several strategies to mark pronominal possession, including a post-positioned possessive pronoun. A construction with an enclitic dative pronoun is also possible with a definite form of the noun. However, this use is considered literary, if not archaic, and, in the spoken languages, it is limited to some fixed expressions. Yet while in first and second person a possessive pronoun is used, with the third person the full dative–genitive personal pronoun is possible:

(61) Romanian

părinții	***ei***
parents.DEF.NOM/ACC	she.DAT/GEN.SG

'her parents'

However, with kinship terms, a clitic form of the possessive pronoun can be used, although without the intervening definite article (Mišeska Tomić 2006, 145):

(62) Romanian

maică-sa
mother-his/her
'his/her mother'

In Albanian, unmarked pronominal possession is expressed with a possessive pronoun:

(63) Albanian

prindërit	*e*	*tij*
parents.DEF.NOM.M.PL	LINK.NOM.M.PL	his

'his parents'

Here, as in example **(45)**, a linking article, agreeing with the head, intervenes between the possessee and the postposed possessor.[114] While not used with the word "parents," with some kinship terms this linking article can be used preposed, in third person, to indicate possession (Hetzer 1978, 172):

(64) Albanian

e	*ëma*
LINK.NOM.F.SG	mother.NOM.F.SG.DEF

'his/her mother'

It emerges from this analysis that Greek, Balkan Slavic, and, to a limited degree, Balkan Romance employ a pronominal possession-marking strategy with a dative–genitive clitic attached to the possessee, that is, the head of the NP. Nevertheless, Albanian and Balkan Romance have a frequent strategy to mark possession through a clitic element, although not a personal pronoun, attached to the head. The modern Romanian construction uses the clitic possessive pronoun, while Albanian employs the linking article. In syntactic terms, the clitics in examples **(59)**, **(60)**, **(62)**, and **(64)** must be analyzed as possessors in the absence of another explicit possessor within the same clause.

Since the possessive clitic is phonetically attached to the possessee, that is, the head, it could be asked whether these constructions have some other common characteristics with the head marking of possession, which is a cross-linguistically common strategy to indicate adnominal possession (e.g., through possessive suffixes, typical of the Uralic and Turkic languages), although mostly absent from Indo-European languages. Consider again example **(15)**:

sin	*mu*	*Stojan-u*	*pošol*	*vojnik*
son	he.DAT.M.SG	Stojan-DAT.M.SG	started.out.as	soldier

'Stojan's son became a soldier'

[114] Excluding the first- and second-person possessors when the possessee is in the singular. These structures omit the linking article.

Here, the possessee is marked with the dative clitic, but the possessor in the dative case is also present. Both in the Romanian example **(62)** and the Albanian example **(64)**, a possessor marked with the genitive–dative can be added. In Greek, however, such marking seems to be somewhat more limited (see Mišeska Tomić 2006, 118). Because these constructions involve both the marking of the dependent and the head, they must, in fact, be called a double marking of possession.[115]

A further characteristic of these constructions in Balkan Romance, Albanian, and at least some varieties of Balkan Slavic is to limit the strategy (or at least one of the strategies) of clitic head marking of possession to kinship terms. In addition, Bulgarian, where all kinds of possessees may take the clitic marking, has a differential marking for kinship terms, as the words "father" and "mother" cannot take the definite article. With some of the other kinship terms, there is vacillation in the use of the definite article (Mišeska Tomić 2006, 101).[116] In typological terminology, such structures with differential marking constitute split systems of possession, which are amply attested cross-linguistically. It has been claimed that they are less common, however, in Eurasia (Nichols & Bickel 2008), yet similar splits have been discovered in the Celtic languages, Portuguese, Italian, Basque, and Turkish (see e.g. Stolz et al. 2008).

The splits often involve one strategy for marking alienable possession and another for inalienable possession. Kinship relations pertain to inalienable possession, that is, irrevocable forms of possessive relationships or belonging, often including body parts. Generally, as the Balkan languages show, it is the inalienable possession that is head-marked, while alienable possession is dependent-marked (Nichols 1992, 117). However, what is noteworthy in the case of the Balkans, with the exception of Bulgarian, is that the split seems to be similarly conditioned and realized in a similar

[115] For the sake of argument, it could be said that Gołąb's insistence on "object doubling" of the possessors is fulfilled in constructions in which the head of a NP is marked for possession, similar to when the predicate, that is, the clausal head, receives the indirect object marking. There is also another logical option: instead of head marking, a parallel to the doubling of the direct objects could be claimed when the possessor occurs on the clausal level, i.e., in constructions of external possession, if, at the same time, the adnominal possessor remains overtly expressed. Nevertheless, this is obviously splitting hairs and demonstrates yet again the unfeasibility of demanding a perfect parallel between two functions that operate on different domains of the syntax.

[116] However, not necessarily on semantic grounds: In the standard language, the word *brat* 'brother' is used without the definite article, while *sin* 'son' takes the article, most likely because if an enclitic element, such as a possessive clitic or a definite article, is added, the stress falls on that element. The resulting stress pattern *sin mí* would not be grammatical, whereas *sinǎt mi* is acceptable.

structural manner. For example, in Italian the split involves the omission of definite articles with kinship terms; in Portuguese, the split involves only predicative possession (Stolz et al. 2008, 471–472).

Moving further away from possession proper, we find another adnominal construction, one intimately related to adnominal possession, but diverging in its realization from the dative–genitive, the so-called pseudo-partitive constructions, which are NPs involving two nouns, one of which acts as a quantifier:

(65) Romanian

un	*pahar*	*de*	*vin*
a	glass.NOM/ACC.SG	PREP	wine.NOM/ACC.SG

‘a glass of wine’

The pseudo-partitive construction, shown in Romanian in example **(65)**, differs from the adnominal partitive construction—for example, *a glass of that wine*—in that the quantified noun expresses only a type of entity, quantified by the other noun, not a definite, specific, or presupposed entity like the partitive construction (Koptjevskaja-Tamm 2009, 329–330). While the adnominal partitive construction is usually expressed in the Balkan languages with prepositions indicating source, Albanian, Balkan Slavic, and Greek employ a juxtapositional construction (note that while the Albanian and Greek nouns are glossed as nominatives, the accusative and the nominative are indistinguishable in these examples):[117]

(66) Albanian

një	*gotë*	*verë*
a	glass.NOM.SG	wine.NOM.SG

‘a glass of wine’

(67) Bulgarian/Macedonian

edna	*čaša*	*vino*
a/one	glass	wine

‘a/one glass of wine’

[117] Flavia Pompeo (2012), who, probably for the first time in the literature, acknowledges this feature as a potential Balkanism, does not recognize the same structure in Albanian. Instead, she quotes an example of a plural construction *një grup punëtorësh*[abl.] ‘a group of workers’ (ibid., 536). Indeed, it seems that with plural nouns, Albanian employs the ablative, whereas in the singular it conforms to the same strategy as Greek and Balkan Slavic.

(68) Greek

éna	*potíri*	*krasí*
a	glass.NOM.SG	wine.NOM.SG

'a glass of wine'

4.2.5 Areal and typological limits of the merger

Here, I attempt to establish whether the recipient–possessor merger makes the Balkan languages stand out from their immediate neighbors and thereby set the areal limits of the merger. In the absence of comprehensive typological research into the occurrence of the merger cross-linguistically, I will examine similar developments in other European languages. I argue that despite not being rare or typologically uncommon, the recipient–possessor merger is unlikely to have resulted in the Balkans by chance, but must be seen a contact-induced development. Nevertheless, the examples from Hungarian and Turkish reveal that both the double marking of possession, discussed in **4.2.4**, and the marking of pseudo-partitives through juxtaposition should be viewed in a wider areal context.

Overlap between some of the functions of the "dative" and the "genitive" is attested in several languages of the world (Næss 2008, 578). Moreover, case systems with one case dedicated to mark both the recipient and the adnominal possessor are not unknown; outside the Balkans, these languages include, for example, Armenian and some Australian languages (ibid.). However, sometimes the frequency of such a merger in Europe is exaggerated, based on a misunderstanding. Consider the well-known example:

(69) German (Haspelmath 1999b, 109)

Die	*Mutter*	*wusch*	*dem*	*Kind*	*die*	*Haare*
DEF	mother	washed	DEF.DAT	child.DAT	def.ACC	hair.ACC

'the mother washed the child's hair'

The German dative indeed seems to correspond in translation to an adnominal possessor, marked with the English genitive clitic. However, the German example represents a structure called dative external possession, which, according to Martin Haspelmath (1999b, 109), constitutes a feature of SAE. The dative encodes a possessor, which must be a mentally-affected participant in the described event where the possessee is a direct object, a locative argument, or an unaccusative subject (ibid., 110–111). These constructions, characteristic of the Balkan languages as well, must be distinguished, however, from adnominal possession: The dative is

essentially a verbal complement and thus pertains to the clause level. Therefore, the languages which have this type of external possessor, but no adnominal use of the dative do not allow sentences such as **die Haare dem Kind ist schmutzig* (Dürscheid 2007, 104).[118]

Case systems which display the merger of a dative-like and a genitive-like case—beyond the external possessor constructions—are rare in Europe. Yet a striking parallel with the three-member Balkan case system is found in Scandinavia. The only major variety of Scandinavian or North Germanic that preserves case inflection is Icelandic, which distinguishes four cases: nominative, accusative, genitive, and dative. However, Elfdalian (autonym *övdalska*, Swedish *älvdalska*), a North Germanic linguistic variety spoken in Sweden, yet lacks mutual intelligibility with Swedish, has almost completely lost the genitive, although its nouns preserve a distinction between the nominative, accusative, and dative. However, nominative–accusative and somewhat less often accusative–dative syncretism are common in several inflectional paradigms, the definite and plural forms preserving the most distinctions (Dahl & Koptjevskaja-Tamm 2006, 63). In Elfdalian, the case that marks the recipients, the dative, is the unmarked way to express adnominal possession (ibid., 66):[119]

(70) Elfdalian (Levander 1909, 97)

fjǫsbuðę *stumas-ųm*
cowshed Stormas-DAT.PL
‘cowshed of the Stormas-people’

Remarkably from a Balkan perspective, Elfdalian has, a postposed definite article, as do all Scandinavian languages, although in Elfdalian the article is attached to the noun unlike the Balkan languages where it is enclitic, with the exception of Greek.[120] Elfdalian fuses the number, the definite article, and the case into one suffix morpheme, as do Albanian, Romanian, and earlier Balkan Slavic. This undoubtedly reflects similar

[118] However, note that the predicative use of the dative may be acceptable in spoken German; see Heine & Kuteva 2002, 105. For a different view of the acceptability of the adnominal dative in spoken German, see Haspelmath & König 1997, 586–587.

[119] Yet goal and location are expressed with a preposition and the accusative for goal and the genitive for location are different from the Balkan languages.

[120] Another analogous development pertaining to the place of pronominal determiners in the NP and connecting Scandinavia and the Balkans is the sometimes postposed possessive pronoun. While almost absent in standard Swedish, this word order occurs in several other Scandinavian varieties typical of pronominal possession of kinship words. It is likely that, both in Scandinavia and in the Balkans, the postpositioned possessive pronoun is a relic of a period that gave rise to the definite postposed article, which developed from a postposed demonstrative pronoun.

pathways of grammaticalization: The inflected demonstrative pronoun is the source for the new fused morpheme. Elfdalian is also particularly interesting from the point of view of the uniqueness of the Balkan system. On the one hand, it proves that Indo-European languages can develop three-case systems outside of the Balkans; on the other hand, as Östen Dahl and Maria Koptjevskaja-Tamm (ibid., 60) note, not only are bound definiteness markers cross-linguistically rare, but languages with both cases and bound definiteness markers are even rarer.[121]

The major languages adjacent to the Balkan languages under scrutiny here include BCMS, Hungarian, and Turkish. As stated before, the distinction between the genitive and the dative is preserved in most varieties of BCMS; where the distinction has not been preserved is in dialects closest to Bulgarian and Macedonian on the South Slavonic dialectal continuum. In standard Turkish, the dative marks both the recipient and the goal of movement. The possessive construction involves a dedicated genitive case, but not in all situations: The definite possessor is marked with the genitive case and the possessee, which follows the possessor, with a possessive suffix agreeing with the possessor. The indefinite possessor, however, is in the non-marked nominative case, and the possessee gets the possessive suffix. (Golstein 1999, 57–60.) In Turkish too there is a pseudo-partitive construction, expressed through mere juxtaposition:

(71) Turkish (Golstein 1999, 60)

iki	*yudum*	*su*
two	gulp.NOM.SG	water.NOM.SG

‘two gulps of water’

The Turkish varieties spoken in the Balkans have several characteristics attributed to contact with the other Balkan languages, and some of them involve the nominal morphosyntax. These phenomena include changes in the preferred word order, such as a change from SOV to SVO and from possessor–possessee to possessee–possessor order, and the merger of goal and location, generalizing the locative case in this role (Friedman 2006, 35–38). However, no changes involving the marking of recipients and possessors have been attested.

[121] The source for the latter generalization is, unfortunately, unpublished, but it is credited to Matthew Dryer, who reports only five or six languages with both cases and bound definiteness markers in a sample of 700 languages. Yet the status of the Balkan postposed definite article as a “bound” marker is not completely clear, since it often displays the characteristics both of a clitic and of a bound morpheme.

Hungarian is known to display homonymous marking for the recipients and possessors by using the dative case. However, the unmarked possessive construction involves the nominative and the possessive suffix attached to the possessee, which follows the possessor, while the dative is used only in certain contexts, these contexts and those requiring the nominative being mutually exclusive. The possessor is marked with the dative if it is a possessee of another possessor, if it has a demonstrative, indefinite, or interrogative pronoun as a determiner, or if the possessee does not immediately follow the possessor (Csepregi 1991, 115).[122] Hungarian too has a juxtapositional pseudo-partitive construction:

(72) Hungarian

öt	*pohár*	*víz*
five	glass.NOM.SG	water.NOM.SG

'five glasses of water'

Hungarian does show limited identity between the marking of the recipient and the adnominal possessor. Yet in other respects, the Hungarian expression of possession is very different from that of the Balkan languages, relying mostly on a head-marking strategy. Nevertheless, in both Hungarian and Turkish, the double marking in expressions of adnominal possessors is common, as with the Turkish definite possessors and Hungarian possessors marked with the dative. Both languages also display a split in the marking of adnominal possession; in Turkish this is based on definiteness, and in Hungarian it depends on the presence of pronominal determiners. As shown in **4.2.4**, although normally limited to kinship relations, the double marking is also common in the Balkan languages, bringing them closer to their non-Indo-European neighbors and distancing them from other European languages, where such constructions do not exist. In addition, both languages display a pseudo-partitive construction. Note, too, that in examples **(71)** and **(72),** both nouns, regardless of whether they are quantified by a noun or a numeral, receive no case or plural endings.

[122] In constructions of predicative possession, the possessor is likewise encoded with the dative and, because Hungarian lacks a *habeo*-construction, the predicate is the verb "to be." This might suggest that the grammaticalization of the dative into the marking of the adnominal possessor spread through the predicative possession: Since the verb "to be" is omitted in the third person in non-possessive predicative expressions, this renders both possessive constructions almost homonymous, because the possessee would be marked with the possessive suffix in any case. However, with the adnominal possessor marked with the dative, the possessee must always be preceded by the definite article.

To sum up, what Elfdalian proves is that a three-case system with one case dedicated to expressing both the recipient and the adnominal possessor exists outside of the Balkans, found in a language with a common Indo-European source, but with no later areal connection to the Balkans. A credible hypothesis of the diffusion of a feature through language contact, unlike Catasso (2011, 87–88) suggests, does not call for the typological rarity of the feature (see Bickel & Nichols 2006, 2, and the discussion in **4.1.2**). Similar systems, if not quite as strikingly analogous, are no doubt to be found elsewhere. Yet crucially, although systematic study on the subject is lacking, these systems cannot be so common that the odds for a clustering, such as in the Balkan languages, were particularly high. If the merger is found, let's say, in 25 percent of the languages of the world (which would be, of course, a gross exaggeration), then the likelihood of the feature being present by chance in four adjacent, but otherwise independent, linguistic varieties is rather low, although such a frequency could be called "typologically not uncommon." What the feature present in adjacent varieties then implies is an areal explanation: Either the feature was inherited from a common ancestor language, known or unknown, or the feature has been diffused through contact. Despite some overlap among the functions of the Indo-European genitive and dative, a common inheritance does not explain the merger.

However, while such a close parallel in Scandinavian somewhat diminishes the uniqueness of the merger of the two functions, the relatively tightly-bound fused case and definiteness markers in Albanian, Romanian, and earlier Balkan Slavic seem to make the Balkan case system more exceptional as a whole. The Hungarian use of the dative in constructions of adnominal possession seems to challenge the neat clustering of the feature in the Balkans. Nevertheless, the Hungarian use of the dative is rather limited and significantly less developed than in the Balkan languages, which show tendencies similar to Hungarian and Turkish in the double marking of possession and the pseudo-partitive construction. These similarities are striking indeed, and although any speculations about their contact-induced origins are, no doubt, preliminary, the features do show an areal clustering in the European context.

4.2.6 Conclusions

The identical marking of the recipient and the adnominal possessor in the Balkan languages, known in many accounts as the dative–genitive merger, is

a complex phenomenon. Closer scrutiny reveals several related subphenomena, which can all, arguably, be evaluated against an areal hypothesis, that is, the assumption that language contact has contributed to the similarity of the expression of these structures among the languages involved. Based on the analysis here, two subphenomena emerge, both of which can be categorized under the single heading of the recipient–possessor merger:

1. In all languages studied here (Albanian, Balkan Romance, Balkan Slavic, Greek), there is or there has been a single morphological case that expresses both the recipient of a ditransitive verb and an adnominal possessor. The sources of the new case and the paths of its grammaticalization vary from language to language to the extent that "merger" of the genitive and the dative becomes somewhat misleading as a term. The major exceptions to this among the linguistic varieties studied here are the northern Greek dialects, where the recipient function has been overtaken by the accusative case, and those varieties that have lost their case inflection entirely.
2. The varieties that have lost their case inflection altogether have an analytical marking for the recipient and the adnominal possessor. Balkan Slavic, Aromanian, and some dialects of Romanian and Greek employ a construction with a goal/location preposition to express both functions.

However, the analysis reveals three additional phenomena that, while not classifiable as part of the recipient–possessor merger, belong inseparably to the adnominal use of nouns in the Balkans and therefore cannot be left out of a comprehensive analysis:

3. Examining the marking of adnominal possession reveals another phenomenon, areally limited to the geographical region of Macedonia, namely, the expression of adnominal possession with a preposition indicating source. This phenomenon pertains to the south and southwestern Balkan Slavic dialects, Megleno-Romanian, Aromanian, Romanian, and the northern dialects of Greek. However, among these varieties the most remote geographically, Romanian, shows the most limited use of this construction.
4. In the expression of adnominal possession, all of these languages have a double-marking strategy: they can mark both the possessor—that is, the dependent—and the possessee, the head of the noun phrase, with a clitic. The head marking is more widespread in Bulgarian and Greek. In other varieties, at least one of the head-marking strategies is limited to

inalienable possession, and more specifically, is used to mark the nouns expressing close kinship. While not uncommon in Europe, this particular split in possession marking is areally limited to the Balkans, including when compared with Hungarian and Turkish, which, although they resort to head marking, display a split based on different criteria.

5. There is a juxtapositional pseudo-partitive construction in Albanian, Balkan Slavic, and Greek. A similar structure is found in Hungarian and Turkish, where the juxtapositional strategy nevertheless pertains to the general treatment of quantified nouns. In the Balkan languages, numerals above one receive a plural ending (in Balkan Slavic, with masculine nouns, a historical nominative dual ending, optional in Macedonian), whereas all quantified nouns in Hungarian and Turkish employ the nominative singular. In Europe, a similar strategy is found only in some Germanic languages, while the majority of European languages, including other Slavic tongues and languages with no case inflection like English, employ a strategy with an explicit marker (Koptjevskaja-Tamm 2009, 335).

My main goal has been to analyze some of the divergent and convergent fine-grained phenomena related to a classical Balkan linguistic area feature commonly thought to be robust and relatively unproblematic. Although the areality of the head marking of possession, for instance, requires more study, I believe that the overall areal hypothesis regarding the recipient–possessor merger has been solidified by showing that the deviations from the "completeness" of the merger in fact involve other areally significant phenomena. Nevertheless, the functional definitions of the features show intriguing distributions, as the double marking of possession and juxtapositional pseudo-partitive constructions also appear in Hungarian and Turkish. Here, it is crucial that the morphosyntactic manifestation of the features, their conditioning criteria, and any analogous structures in the language are taken into account, since these usually show that the languages of the Balkan linguistic area stand out as particular among other languages displaying similar phenomena. In this sense, one cannot entirely agree with Howard Aronson's (2007) view, discussed in **3.1.1**, that functional criteria must be given center stage in the study of the Balkan linguistic area: The "intertranslatability" of Balkan languages is, for instance, not based merely on functional similarities, but also on their similar morphosyntactic realizations.

4.3 The fall of case inflection, the rise of the definite article: Evidence for a universal tendency?

The case loss in Balkan Slavic is accompanied by another change in the marking of the NP, namely, the rise of the definite article.[123] Since both phenomena involve nominal syntax and took place at the same time, it is sometimes claimed that the emergence of the definite article and the case loss are connected. Through a statistical analysis of typological data, this section seeks to establish whether there is a cross-linguistically observed connection between case and definite article and what the possible mechanisms are that may operate between the two. I will argue that there is possibly an inverse correlation between the number of cases and the occurrence of the definite article, thus indicating a universal tendency, which, nevertheless, does not imply direct causation. I propose that the domain where the mechanisms of this interconnection may operate is found in the marking of information structure. Yet, since in the best scenario the definite article can be explained only partly by the loss of case inflection, this possible universal tendency does not challenge the diffusion of the definite article through language contacts. In addition, in introducing the concept of universal areality, Balthasar Bickel (2007, 243) notes that hardly any typological variable is evenly distributed throughout the world, which suggests that in any case a universal tendency is unlikely to be independent of areal factors.

4.3.1 The definite article in Europe

Gennadij Tiraspol'ski (1980, 73–74) argues that the introduction of the definite article, inflected for case, was directly responsible for the demise of case in Balkan Slavic by marginalizing the inflected case forms. On the other hand, it has been proposed that the loss of case inflection in Western European languages may also be connected to the rise of the definite article, but through an opposite mechanism, because it is the loss of case inflection that supposedly brought about the grammaticalization of the definite article (for a critical discussion of this view, see Anward & Swedenmark 1997, 32–33; Barðdahl 2001, 192–193). In light of the conclusion in **Chapter 2**, Tiraspol'ski's theory is unconvincing, since, for example, the attested definite articles, inflected for case, always combine with the inflected case forms of the noun. Yet as shown in **3.3**, case marking is closely connected to

[123] For the rise of the definite article in Balkan Slavic, see Kurz 1938; 1946, Gălăbov 1951; 1952, Svane 1961; 1962, and Mladenova 2007.

definiteness in the Balkan languages, because definite NPs are more likely to express more case distinctions than nouns, and clitic doubling typically occurs with definite objects, although definiteness may not be the only criterion for the doubling. According to Christa König (2011, 514–515), a universal interrelationship between case and the expression of definiteness is obvious, since, for instance, definiteness may serve as a criterion for a split in case marking, as in the Balkans. In addition, König gives examples in which the marker of definiteness has developed into a nominative or ergative case marker.

In addition to Balkan Slavic, in Balkan Romance too the emergence of the definite article and the case loss coincide chronologically (Dimitrescu et al. 1978, 226; see also **3.2.2**). In Greek, the definite article is not yet attested in Mycenaean Greek (16th–12th centuries BCE), but some signs of its emergence may be seen in Homeric Greek (8th century BCE) until it is observed fully grammaticalized in the Classical period (5th–4th centuries BCE) (Manolessou & Horrocks 2007, 228–229). By the time the definite article had grammaticalized, the number of cases in Greek had been reduced to four (see **3.2.3**). With Albanian, we cannot trace the emergence of the definite article with any precision, because the article was already fully grammaticalized in the first written texts (see **3.2.1**). In addition to these observations from the Balkans, a similar development has taken place in a large number of Indo-European languages spoken in Europe: With the exception of Slavic (excluding Bulgarian and Macedonian) and the Baltic languages,[124] the case inflection has either been lost or significantly reduced,[125] and a definite article has meanwhile emerged.

The emergence of the definite article in Europe has also been explained independently. As in the Balkans, in the rest of Europe as well the rise of the definite article is thought to have resulted from diffusion through language contacts (Heine & Kuteva 2006, 106–119).[126] Therefore, it may be perfectly reasonable to assume that the introduction of the definite article into these languages and the simultaneous loss of case inflection were mere coincidence. A further question is whether the diffusion of the "European" definite article is best characterized as the spread of a mere functional

[124] In Latvian, the adjectives retain an archaic definite form, whose characteristics are discussed in **2.2.1**.
[125] The peripheral Icelandic and Faroese are a counter-example; see Barðdahl 2001, 192–193.
[126] Note that Heine and Kuteva (2006, 106) reiterate the old misconception that there is a definite article in some North Russian dialects. See Kasatkina (2007) for the so-called pseudo-article in these dialects.

category or whether the diffusion was accompanied by the morphosyntactic realization of the marking, that is, the preposed article, grammaticalized from a demonstrative. Scandinavian may be indicative of the latter possibility. All Scandinavian languages have a postposed definite article, already present in Old Norse, which displayed four, often syncretic, cases in contrast to the Proto-Germanic five (+ the vocative). In addition to the postposed article, a preposed article emerged later, in complementary distribution with the postposed article in Danish, contributing to double marking in Swedish and Norwegian and being facultative in Icelandic (see e.g. Anward & Swedenmark 1997, 25–26). This geographical continuum of the level of grammaticalization suggests that the preposed article was most likely the "European" definite article, but the postposed article, unparalleled in the neighboring languages of the region, was an earlier internal development.

There are two languages spoken in Europe with more than four cases that display the definite article, namely, Basque and Hungarian. While Hungarian employs a preposed, uninflected definite article, Basque has a postposed article. Yet their case systems are not characterized by inflection, unlike that of the other European languages, including Finnic. In fact, the Basque case desinences attach to the last component of the NP, regardless of whether it is a noun, an adjective, or a determiner, such as the definite article. In Hungarian, only nouns may have case endings, and there is no case agreement of any kind in the NP, one of the facts that encouraged Andrew Spencer (2008) to characterize Hungarian as not having case at all, but rather a set of "fused postpositional portmanteaus." Whether these morphosyntactic characteristics which distinguish the two languages in the European context have any significance regarding the emergence of the definite article is, of course, hard to say. Nevertheless, the Baltic and most Slavic languages, which typically display no fewer than six cases, resist the definite article, despite being mostly situated amidst European languages that use the definite article. According to Heine and Kuteva (2006, 106–119), there is, nevertheless, an incipient-stage definite article in West and South Slavic languages other than Macedonian and Bulgarian. They base

their argument mostly on the use of adnominal demonstratives expressing anaphoric reference.[127]

In the following subsection, I undertake to test whether there is typological support to posit that the co-occurrence of case loss and the rise of the definite article in the Balkans is not coincidental. Formulated as a hypothesis, my proposal is that the fewer cases there are in a language, the likelier it is that there is a definite article. To test the hypothesis, I perform a statistical analysis of typological data drawn from a typological database.

4.3.2 Statistical analysis of the data

The material used in the statistical analysis is taken from the typological database *World Atlas of Language Structure* (Haspelmath et al. [ed.] 2013, abbreviated as WALS). For the marking of definiteness, I use Matthew Dryer's (2013) sample, which encompasses 620 languages. Dryer uses a five-part classification:

1) a definite word distinct from the demonstrative (216 languages)
2) a demonstrative word used as a marker of definiteness (69 languages)
3) a definite affix on nouns (92 languages)
4) no definite article, but an indefinite article (45 languages)
5) neither a definite nor an indefinite article (198 languages).

I use a binary classification into languages with a definite article, comprising classes 1–3, and languages with no definite article, classes 4 and 5.

For the number of cases, Oliver Iggesen's (2013) sample of 261 languages is used. Iggesen classifies the languages into nine categories:

a) no morphological case marking (100 languages)
b) 2-case categories (23 languages)
c) 3-case categories (9 languages)
d) 4-case categories (9 languages)
e) 5-case categories (12 languages)
f) 6- to 7-case categories (37 languages)
g) 8- to 9-case categories (23 languages)
h) 10 or more case categories (24 languages)
i) exclusively borderline morphological case marking (24 languages).

[127] Note, however, that the authors give the false impression that this use is not possible in East Slavic languages. In addition, their analysis of Serbian is based on one dubiously-elicited example. The anaphoric use of the adnominal demonstratives is a highly discourse- and genre-sensitive phenomenon, which renders it a very complicated research subject. For example, a language with a grammaticalized definite article employs the article in several contexts where other languages use demonstratives; see, e.g., Juvonen 2000, 21. Yet, crucially, none of the incipient-stage languages in Heine & Kuteva's analysis display obligatory markings of definite NPs.

Iggesen bases his division on Blake's (2001, 1) definition, discussed in **1.3.2**:

> Case is a system of marking dependent nouns for the type of relationship they bear to their heads. Traditionally the term refers to inflectional marking, and, typically, case marks the relationship of a noun to a verb at the clause level or of a noun to a preposition, postposition or another noun at the phrase level.

Two issues arising from this definition are potentially problematic for testing the hypothesis presented in **Subsection 4.3.1**. First, category b), languages with two cases, includes languages like English and Swedish, whose possessive endings fall into Blake's definition of case. Unlike Romanian, for example, which is included in the same category, these languages can be considered to have lost their case inflection completely, since the possessive ending attaches to the last constituent of a NP, which is not necessarily the NP's head.[128] Category i), exclusively borderline morphological case marking, is also confusing in the sense of the number of inflected cases. It seems that these languages are generally ill-suited for the traditional definition of case, because they are polysynthetic, for example, and involve noun incorporation, such as Wichita (Caddoan language) and Ainu (isolate). The languages in category i) may have several adverbially-used cases, if none of these contributes to the marking of dependent nouns.

For the statistical analysis, I include category b), despite some of its languages being more reminiscent of category a) than, for instance, Romanian, which is included in category b) with its two inflected cases. I leave out category i) languages from the sample, since it is unclear exactly how the hypothesis regarding the number of cases should respond to such a heterogeneous and problematic group.[129] The sample is further controlled for the effect of closely related languages, potentially skewing the results by using a sampling method, introduced by Matthew Dryer (1989; 1992), which excludes data points that belong to the same genealogical subfamily (or its equivalent, called a genus by Dryer) and which displays the same values for both the definite article and the number of cases. Therefore, Polish and Russian are counted as a single data point, since both belong to the Slavic subfamily of Indo-European languages, have six cases, and no definite article. After these procedures, the number of languages categorized both for the definite article and the number of cases is 144, forming the

[128] Cf. "the Queen of England's birthday."

[129] Category i) also seems to be rather indecisive regarding the occurrence of the definite article: Of the 9 languages that remain in the sample, 5 have the definite article.

sample used in the statistical analysis. The distribution of these features among the languages is shown in **Table 10**:

Number of cases in the language		No definite article	Definite article
	0	22	35
	2	3	12
	3	2	5
	4	1	3
	5	5	1
	6–7	18	6
	8–9	9	7
	>10	9	6

Table 10. Number of cases vs. definite article. N = 144 languages.

The correlation between the number of cases and the definite article is evaluated with Kendall's tau,[130] which gives the coefficient -0.20. P-value < 0.00001. This indicates that, in support of the hypothesis, there is an inverse correlation between the number of cases and the occurrence of the definite article. In other words, the fewer cases there are in a language, the greater the likelihood that there is also a definite article in that language; conversely, the more cases there are, the less likelihood there is of having a definite article.

According to the hypothesis, the grammaticalization of a definite article and the loss of case are partly dependent on each other. However, the examples from Indo-European languages spoken in Europe indicate that the definite article also emerged in languages that preserve some case inflection, although no more than four cases. In **Table 11** the organization of data assumes that having no more than four cases is significant for the occurrence of a definite article:

[130] Kendall's tau is a coefficient that shows the degree of concordance between two ranked variables within a set of data. The values vary between 1 and -1, where 0 indicates a total lack of correlation and 1 represents a perfect positive correlation (-1 standing for a perfect inverse correlation). The coefficient is calculated by counting the proportion of concordant pairs minus the proportion of discordant pairs. Concordance means that a data point's value for one of the variables ranks higher than the values of the other data points which are ranked lower based on the other variable.

		more than four cases	
		no	*yes*
definite article	*no*	27	46
	yes	36	20

Table 11. The occurrence of the definite article in a language with more than four cases. N = 129 languages.[131]

The table shows that organizing the data into two groups and setting the dividing line between four and five cases yields a more significant distribution of the occurrence of the definite article than any other division based on the number of cases, presented in **Table 10**. Fisher's exact text gives $p = 0.000094$. The odds ratio drawn from **Table 11** tells us that if there are fewer than five cases in a language, then the odds of the language not having the marking of definiteness are three times less than if the language had five cases or more. Conversely, if the language has five or more cases, the odds of not having the marking of definiteness are three times greater than if the language had fewer than five cases.

Finally, to make sure that the observed correlation is universal and not dependent on the effect of areality, I tested the data for one areal division, discussed in **4.1.2**, the six macroareas proposed by Dryer (1992), with some minor changes introduced by WALS (Haspelmath et al. [ed.] 2013).

	Africa	Eurasia	Southeast Asia & Oceania	Australia-New Guinea	North America	South America
definite article, fewer than five cases	15	7	7	3	10	4
no definite article, fewer than five cases	4	2	6	3	7	5
definite article, more than four cases	4	5	1	6	4	2
no definite article, more than five cases	2	16	2	5	2	7
Fewer than five cases	0.79	0.78	0.54	0.50	0.59	0.44
More than four cases	0.67	0.24	0.33	0.55	0.67	0.22

Table 12. The distribution of data in six macroareas and the proportion of genera containing languages with definite article as opposed to no definite article.

The inverse correlation between the number of cases and definite articles in languages with fewer than five cases is shown in four macro areas: Africa, Eurasia, Southeast Asia and Oceania, and North-America. The tendency is observed in languages with more than four cases in Eurasia,

[131] This sample too is controlled for the effect of closely related languages by Matthew Dryer's (1992) method, leaving 129 languages to be analyzed.

Southeast Asia and Oceania, and South-America. However, the division into six macro areas reduces the number of languages in some of the areas in a way that is too small to be conclusive; this is the case especially with Australia-New Guinea and North America, which do not conform to the tendency. With these areas, the addition of a single data point may change the result in favor of the hypothesis regarding the order of proportions of genera containing languages with definite article as opposed to no definite article. Nevertheless, the most important observation is that the tendency is not limited to Europe, although, admittedly, Eurasia shows the most significant correlation.

4.3.3 Discussion

With the hypothesis proposed in **4.3.1**, an inverse correlation between the number of cases and the existence of a definite article seems possible in the languages of the world.[132] Yet this finding does not imply a direct causal link, but rather suggests that a mechanism or mechanisms operating between case and the marking of definiteness may be nevertheless taking place in a large number of languages. In terms of typological universals, discussed in **4.1.3**, this tendency could be described as a restricted probabilistic universal: The likelihood of the occurrence of the definite article increases the fewer cases there are in a language. However, it appears that, while this tendency is found to some extent in a major part of the world, it is not found everywhere: in some places, either it does not show up at all or it is not found in languages with more than four or fewer than five cases. When such areal dependency is established with individual features or their co-occurrence, the primary hypothesis is that the feature is not universal in the sense of stemming from the inherent qualities of human cognition or other sources (for potential sources for linguistic universals, see Christiansen & Chater 2008), but rather that it has probably been diffused through language contacts (see **3.3.4**). Yet it is not clear exactly what is disseminated through language contact that would result in the increased likelihood of the occurrence of one feature versus fewer distinctions being made in another category. However, I attempt to give at least a partial answer to this in the present subsection.

[132] To solidify further the hypothesis analyzed here requires the introduction of more data and perhaps the use of alternative sampling methods, including a wider range of statistical tests.

If we follow Moravcsik's (2011, 71–74) analysis of typological universals as an explanation, discussed in **4.1.3**, for the co-occurrence of case loss and the emergence of the definite article by a *universal structural explanation*, we encounter a problem: First, what is observed cross-linguistically here is not an absolute restricted universal, but a potential tendency, one that, moreover, is not found in certain parts of the world. It follows that even if Balkan Slavic and other European languages conform to the tendency, a single *universal historical or functional generalization* may not be relevant for all of these languages, since the probabilistic universal, observed here, means that there is a possibility that the situation did not result from this universal exerting pressure to develop toward or maintain the preferred pattern, but rather came about for other reasons. Yet in the languages that conform to the tendency, the phenomenon can be coincidental only in some of them, as the analysis in **4.3.2** showed. In this respect, a probabilistic restricted universal, like the tendency observed here, differs from a single feature whose distribution follows an areal pattern. The single feature can be dismissed by stating that it must have spread through language contact or that it resulted from common inheritance, while in the case of the co-occurrence of case loss and the emergence of the definite article, we do not know what, if anything,[133] was the precise effect of the contact or what common inherited properties affected it. Therefore, the mechanism behind the tendency still requires an explanation.

We thus encounter two alternatives in trying to account for the co-occurrence of case loss and the emergence of the definite article: Either a language that has fewer cases is more susceptible to the emergence of the definite article (or to acquiring it through language contacts), or a language that develops a definite article (possibly through contact) becomes more prone to lose case inflection. Although the chronological chain of events remains typically opaque in the European languages, the first option seems likelier than the second. If we consider further Tiraspol'ski's (1980) proposition that the definite article, initially inflected for case, pushed aside the original case inflection, what remains to be explained is why the definite article too ceased to inflect. Also, how does the definite article explain the case loss in non-articulated words? It seems more logical that the emergence of a new category is related to the need to compensate somehow for the diminishing number of categories resulting from the case loss rather than

[133] It could be that the explanandum in fact becomes the element that prevents the tendency to manifest in certain parts of the world.

trying to argue that the emergence of a new category leads to the loss of existing categories. It has been rightfully questioned whether there is an all-encompassing phenomenon whereby the loss of complexity in one part of a grammar is countered by the increase in complexity in another (see, e.g., Sinnemäki 2014; 2008; Shosted 2006). Yet in this case, a mechanism, one that is also suggested by the so-called *complexity trade-off hypothesis*, seems more fruitful.

The best-known effect of the loss of case inflection is the increased analytical marking of grammatical relations (discussed in **3.3**). Also, in a caseless language the core functions of grammar are more often expressed by word order than they are in languages that have cases, a common observation supported by quantitative typological evidence (Sinnemäki 2008, 82). This was the very phenomenon that David Crystal (1987, 6) chose to echo the complexity trade-off hypothesis: "All languages have a complex grammar: there may be relative simplicity in one respect (e.g., no word-endings), but there seems always to be relative complexity in another (e.g., word-position)." This latter consequence of case loss is shown in the relatively rigid word order of most of the West European languages by comparison, for instance, with those Slavic languages that have case inflection (for Germanic and Romance languages, see Bentz & Christiansen [In print]). In the Balkan Slavic languages, deviation from the default SVO word order is constrained by the identifiability of the subject and the object, as examples **(73)** and **(74)** show:

(73) Bulgarian (Gebert 2009, 315)

Marija	*običa*	*Ivan*
Marija	loves	Ivan

'Marija loves Ivan'

(74) Bulgarian (Gebert 2009, 316)

Marija	*ja*	*običa*	*Ivan*
Marija	it.F.ACC.SG.CL	loves	Ivan

'Marija is loved by Ivan'

In example **(73)**, only one reading pertaining to the default SVO word order is possible, since the nouns are not marked for case. Nevertheless, in example **(74)**, the object may be fronted when the doubling object clitic is introduced. In other Slavic languages, the fronting of the object is usually possible, since nouns are marked for the accusative case:

(75) Polish

Marię	*kocha*	*Jan*
Maria.ACC	loves	Jan.NOM

‘Maria is loved by Jan’

The *language-specific historical generalization* (in Moravcsik’s terms) about the relationship between the loss of case inflection and the rise of the definite article, or put more modestly, some reasons for it in Balkan Slavic are found, in my view, in the pragmatic functions of word-order variation. The fronting of the object, shown in examples **(74)** and **(75)**, pertains to the marking of the information structure of the sentence. The fronting of one part of the sentence adheres to a widely observed, left-to-right progression from given knowledge to new information (Mereu 2009, 77–78). This fronted, given knowledge, the topic of the sentence, is contrasted with the new information, the focus of the sentence.[134] The loss of case inflection in Balkan Slavic and, most important, the weakening of the nominative–accusative distinction led to increasing difficulties in using fronting—also called topicalization—as a strategy to mark information structure. The new restrictions on word order were compensated for by an increased adnominal use of demonstratives to mark topical NPs. Thereafter, the increased frequency of the inherently definite adnominal demonstratives contributed to the grammaticalization of the definite articles to mark non-topical definite referents as well. Later, the fronting strategy re-gained prominence, enabled by clitic doubling. It must be noted that the stages of grammaticalization in clitic object doubling can be observed synchronically in Balkan Slavic: In Bulgarian, it is still connected with the marking of information structure, whereas in a number of Macedonian dialects, including standard Macedonian, one of its key conditioning criteria is definiteness (Gebert 2009; see also **3.3.2**). Thus, in some of the dialects both adnominal demonstratives and clitic doubling developed from a marker of information structure into contributing to the marking of definiteness.

As was briefly mentioned in **4.3.1**, Heine and Kuteva (2006, 106–119) argue that the anaphoric use of demonstratives is characteristic of an incipient-stage definite article. This is in line with the proposed chain of events connecting the definite article and case loss, since topical and anaphoric NPs often coincide. However, Diessel (1999, 128–129) notes that

[134] Other pairs of terms, roughly coinciding with topic and focus, include topic–comment, theme–rheme, given–new, presupposition–focus, background–foreground, and presupposition–assertion (Mereu 2009, 77).

definite articles tend to develop from anaphoric demonstratives, while anaphoric adnominal demonstratives do not usually refer to the topic of the previous sentence, but reintroduce another discourse-old NP—unlike definite articles, which may refer to the topic of the previous sentence. The relationship between anaphoricity and topicality is complex, and requires a distinction between a discourse-level topic and a sentence-level topic to be made. While the model to account for the co-occurrence of the definite article and case loss involves sentence-level topics, it could be hypothesized that the first occurrences of demonstratives used as definite articles could be observed in the anaphoric use of the demonstratives, which refer to a discourse topic introduced in the immediately preceding sentence.

4.3.4 Conclusions

In the previous subsection, I tentatively proposed a language-specific historical generalization for Balkan Slavic to account for the possible universal tendency to an inverse correlation between the number of cases and the occurrence of a definite article, observed in **Subsection 4.3.2**. While some aspects of the proposed development need more study, it could be asked whether a similar explanation may also be valid as a *universal historical generalization*. The first problem is related to the number of cases. Based on what was observed, the increased odds of having a definite article occurs with languages that have fewer than five cases. Nevertheless, it is unlikely that the number of cases would be the ultimate deciding factor. Rather, in languages with only a few cases, there is a greater chance that some of the key distinctions may remain unmarked or homonymic with other cases. The proposed model emphasizes the significance of homonymy between the marking of subjects and objects. Given that four or fewer cases in the sample languages may mark any grammatical roles, it would be extremely useful for testing the model to determine the role of nominative–accusative type homonymy in the occurrence of definite articles. Another question is how well the proposed model suits non-European languages. It might well be thought, for example, that non-configurational languages, that is, languages whose word order is dictated largely by pragmatic, not syntactic, considerations may behave very differently from the European languages (see e.g. Mereu 2009, 83–95).

Although the proposed model can be read as an independent, language-internal explanation for the observed tendency, it does not exclude the role of language contacts in the process. As discussed in **4.3.1**, the definite article

seems to be prone to diffusion through language contacts. It is likely that language-internal susceptibility resulting from case loss, as suggested by the model, also speeds up the acquisition of the definite article in language contact. More specifically, in the case of Balkan Slavic, the spread of the postposed demonstratives to mark definiteness—of recently introduced discourse topics, for instance—may be significantly eased if similar functional extension has occurred in the contact languages that realize the definite article in a structurally similar manner, as do Balkan Romance and Albanian.

Finally, it must be asked what a probabilistic universal that affects a feature of a hypothesized linguistic area does for the credibility of that feature as evidence for the convergence between the languages. As Balthasar Bickel and Johanna Nichols (2006, 3–4), cited at the outset of this chapter, state: “Our impression is that the classic Balkan features include a few variables of sufficiently low [cross-language] frequency to be of diagnostic value” and “[T]he classic Balkanisms do not do a very complete job of defining the shared grammar that makes for the notable intertranslatability of Balkan languages.” It is easy to agree with the authors on the first point: The definite article, for example, is not a particularly unexpected feature, and even less so if it is affected by a typological universal that predicts its occurrence as being more likely because of another feature of the language. However, the fact that the definite article is realized as postposed makes it more exceptional. Definite articles similarly realized as in Balkan Slavic, Albanian, and Balkan Romance can be found in Europe only in Basque and Scandinavian languages, all of which are situated on the opposite peripheries of the continent. Yet, in my view, this excursion into the interrelation between case loss and the definite article, as well as the analysis of the recipient–possessor merger have both contributed to the understanding of the “notable intertranslatability of the Balkan languages” in different ways. In **4.2**, a number of smaller, potentially contact-induced phenomena related to the recipient–possessor merger were introduced. On the other hand, as was argued in **3.4**, a large number of L2 speakers was a key factor in the loss of case inflection in Balkan Slavic. Further, based on the model proposed here, case loss potentially sped up the acquisition of the definite article. Therefore, in the recipient–possessor merger and the emergence of the definite article, we observe complex, yet nonetheless contact-induced phenomena that account for some key

characteristics of the Balkan linguistic area, although these characteristics are not cross-linguistically unique.

5 Conclusions

In this final chapter, I provide a synthesis summarizing the key findings of **Chapters 2**, **3**, and **4**, which analyzed the loss of case inflection in Balkan Slavic from diachronic, areal, and typological perspectives, respectively. The evolution of the field of linguistics is directly reflected in the study's three different approaches. As an independent research topic, the loss of case inflection in Bulgarian and Macedonian has a century-long history, one that began with Karl Meyer's (1920) monograph. Yet early notions about the Balkan Slavic case loss go back to the emergence of Slavic philology and the beginnings of the study of the Balkan linguistic area in the early 19th century. This span of almost two centuries also witnessed the diversification of the field of linguistics, leading to the emergence of numerous specialized subfields. In the 19th century, philology, dialectology, and comparative historical linguistics dominated linguistic research, leading to the study and publication of critical editions of old manuscripts, including those that could shed light on the earlier stages of Balkan Slavic. Therefore, the great Bulgarian linguists of the late 19th and early 20th centuries, Ljubomir Miletič and Benjo Conev, already had access to a major part of the relevant sources when they formulated their views on the development of the Balkan Slavic case system.

The 1920s saw the introduction of a new perspective on the loss of case inflection in Balkan Slavic. Two events, Kristian Sandfeld's (1926) urging for a new field, "Balkan philology," and Nikolai Trubetzkoy's (1928) proposed definition of a *Sprachbund*, illustrate the increased interest in language contact as a source for language change. Previously, any similarities between adjacent languages not attributable to common inheritance were usually considered as resulting from a substrate that must have influenced the languages of that area. It was then the linguist's job to come up with the most credible hypothesis about the substrate language in question. Since World War II, our understanding of the dynamics of language contacts has increased tremendously, culminating, on the one hand, in Sarah G. Thomason and Terrence Kaufmann's (1988, 14) laconic statement that "any linguistic feature can be transferred from any language to any other language." On the other hand, an understanding of the results of

extreme contact situations, like creolization, has illuminated the role that a particular sociolinguistic setting plays in the effect of the language contact.

Part of linguistics has always been occupied with devising restraints on what is possible or likely in a language. These formalist approaches, in their search for a systematic description of language, have been racing against the accumulation of better data from an ever increasing number of languages. The astonishing diversity of the languages of the world has proved time and again that many generalizations about the limits of language are far too limited. Similarly, by comparison with scholarship one hundred years ago, what the study of the Balkan Slavic case loss faces today is a vast amount of data about similar phenomena around the world. In addition, the contact-induced origin of many features of the Balkan linguistic area is now questioned on the basis of what we currently know about what is typical or common in languages. At the same time, this data, when organized into databases, allows a wide range of new questions to be asked about the co-occurrence of features or their susceptibility to diffusion through language contacts.

5.1 Sound changes and the case system: The inability to regenerate

A topical question in **Chapter 2**—and also in the earlier literature—was the effect of sound changes on the Balkan Slavic case system. Could the case homonymy caused by the sound changes be significant enough to have derailed the entire case system? A positive answer to this question seems to entail the further claim that, since the sound changes must be viewed as an internal development in the language, no external explanations for the case loss need to be sought. However, as the discussion in **Section 2.4** demonstrated, the whole question may be wrongly formulated. Even though numerous authors have maintained that sound changes of the same magnitude did not affect the case systems of the other Slavic languages, their claim is not entirely true. While some resulting syncretism is clearly tolerated, it is particularly striking that all of the Slavic languages except Balkan Slavic resorted to one or more strategies in order to revive, at least in part, the accusative–nominative homonymy caused by the sound changes. Thus, the notable difference with Balkan Slavic is that it was unable to resort to alternative strategies to revive the lost distinctions, and thus the sound changes proved fatal for the inflectional strategy to mark grammatical roles. Therefore, the question we should be asking is why were the detrimental effects of the sound changes not countered in Balkan Slavic? I

argued in **Section 3.4** that the failure to maintain the inflectional case system through new innovations most likely resulted from a particular mechanism of language contact. Balkan Slavic had a significant number of L2 speakers, who favored explicit and cross-linguistically easily identifiable analytical strategies in dealing with the lost case distinctions. This indiscriminate effect of L2 speakers must be seen as one of the key causes of the typological change from case inflection toward non-inflection.

However, the dialectally-preserved vestiges of case inflection show that not all cases were treated in an equal manner in the process of case loss. From the analysis in **Chapter 2**, it emerged that the accusative singular tends to be preserved for a very long time if a form distinct from the nominative exists. The frequent *(j)ā*-stem singular accusative forms were rendered homonymous with the nominative in most dialects. Yet, as I argued in **Section 2.2.2**, the 17th-century Damascene literature suggests that, if a stress fell on the case ending, a distinct accusative was preserved in a large number of mainly Bulgarian dialects until a relatively late date. The particular status of the accusative singular is also reflected in the best-preserved inflected case form, the genitive–accusative of the animate masculine singulars.

On the level of word forms, two key observations were made in **Chapter 2**: Of all the examples of erroneous use of case forms in the manuscripts, the neuters occupy a much more significant part than would be expected on the basis of their frequency in the texts. If indeed the neuters lost their case inflection earlier than other nouns, this could be partly explained by the fact that their referents are typically inanimate and the contexts where they appear would exert less pressure on the disambiguation of their grammatical roles through case inflection. In addition, it is often said that the Balkan Slavic case inflection, before its loss, showed a preference for hard-stem endings. Yet the dialectally preserved *(j)ā*-stem dative singular forms show that the process must have been more complicated, since it must be assumed that either both the hard *ā*-stem and the soft *jā*-stem forms contributed to these forms or that the origin of some of the forms is in the *ā*-stem genitive.

5.2 The loss of case inflection as a Balkan development

The more specific mechanism of contact operating between the languages of the Balkan linguistic area is based on the mutual bilingualism of the speakers of the different languages. The result of these contacts is shown by a grammatical convergence between the languages. Because this

convergence is still best understood as a bundle of areally significant occurrences of individual features, it is different from the overall effect of the L2 speakers, which hampered the regeneration of the inflectional case system. However, the reduced Balkan Slavic case system was also highly influenced by specific contact-induced change, that is, the spread of individual features, since not all loss of case inflection resulted from sound changes. In **Section 2.2.4**, it was shown that, in the manuscripts, the locative and instrumental display signs of weakening early. The loss of these two cases is connected to two functional mergers, the merger of goal and location and the merger of instrument and accompaniment, both of which are also well-established Balkan features.

As demonstrated in **Section 3.3.4**, the accusative is one of the best-preserved case forms, not only in Balkan Slavic, but also in the remaining Balkan languages, yet the likelihood of there being an accusative form increases if the noun is singular, masculine, or definite. Although there is evidence that these criteria, collectively called the effect of referential scales, may not be universal, but rather prone to spread through language contact, there is no clear indication that the preserved accusative forms in the Balkan linguistic area were influenced by the processes of mutual re-enforcement. Nevertheless, there are other phenomena pertaining to the marking of direct objects which show more uniform patterning across the different varieties.

The most notable shared feature of the Balkan case systems is the merger of the marking of recipient and adnominal possessor. Discussed from different perspectives in **Subsection 2.2.4** and **Sections 3.2** and **4.2**, the recipient–possessor merger is also found in the Middle Balkan Slavic case system, proposed in Section **2.4**. While Balkan Slavic shows a uniform development, where the source of the merged case was exclusively the dative (with the possible exception of *(j)ā*-stem dative singular forms), other Balkan languages display mixed sources within one language. There is evidence that points to a rather long development in the formation of the merged case in the Balkans, which, as I argue, should be assumed to extend from the 6th to the 16th century. Yet the dative–accusative merger in the Northern Greek dialects, which deviate from the general Balkan development, may result from a rather early replacement of the dative by the accusative, predating the period of the formation of the general Balkan recipient–possessor merger. The later uniform analytical marking of adnominal possessors and recipients in Balkan Slavic is not unique, but has

full or partial parallels in the other Balkan languages. I argued that some of the differences in the realization of the merger between the Balkan languages are due to the inherently different structural properties of the languages, and therefore absolute parallelism cannot be expected as result of language contact unless the contact affects those structural properties as well.

5.3 The loss of case inflection from a cross-linguistic perspective

The contact-induced origin of one of the best-known Balkanisms, namely, the recipient–possessor merger, has been disputed by some researchers along with certain other assumed features of the Balkan linguistic area. The critics point out that these features are cross-linguistically too common and thus could have resulted by chance. In **Section 4.2**, I argued that, despite not being altogether unique among the European languages, the nearly uniform realization of the merger in a contiguous geographical area is still striking, and that the odds of such a merger coinciding in all the languages of the Balkan linguistic area by chance are very small. In addition, I demonstrated that the notion of typological commonness is significantly affected by the level of observation. If the feature is defined in functional terms, then the structural morphosyntactic similarities, typical of the Balkan linguistic area, may go unnoticed. Moreover, in **Subsection 4.2.4**, I introduced two additional phenomena related to the recipient–possessor merger that show areal clustering in the Balkans, which further demonstrates that that particular area of grammar is prone to contact-induced changes. This too shows that studying one isolated feature means that the reasons for such a high level of "intertranslatability" among the Balkan languages may be missed.

From a typological perspective, case loss seems to go hand in hand in Western and Central Europe with another feature, pertaining to the marking of NP. Like in Balkan Slavic, also in numerous other European languages the loss of case inflection coincides timewise with the emergence of definite article. **Section 4.3** formed an attempt to find, whether these two phenomena are connected universally, or whether their co-occurrence is observed only areally. In this preliminary study I demonstrated that there may exist a universal tendency, pending further research, where the number of cases in a language is inversely correlated with the occurrence of a definite article. Although their correlation does not imply direct causation, on the basis of Balkan Slavic, I proposed a potential historical mechanism that may connect

the two features. Whether this mechanism, pertaining to the marking of information structure, is relevant also cross-linguistically, remains to be studied in more detail. However, if these two features are connected universally this weakens somewhat their role as evidence for language contacts. Yet, it is probable that the loss of case inflection may only make the language more susceptible for the adoption of the definite article—which is known to diffuse easily in language contact.

***

This study has been an attempt to deal with a classical problem in the study of Slavic languages, namely, the Balkan Slavic case loss, an issue that furthermore is related to a significant number of phenomena that have been studied in the context of the Balkan linguistic area. I have tried to demonstrate that a question such as loss of case inflection cannot be approached from only one viewpoint, but that the diachronic, areal, and typological perspectives are all crucial to understanding the dynamics of the historical process. While several questions remain to be addressed by future research, I believe that we can say with relative certainty that the loss of case inflection in Bulgarian and Macedonian resulted from the combined effect of different types of language contact that the languages have undergone since the arrival of their speakers in the Balkans. While the primary reason for the inability to maintain case inflection may have been rather indiscriminate and simple, the form the process took is evidence for a complex and fascinating interplay among internal language processes, universal qualities of language, and changes resulting from contacts with speakers of the other languages of the region.

References

Adams, James Noel. 2013. *Social variation and the Latin language.* Cambridge: Cambridge University Press.

Aikhenvald, Alexandra Y. 2006. Grammars in contact: A cross-linguistic perspective. In Alexandra Y. Aikhenvald & R[obert] M. W. Dixon (eds.), *Grammars in contact: A cross-linguistic perspective*, vol. 4, 1–66. Oxford: Oxford University Press.

Anderson, Stephen. 1985. Inflectional morphology. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 3, *Grammatical categories and the lexicon*, 150–201. Cambridge: Cambridge University Press.

Anward, Jan & John Swedenmark. 1997. *¡Kasus nej, bestämdhet ja! Om Möjliga modeller av nominalböjningens utveckling i svenskan.* (Studier i svensk språkhistoria 4). Stockholm: Stockholms universitet.

Aronson, Howard I. 2007. *The Balkan linguistic league, "orientalism", and linguistic typology.* Ann Arbor/New York: Beech Stave Press.

Arumaa, Peeter. 1985. *Urslavische Grammatik: Einführung in das vergleichende Studium der slavischen Sprachen. Formenlehre*, vol. 3. Heidelberg: Winter.

Asenova, Petja. 2002. *Balkansko ezikoznanie: Osnovni problemi na balkanskija ezikov săjuz.* 2nd edition edn. Veliko Tărnovo: Faber.

Baerman, Matthew, Dunstan Brown, & Greville G. Corbett. 2005. *The syntax-morphology interface: A study of syncretism.* Cambridge: Cambridge University Press.

Barðdahl, Jóhanna. 2001. *Case in Icelandic: A synchronic, diachronic and comparative approach.* Lund: Lund University.

Barford, P. M. 2001. *The early Slavs: Culture and society in early medieval Eastern Europe.* London: British Museum Press.

Bauer, Laurie. 2009. IE, Germanic: Danish. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press.

Beekes, Robert S. P. 2011. *Comparative Indo-European linguistics: An introduction.* 2nd edn. Amsterdam: John Benjamins Publishing Company.

Bentz, Christian & Bodo Winter. 2013. Languages with more second language learners tend to lose case. *Language Dynamics and Change* 3. 1–27.

Bentz, Christian & Morten H. Christiansen. [In print]. Linguistic adaptation: The trade-off between case marking and fixed word orders in Germanic and Romance languages.

Bernštejn, Samuil Borisovič. 1948. *Razyskanija v oblasti bolgarskoj istoričeskoj dialektologii*, vol. 1, *Jazyk Valašskih gramot XIV–XV vekov.* Moscow: AN SSSR.

Bickel, Balthasar. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11(1). 239–251.

Bickel, Balthasar. 2008. A general method for the statistical evaluation of typological distributions [manuscript], http://www.spw.uzh.ch/bickel-files/papers/testing_universals_bickel2008.pdf, (22 August 2013).

Bickel, Balthasar. 2011. Putting variation center stage: Beyond "language" (or "dialect") as the basic data unit in typology (and elsewhere). Paper presented at Variation and Typology: New Trends in Syntactic Research, Helsinki, 25–27 August 2011.

Bickel, Balthasar & Johanna Nichols. 2006. Oceania, the Pacific Rim, and the theory of linguistic areas. Paper presented at Proceedings of the 32nd Annual Meeting of the Berkeley Linguistics Society, Berkeley, http://www.spw.uzh.ch/bickel-files/papers/bickel_nichols_bls32.pdf, (21 October 2012).

Bickel, Balthasar & Alena Witzlack-Makarevich. 2008. Referential scales and case alignment: reviewing the typological evidence. In Andrej Malchukov & Marc Richards (eds.), *Scales* (Linguistische Arbeitsberichte 86), 1–37. Leipzig: Institut für Linguistik.

Bickel, Balthasar, Alena Witzlack-Makarevich, & Taras Zakharko. [Submitted draft]. Typological evidence against universal effects of referential scales on case alignment.

Blake, Barry J. 2001. *Case.* 2nd edn. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Boel, Gunnar de. 2008. The genesis of clitic doubling from Ancient to Medieval Greek. In Dalina Kallulli & Liliane Tasmowski (eds.), *Clitic doubling in the Balkan languages*, 89–103. Amsterdam: Benjamins.

Bogorov, Ivan. 1986 [1848]. *Părvička bălgarska gramatika: Fototipno izdanie.* Sofia: Nauka i izkustvo.

Brunot, Ferdinand & Charles Bruneau. 1949. *Précis de grammaire historique de la langue française.* 3rd edn. Paris: Masson.

Campbell, Lyle. 1985. Areal linguistics and its implications for historical linguistics. In Jacek Fisiak (ed.), *Papers from the 6th International Conference on Historical Linguistics*, 25–56. Poznań: John Benjamins Publishing Company.

Campbell, Lyle. 2006. Areal linguistics: A closer scrutiny. In Yaron Matras, April MacMahon, & Nigel Vincent (eds.), *Linguistic areas: Convergence in historical and typological perspective*, 1–31. New York: Palgrave Macmillan.

Campbell, Lyle & Richard Janda. 2001. Introduction: Conceptions of grammaticalization and their problems. *Language sciences* 23(2). 93–112.

Campos, Héctor & Melita Stavrou. 2004. Polydefinite constructions in Modern Greek and in Aromanian. In Olga Mišeska Tomić (ed.), *Balkan syntax and semantics*, 137–173. Amsterdam: John Benjamins Publishing.

Catasso, Nicholas. 2011. Genitive-Dative Syncretism in the Balkan Sprachbund: An Invitation to Discussion. *SKASE Journal of Theoretical Linguistics* 8(2), http://www.skase.sk/Volumes/JTL19/pdf_doc/04.pdf. (17 June 2012.)

Češko, E. V. 1970. *Istorija bolgarskogo sklonenija.* Moscow: Nauka.

Christiansen, Morten H. & Nick Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31(5). 489–509.

Civ'jan, Tat'jana. 1965. *Imja suščestvitel'noe v balkanskih jazykah: k strukturno-tipologičeskoj harakteristike balkanskogo jazykovogo sojuza.* Moscow: Nauka.

Comrie, Bernard. 1981. *Language universals and linguistic typology: Syntax and morphology.* Oxford: Blackwell.

Comrie, Bernard. 1986. On delimiting cases. In Richard D. Brecht & James S. Levine (eds.), *Case in Slavic*, 86–106. Columbus, OH: Slavica Publishers.

Comrie, Bernard, Martin Haspelmath, & Andrej Malchukov. 2006. Questionnaire on ditransitive constructions, http://www.eva.mpg.de/lingua/tools-at-lingboard/pdf/Malchukov_DitrQuest_3Andrej.pdf. (23 September 2013.)

Conev, B[enjo]. 1914. *Vračansko evangele.* (Bălgarski starini. Book 4). Sofia: Bălgarskata akademija na naukitě.

Conev, Benjo. 1984 [1919]a. *Istorija na bălgarskija ezik*, vol. 1. Sofia: Nauka i izkustvo.

Conev, Benjo. 1984 [1919]b. *Istorija na bălgarskija ezik*, vol. 2. Sofia: Nauka i izkustvo.

Conev, Benjo. 1985 [1919]. *Istorija na bălgarskija ezik*, vol. 3. Sofia: Nauka i izkustvo.

Crystal, David. 1987. *The Cambridge encyclopedia of language.* Cambridge: Cambridge University Press.

Csepregi, Márta. 1991. *Unkarin kielioppi.* Helsinki: Finn Lectura.

Cvetanovski, Goce. 2010. *Govorot na makedoncite vo Mala Prespa.* Skopje: Institut za makedonski jazik "Krste Misirkov".

Cyhun, Gennadij. 1981. *Tipologičeskie problemy balkanoslavjanskogo jazykogo areala.* Minsk: Nauka i tehnika.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity.* (Studies in language companion series 71). Philadelphia, PA: John Benjamins Publishing Company.

Dahl, Östen & Maria Koptjevskaja-Tamm. 2006. The resilient dative and other remarkable cases in Scandinavian vernaculars. *STUF - Sprachtypologie und Universalienforschung* 59(1). 56–75.

Daniliuc, Laura & Radu Daniliuc. 2000. *Descriptive Romanian grammar: an outline.* Munich: LINCOM.

Dardel, Robert de. 1964. Considérations sur la déclinaison romane a trois cas. *Cahiers de Ferdinand Saussure* 21. 7–23.

Dëmina, Evgenija I. 1985a. *Tihonravovskij damaskin: bolgarskij pamjatnik XVII v.: issledovanie i tekst*, vol. 1. Sofia: Izd[atelstvo]. na Bălg[arskata]. Akad[emija]. na Naukite.

Dëmina, Evgenija I. 1985b. *Tihonravovskij damaskin: bolgarskij pamjatnik XVII v.: issledovanie i tekst*, vol, 2. Sofia: Izd[atelstvo]. na Bălg[arskata]. Akad[emija]. na Naukite.

Demiraj, Shaban. 1993. *Historische Grammatik der albanischen Sprache.* (Schriften der Balkan-Kommission, 34). Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Diessel, Holger. 1999. *Demonstratives: Form, function, and grammaticalization.* Amsterdam: Benjamins.

Dimitrescu, Florica, Viorica Pamfil, Elena Barborică, Maria Cvasnîi, Mirela Theodorescu, Cristina Călăraşu, Mihai Marta, Elena Toma, & Liliana Ruxăndoiu. 1978. *Istoria limbii române: Fonetică, Morfosintaxă, Lexic.* Bucharest: Editura Didactică şi Pedagogică.

DIRM. 1991. *Declaraţia de independenţă a Republicii Moldova.* http://www.istoria.md/articol/573/ (4 May 2014).

Dixon, R[obert] M. W. 2010a. *Basic linguistic theory*, vol. 1, *Methodology.* Oxford: Oxford University Press.

Dixon, R[obert] M. W. 2010b. *Basic linguistic theory*, vol. 2, *Grammatical topics.* Oxford: Oxford University Press.

Dixon, R[obert] M. W. 2012. *Basic linguistic theory*, vol. 3, *Further grammatical topics.* Oxford: Oxford University Press.

Donohue, Mark. 2012. Typology and Areality. *Language Dynamics and Change* 2(1). 98–116.

Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13(2). 257–292.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1). 81–138.

Dryer, Matthew S. 2013. Definite Articles - Chapter 37. In Martin Haspelmath, Matthew S. Dryer, David Gil,, & Bernard Comrie (eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. http://wals.info/feature/37 (10 August 2014).

Dumitrescu, Andrei. 2008. Romanian kielen opintopiiri [Study material for the basic course of Romanian]. http://www.ling.helsinki.fi/~dumitres/curs/ (7 May 2014).

Duridanov, Ivan. 1956. Kăm problemata za razvoja na bălgarski ezik ot sintetizăm kăm analitizăm. In *Godišnik na sofijskija universitet, Filologičeski fakultet*, 1955, vol. 51(3). Sofia: Dăržavno izdatelstvo "Nauka i izkustvo".

Duridanov, Ivan. 2004. Die entwicklung des Bulgarischen vom Syntetischen zum analytischen Sprachtyp. In Uwe Hinrichs (ed.), *Die europäischen Sprachen auf dem Wege zum analytischen Sprachtyp*, 213–230. Wiesbaden: Otto Harrassowitz Verlag.

Dürscheid, Christa. 2007. Quo vadis, Casus? Zur Entwicklung der Kasusmarkierung im Deutschen. In Hartmut Lenk & Maik Walter (eds.), *Wahlverwandtschaften: Valenzen, Verben, Varietäten. Festschrift für Klaus Welke zum 70. Geburtstag*, 89–112. Hildesheim, Zürich & New York: Olms.

Feuillet, Jack. 1996. *Grammaire synchronique du bulgare.* Paris: Institut d'études slaves.

Feuillet, Jack. 1999. *Grammaire historique du bulgare*. Paris: Institut d'études slaves.

Franks, Steven. 2009. Macedonian pronominal clitics as object agreement markers. In Steven Franks, Vrinda Chidambaram, & Brian D. Joseph (eds.), *A linguist's linguist: Studies in south Slavic linguistics in honor of E. Wayles Browne*, 189–221. Bloomington, IN: Slavica.

Franks, Steven & Tracy Holloway King. 2000. *A handbook of Slavic clitics.* New York: Oxford University Press.

Friedman, Victor A. 1991. Case in Romani: Old grammar in new affixes. *Journal of the Gypsy Lore Society* 5(1). 85–102.

Friedman, Victor A. 1998. The implementation of standard Macedonian: Problems and results. *International Journal of the Sociology of Language* 131(1). 31–58.

Friedman, Victor. 2001. *Macedonian.* http://www.seelrc.org:8080/grammar/pdf/stand_alone_macedonian.pdf (9 July 2014).

Friedman, Victor A. 2006. West Rumelian Turkish in Macedonia and adjacent areas. In *Turkic Languages in Contact* (Turcologica 61), 27–45. Wiesbaden: Harassowitz.

Friedman, Victor A. 2008. Balkan object reduplication in areal and dialectological perspective. In Dalina Kallulli & Liliane Tasmowski (eds.), *Clitic doubling in the Balkan languages* (Linguistik aktuell vol. 130), 35–63. Amsterdam: Benjamins.

Friedman, Victor A. 2011. The Balkan Languages and Balkan Linguistics. *Annual Review of Anthropology* 40. 275–291.

Friedman, Victor A. & Brian D. Joseph. 2014. Lessons from Judezmo about the Balkan Sprachbund and contact linguistics. *International Journal of the Sociology of Language* 2014(226). 3–23.

Gălăbov, Igor. 1951. Za člena v bălgarski ezik. *Izvestija na Narodnija muzej* 2. 171–227.

Gălăbov, Igor. 1952. [Review of] Josef Kurz: Kăm văprosa za člena v slavjanskite ezici s osoben ogled kăm starobălgarski ezik. *Izvestija na instituta za bălgarski ezik* 2. 356–371.

Galton, Herbert. 1967. The evolution of Bulgarian syntax – A phenomenological study of "analytism". *Linguistique balkanique* 12. 45–99.

Gebert, Lucyna. 2009. Information structure in Slavic languages. In Lunella Mereu (ed.), *Information Structure and its Interfaces*, 307–324. Berlin: Mouton de Gruyter.

Georgiev, Vladimir. 1952. Opit za periodizacija na istorijata na bălgarskija ezik. *Izvestija na instituta za bălgarski ezik* 2. 71–116.

Gołąb, Zbigniew. 1984. *The Arumanian dialect of Kruševo in SR Macedonia, SFR Yugoslavia.* Skopje: Macedonian Academy of Sciences and Arts.

Gołąb, Zbigniew. 1997. The ethnic background and internal linguistic mechanism of the so-called Balkanization of Macedonian. *Balkanistica* 10. 13–19.

Golstein, Bernard. 1999. *Grammaire du turc: Ouvrage pratique à l'usage des francophones.* Paris: l'Harmattan.

Gooskens, Charlotte & Nanna Haug Hilton. 2013. The effect of social factors on the comprehension of a closely related language. In Jani-Matti Tirkkonen & Esa Anttikoski (eds.), *Proceedings of the 24th Scandinavian Conference of Linguistics*, 201–210. Joensuu: University of Eastern Finland.

Granqvist, Kimmo. 2011. *Lyhyt Suomen romanikielen kielioppi.* http://scripta.kotus.fi/www/verkkojulkaisut/julk24/lyhyt_romanikielen_kielioppi.pdf (25 May 2014).

Gruev, Joakim. 1983 [1858]. *Osnova za bălgarska gramatika.* Sofia: Nauka i izkustvo.

Grünthal, Riho. 2007. Morphological change and the influence of language contacts in Estonian. In Hans Fix (ed.), *Beiträge zur Morphologie. Germanisch, baltisch, ostseefinnisch*, 403–432. Odense: Press of Southern Denmark.

Gyllin, Roger. 1991. *The genesis of the modern Bulgarian literary language.* (Acta Universitatis Upsaliensis – Studia Slavica Upsaliensia 30). Uppsala: Uppsala University.

Hakulinen, Lauri. 1979. *Suomen kielen rakenne ja kehitys.* 4th edn. Helsinki: Otava.

Halla-aho, Hilla. 2009. *The non-literary Latin letters: A study of their syntax and pragmatics.* (Commentationes Humanarum Litterarum 124). Helsinki: Societas Scientiarum Fennica.

Halla-aho, Jussi. 2006a. *Muinaiskirkkoslaavin käsikirja.* http://www.halla-aho.com/ocs/ (20 May 2014).

Halla-aho, Jussi. 2006b. *Problems of Proto-Slavic historical nominal morphology on the basis of old church Slavic.* (Slavica Helsingiensia 26). Helsinki: University of Helsinki, Department of Slavonic and Baltic Languages and Literatures.

Hamp, Eric P. 1977. On Some Questions of Areal Linguistics. Paper presented at Proceedings of the 3rd Annual Meeting of the Berkeley Linguistic Society, Berkeley, 19–21 February 1977.

Hamp, Eric P. 1989. Yugoslavia: A crossroads of Sprachbünde. *Zeitschrift für Balkanologie* 25(1). 44–47.

Hancock, Ian. 1995. *A handbook of Vlax Romani.* Columbus (Ohio): Slavica.

Haspelmath, Martin. 1998. How young is Standard Average European? *Language Sciences* 20(3). 271–287.

Haspelmath, Martin. 1999a. External possession in a European areal perspective. *Typological Studies in Language* 39. 109–136.

Haspelmath, Martin. 1999b. External possession in a European areal perspective. In Doris L. Payne & Immanuel Barshi (eds.), *External possession*, 109–134. Amsterdam: John Benjamins.

Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, vol. 1, *1492–1510*. Berlin: De Gruyter.

Haspelmath, Martin. 2010. Framework-free grammatical theory. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 341–365. Oxford: Oxford University Press.

Haspelmath, Martin, Matthew S. Dryer, David Gil, & Bernard Comrie (eds.). 2013. *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. http://wals.info (10 August 2014).

Hawkins, John A. 2003. Efficiency and complexity in grammars: Three general principles. In John Moore & Maria Polinsky (eds.), *The nature of explanation in linguistic theory*. Stanford: CSLI Publications.

Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization.* Cambridge: Cambridge University Press.

Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change.* (Cambridge approaches to language contact). New York: Cambridge University Press.

Heine, Bernd & Tania Kuteva. 2006. *The changing languages of Europe.* (Oxford linguistics). Oxford: Oxford University Press.

Hetzer, Armin. 1978. *Lehrbuch der vereinheitlichten albanischen Schriftsprache.* Hamburg: H. Buske.

Hinrichs, Uwe. 2004. Ist das Bulgarische kreolisiertes Altbulgarisch? In Uwe Hinrichs (ed.), *Die europäischen Sprachen auf dem Wege zum analytischen Sprachtyp*, 231–242. Wiesbaden: Otto Harrassowitz Verlag.

Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.

Horrocks, Geoffrey C. 2010. *Greek: A history of the language and its speakers.* Oxford: Wiley-Blackwell.

Humbert, Jean. 1930. *Le disparition du datif en grec.* (Collection linguistique publiée par la Société de linguistique de Paris 33). Paris: Société de linguistique de Paris.

Hupchick, Dennis P. & Harold E. Cox. 2001. *The Palgrave concise historical atlas of the Balkans.* New York: Palgrave.

Iggesen, Oliver A. 2005. *Case-asymmetry: A world-wide typological study on lexeme-class-dependent deviations in morphological case inventories.* (Lincom studies in language typology 9). Munich: Lincom Europa.

Iggesen, Oliver A. 2013. Number of Cases - Chapter 49. In Martin Haspelmath, Matthew S. Dryer, David Gil, & Bernard Comrie (eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. http://wals.info/feature/49 (10 August 2014).

Ivanova, Ana. 1967. *Trojanski damaskin: Bălgarski pametnik ot XVII vek.* Sofia: Izdatelstvo na Bălgarska Akademija na naukite.

Ivanova-Mirčeva, Dora & Ivan Haralampiev. 1999. *Istorija na bălgarskija ezik.* Veliko Tărnovo: Faber.

Jagić, Vatroslav. 1894. [Review of] Lavrov, P.: Obzor zvukovych i formal'nych osobennostej bolgarskogo jazyka. Moskva 1893. *Archiv für slavische Philologie* 16.

Jakobson, Roman. 1936. Beitrag zur allgemeinen Kasuslehre: Gesamtbedeutungen der russischen Kasus. *Travaux du Cercle Linguistique de Prague* 6. 240–287.

Janda, Laura A. 1996. *Back from the brink: A study on how relic forms in languages serve as source material for analogical extension.* Munich: Lincom Europa.

Johanson, Lars. 2009. Case and contact linguistics. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 494–501. Oxford: Oxford University Press.

Joseph, Brian D. 2006. On projecting variation in the proto-langauge, with particular attention to Germanic evidence and some thoughts on "drift". In Thomas D. Cravens (ed.), *Variation and Reconstruction*, 103–118. Amsterdam: John Benjamins Publishing Co.

Joseph, Brian D. 2008. On the need for history in doing Balkan linguistics. Lecture at LACITO, Paris, 1 October 2008. http://lacito.vjf.cnrs.fr/ documents/ pdf/ BrianJoseph2008.pdf. (12 November 2012.)

Joseph, Brian D. 2010. Language contact in the Balkans. In Raymond Hickey (ed.), *The handbook of language contact*, 618–633. Malden, MA: Wiley-Blackwell.

Joseph, Brian D. 2013. Demystifying drift. In Martine Robbeets & Hubert Cuyckens (eds.), *Shared grammaticalization: With a special focus on the Transeurasian languages*, 43–66. Amsterdam: John Benjamins.

Joseph, Brian D. 2014. What counts as (an instance of) grammaticalization? *Folia Linguistica* 48(2). 361–384.

Jully, Dželjal' [Xhelal Ylli] & Andrej N. Sobolev. 2002. *Albanskij toskskij govor sela Lešnja (kraina Skrapar).* (Materialien zum Südosteuropasprachatlas, Band 2). Marburg an der Lahn: Biblion.

Juvonen, Päivi. 2000. *Grammaticalizing the definite article: A study of adnominal determiners in a genre of spoken Finnish.* Stockholm: Department of Linguistics, University of Stockholm.

K[opitar, Jernej]. 1829. Albanische, walakische u. bulgarische Sprache. *Jahrbücher der Literatur (Wien)* 46. 59–106.

Kabasanov, Stajko. 1952. *Govorăt na s. Momčilovci, Smoljansko.* 4th edn. (Izvestija na Instituta za bălgarski ezik "Prof. Ljubomir Andrejčin"). Sofia: Institut za bălgarski ezik.

Kallulli, Dalina & Liliane Tasmowski (eds.). 2008. *Clitic doubling in the Balkan languages* (Linguistik aktuell 130). Amsterdam: Benjamins.

Kämmerer, Horst. 1977. *Untersuchungen zur Flexion der Substantive in der bulgarischen Schriftsprache des 14./15. Jahrhunderts.* (Slavistische Arbeiten 3). Munich: Dr. Dr. Rudolf Trofenik.

Kasatkina, Rozalija F. 2007. Psevdoartikl' i predartikl' v severnorusskih govorah. In Juhan i Nuorluoto (ed.), *Topics on the Ethnic, Linguistic and Cultural Making of the Russian North* (Slavica Helsingiensia 32), 101–110. Helsinki: Department of Slavonic and Baltic Languages and Literatures.

Katičić, Radoslav. 1976. *Ancient languages of the Balkans, Part one.* (Trends in linguistics, 4–5). The Hague: Mouton.

Klausenburger, Jürgen. 2000. *Grammaticalization: Studies in Latin and Romance morphosyntax.* (Amsterdam studies in the theory and history of linguistic science 193). Amsterdam: John Benjamins Publishing Company.

Klenin, Emily. 1987. Morphological coding, syntactic change, and the modes of historical attestation: The genitive-accusative in Old Church Slavonic and Medieval East Slavic. *The Slavic and East European Journal* 31(3). 404–419.

Koneski, Blaže. 1965. *Istorija na makedonskiot jazik.* Skopje: "Kočo Racin".

Koneski, Blaže. 1983. *A historical phonology of the Macedonian language.* Heidelberg: Winter.

König, Ekkehard & Martin Haspelmath. 1997. Les contructions à possesseur externe dans les langues d'Europe. In Jack Feuillet (ed.), *Actance et valence dans les langues de l'Europe*, 525–606. Berlin: Mouton de Gruyter.

König, Christa. 2011. The grammaticalization of adpositions and case marking. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 511–521. London: Oxford University Press.

Koptjevskaja-Tamm, Maria. 2000. Romani genitives in cross-linguistic perspective. In Viktor Elšik & Yaron Matras (eds.), *Grammatical Relations in Romani*, 123–150. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Koptjevskaja-Tamm, Maria. 2009. “A lot of grammar with a good portion of lexicon”: Towards a typology of partitive and pseudopartitive nominal constructions. In Johannes Helmbrecht, Yoko Nishina, Yong-Min Shin, Stavros Skopeteas, & Elisabeth Verhoeven (eds.), *Form and function in language research: Papers in honour of Christian Lehmann*. New York: Mouton de Gruyter.

Kottum, Steinar E. 1983. In defense of the vocative: The case of Modern Polish. *Scando-Slavica* 29(1). 135–142.

Kovačec, August. 1962. Notes sur des formes de cas en istroroumain. *Studia romanica et anglica zagrabiensia* 13–14. 75–84.

Kulikov, Leonid. 2009. Evolution of case systems. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 439–457. Oxford: Oxford University Press.

Kurz, Josef. 1938. K otázce členu v jazycích slovanských, se zvláštním zřetelem k staroslověnštině [1st part]. *Byzantinoslavica* 7. 212–340.

Kurz, Josef. 1946. K otázce členu v jazycích slovanských, se zvláštním zřetelem k staroslověnštině [2nd part]. *Byzantinoslavica* 8. 172–288.

Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflection.* Utrecht: LOT.

Labroska, Veselinka. 2008. *Kičevskiot govor.* Skopje: Institut za makedonskiot jazik “Krste Misirkov”.

Leafgren, John. 2002. *Degrees of explicitness: Information structure and the packaging of Bulgarian subjects and objects.* Amsterdam: John Benjamins Publishig Company.

Lee, Ronald. 2005. *Learn Romani: Das-dúma Rromanes.* Hatfield: University of Hertfordshire Press.

Lettenbauer, Wilhelm. 1953. Synthetische und analytische Flexion in den slavischen Sprachen. In Erwin Koschmieder & Alois Schmaus (eds.), *Münchener Beiträge zur Slavenkunde: Festgabe für Paul Diels*, 149–162. Munich: Osteuropa-Institut.

Levander, Lars. 1909. *Älvdalsmålet i Dalarna: ordböjning och syntax.* Uppsala: University of Uppsala Doctoral dissertation.

Lindstedt, Jouko. 2000. Linguistic balkanization: Contact-induced change by mutual reinforcement. In Dicky Gilbers, John Nerbonne, & Jos Schaeken (eds.), *Languages in contact* (Studies in Slavic and general linguistics 28), 231–246. Amsterdam: Rodopi.

Lindstedt, Jouko. 2002. Is there a Balkan verb system? *Balkanistica* 15. 305–318.

Lindstedt, Jouko, Ljudmil Spasov, & Juhani Nuorluoto (eds.). 2008. *The Konikovo Gospel – Konikovsko evangelie: Bibl. Patr. Alex. 268* (Commentationes humanarum litterarum 125). Helsinki: Societas Scientiarum Fennica.

Ljubenova, Elena. 2001. Aleksandăr Teodorov-Balan i teorijata za padeža. Paper presented at Ezikăt i literaturata: Sredstva za (ne)razbiratelstvo, Sofia, 21–22 May 2001. http://georgesg.info/belb/personal/ljubenova/Doklad5.htm (11 March 2013).

Lombard, Alf. 1973. *Rumänsk grammatik.* Lund: Gleerup.

Lopašov, JU A. 1978. *Mestoimennye povtory dopolnenija v balkanskih jazykah.* Leningrad: Nauka.

Lunt, Horace G. 1965. On the loss of declension in Bulgarian and Macedonian. *Die Welt der Slaven* 10. 305–312.

Lunt, Horace G. 2001. *Old Church Slavonic grammar.* Berlin: Mouton de Gruyter.

Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLoS ONE* 5(1), http://www.scopus.com/record/display.url?eid=2-s2.0-77649307637&origin=inward. (6 June 2013.)

Maitz, Péter & Attila Németh. 2014. Language contact and morphosyntactic complexity: Evidence from German. *Journal of Germanic Linguistics* 26(1). 1–29.

Malchukov, Andrej & Andrew Spencer. 2009. Introduction. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 1–12. Oxford: Oxford University Press.

Manolessou, Io & Geoffrey Horrocks. 2007. The development of the definite article in Greek. In *Studies in Greek Linguistics 27*. Thessaloniki: Aristotle University of Thessaloniki.

Markoviḱ, Marjan. 2007. *Aromanskiot i makedonskiot govor od ohridsko-struškiot region vo balkanski kontekst.* Skopje: Makedonska akademija na naukite i umetnostite.

Masica, Colin P. 2001. The definition and significance of linguistic areas: Methods, pitfalls, and possibilities (with special reference to the validity of South Asia as a linguistic area). In Rajendra Singh, Peri Bhaskararao, & Karumuri Venkata Subbararao (eds.), *The yearbook of South Asian languages and linguistics*, 205–267. New Delhi: Sage.

Matasović, Ranko. 2008. *Poredbenopovijesna gramatika hrvatskoga jezika.* Zagreb: Matica Hrvatska.

Matras, Yaron. 2002. *Romani: a linguistic introduction.* Cambridge: Cambridge University Press.

Mazon, André & André Vaillant. 1938. *L'évangéliaire de Kulakia, un parler slave du Bas-Vardar.* Paris: Droz.

McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2). 125–166.

Mereu, Lunella. 2009. Universals of information structure. In Lunella Mereu (ed.), *Information Structure and its Interfaces*, 75–101. Berlin: Mouton de Gruyter.

Mertyris, Dionysios. 2009. Case syncretism in the personal pronouns of Medieval Greek: the loss of the genitive plural. Paper presented at Proceedings of the 9th International Conference on Greek Linguistics, University of Chicago, Chicago, 29–31 October 2009. http://www.ling.ohio-state.edu/ICGL/proceedings/ (11 November 2013).

Mertyris, Dionysios. 2014. *The loss of the genitive in Greek: A diachronic and dialectological analysis.* https://www.academia.edu/10629756/PhD_Dissertation_-_The_loss_of_the_genitive_in_the_diachrony_of_Greek (21 April 2015).

Meyer, Karl H. 1920. *Der Untergang der Deklination im Bulgarischen.* (Slavica 3). Heidelberg: Carl Winter.

Meyer-Lübke, Wilhelm. 1894. *Grammatik der romanischen Sprachen: Formenlehre,* vol. 2. Leipzig: O. R. Reisland.

Miestamo, Matti, Kaius Sinnemäki, & Fred Karlsson (eds.). 2007. *Language complexity; Typology, contact, change* (Studies in language companion series 94). Amsterdam: Benjamins.

Miklosich, Franz. 1856. *Die Sprache der Bulgaren in Siebenburgen.* Vienna: Der Kaiserlich-Königliche Hof- und Staatsdruckerei.

Miletič, Ljubomir. 1890. Staroto sklonenie vъ dnešnitě bъlgarski narěčija. In *Zbornikъ za narodna umotvorenija, nauka i knižnina, Kniga II*, 226–268. Sofia: Ministerstvoto na narodnoto prosvěštenie.

Miletič, Ljubomir. 1905. Interesen srednobălgarski pametnik ot XIII v. *Periodičesko spisanie na Bălgarskoto knižovno družestvo v Sofija* 66. 95–112.

Miletič, Ljubomir. 1925. Zur Geschichte der bulgarischen Nominalflexion. [Review of] Karl H. Meyer: Der Untergang der Deklination im Bulgarischen. *Archiv für slavische Philologie* 39. 267–278.

Miletič, Ljubomir. 1926. *Sedmogradskitě bălgari i tehnijat ezik.* (Spisanie na Bălgarskata akademija na naukitě 33). Sofia: Bălgarskata akademija na naukitě.

Miletič, Ljubomir. 1987 [1896–1900]. *Izsledovanija za bălgarite v Sedmigradsko i Banat.* Sofia: Nauka i iskustvo.

Minčeva, Angelina. 1964. *Razvoj na datelnija pritežatelen padež v bălgarskija ezik.* Sofia: Bălgarska akademija na naukite.

Mirčev, Kiril. 1936. *Nevrokopskijat govor.* 32nd edn. (Godišnik na Sofijskija universitet, Istoriko-filologičeski fakultet). Sofia: Sofijski universitet.

Mirčev, Kiril. 1957. Po văprosa za naj-rannite primeri ot analitičen datelen padež v bălgarskite pametnici. In *Ezikovedski izsledvanija v čest na akademik Stefan Mladenov*, 37–46. Sofia: Bălgarskata akademija na naukite.

Mirčev, Kiril. 1966. Za hronologijata na osnovnite balkanizmi v bălgarski ezik. *Bălgarski ezik* 16(4). 291–293.

Mirčev, Kiril. 1978. *Istoričeska gramatika na bălgarskija ezik.* 3rd edn. Sofia: Nauka i izkustvo.

Mišeska Tomić, Olga. 2006. *Balkan Sprachbund morpho-syntactic features.* (Studies in natural language and linguistic theory; 67). Dordrecht: Springer.

Mišeska Tomić, Olga. 2009. Clitic and non-clitic possessive pronouns in Macedonian and Bulgarian. In Mila Dimitrova-Vulchanova & Olga Mišeska Tomić (eds.), *Investigations in the Bulgarian and Macedonian nominal expression*, 95–120. Trondheim: Tapir academic books.

Mladenov, Stefan. 1979. *Istorija na bălgarskija ezik.* Sofia: Bălgarskata akademija na naukite.

Mladenova, Olga M. 2007. *Definiteness in Bulgarian: Modelling the processes of language change.* Berlin: Mouton de Gruyter.

Moravcsik, Edith A. 2011. Explaining language universals. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 69–89. New York, NY: Oxford University Press.

Næss, Åshild. 2008. Varieties of dative. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*, 572–580. Oxford: Oxford University Press.

Narrog, Heiko & Shinya Ito. 2007. Re-constructing semantic maps: The comitative-instrumental area. *Sprachtypologie und Universalienforschung* 60(4). 273–292.

Naylor, Kenneth. 1981. The dative/locative opposition in the Balkan languages. *Folia Slavica* 4(2–3). 340–345.

Newman, John. 1998. Recipients and "give" constructions. In William van Belle & Willy van Langendonck (eds.), *The dative,* vol. 2, *Theoretical and contrastive studies* (Case and grammatical relations across languages, 3), 1–28. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Nichols, Johanna. 1992. *Linguistic diversity in space and time.* Chicago: The University of Chicago Press.

Nichols, Johanna. 2007. What, if anything, is typology? *Linguistic Typology* 11(1). 231–238.

Nichols, Johanna. 2008. Does language contact simplify grammars? Lecture at Deutsche Gesellschaft für Sprachwissenschaft, Bamberg, February 2008. http://www-personal.umich.edu/~thomason/temp/simple2.pdf. (17 May 2014.)

Nichols, Johanna & Balthasar Bickel. 2008. Feature 59A: Possessive classification. In Martin Haspelmath, Matthew S. Dryer, David Gil, & Bernard Comrie (eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. http://wals.info; (2013-11-22).

Nichols, Lynn. 2001. The syntactic basis of referential hierarchy phenomena: Clues from languages with and without morphological case. *Lingua* 111(4). 515–537.

Nomachi, Motoki & Bernd Heine. 2011. On predicting contact-induced grammatical change: Evidence from Slavic languages. *Journal of Historical Linguistics* 1(1). 48–76.

Nuorluoto, Juhani. 2012a. Towards Sound change in Slavic: The Rise and Decline of the Correlation in Timbre. In Ilona Janyšková & Helena Karlíková (eds.), *Theory and empiricism in Slavonic diachronic linguistics*, 449–464. Prague: Lidové noviny.

Nuorluoto, Juhani. 2012b. *De slaviska standardspråkens framväxt: Idéhistoriska och språksociologiska kontexter och koncept.* Uppsala: Institutionen för moderna språk, Uppsala universitet.

Orel, Vladimir E. 2000. *A concise historical grammar of the Albanian language: Reconstruction of Proto-Albanian.* Leiden: Brill.

Pancheva, Roumyana. 2004. Balkan possessive clitics: The problem of case and category. *Balkan syntax and semantics* 67. 175–223.

Pedersen, Holger. 1894. Bidrag til den albanesiske sproghistorie. In *Festskrift til Vilhelm Thomsen fra Disciple: Udg. i anledning af hans femogtyveårige doktorjubilaeum*, 368. København: Gyldendalske Boghandel.

Petroska, Elena. 2008. Informaciskata struktura i udvojuvanje na objektot vo makedonskiot jazik vo balkanski kontekst. In , 127–133. Skopje: Univerzitet Sv. Kiril i Metodij.

Pompeo, Flavia. 2012. Il sincretismo di genitivo e dativo nella lega balcanica. Una convergenza multipla? *Studi italiani di linguistica teorica e applicata* 41(3). 531–544.

Popova, Antoaneta. 2008. *Za analitizma na bălgarskija ezik.* Sofia: ETO.

Popović, Ivan. 1960. *Geschichte der serbokroatischen Sprache.* Wiesbaden: O. Harrassowitz.

Prendergast, Eric. 2012. Pragmatic dimensions of Macedonian object reduplication. *Balkanistica* 25(2). 153–164.

Qvonje, Jørn Ivar. 1979. Die fehlende Unterscheidung ubi/quo und der Zusammenfall von Genitiv und Dativ. *Zeitschrift für Balkanologie* 15. 134–153.

Qvonje, Jørn Ivar. 1980. Die Grammatikalizierung der Präposition na im Bulgarischen. *Folia Linguistica Historica* 1(2). 317–351.

Roberts, Seán & James Winters. 2012. Social structure and language structure: The new nomothetic approach. *Psychology of Language and Communication* 16(2). 89–112.

Rosetti, Al[exandru]. 1968. *Istoria limbii române: De la origini pînă în secolul al XVII-lea.* Bucharest: Editura pentru literatură.

Ruge, Hans. 1969. *Zur Entstehung der neugriechischen Substantivdeklination.* Stockholm: Almqvist & Wiksell.

Rusek, Jerzy. 1964. *Deklinacja i użycie przypadków w triodzie Chłudowa: Studium nad rozwojem analityzmu w języku bułgarskim.* (Prace komisji słowianoznawstwa 5). Wrocław: Polska akademija nauk.

Sandfeld, Kristian. 1926. *Balkanfilologien: En oversigt over dens resultater og problemer.* Copenhagen: Københavns Universitet.

Sapir, Edward. 1979 [1921]. *Language: An introduction to the study of speech.* London, Toronto, Sydney, New York: Granada.

Schaller, Helmut Wilhelm. 1975. *Die Balkansprachen: Eine Einführung in die Balkanphilologie.* Heidelberg: Winter.

Schenker, Alexander M. 1996. *The dawn of Slavic: An introduction to Slavic philology.* New Haven: Yale University Press.

Seiler, Guido. 2011. Non-Standard Average European: How robust is the European Sprachbund beyond the codified standard varieties? Paper presented at Variation and Typology: New Trends in Syntactic Research, Helsinki, 25–27 August 2011.

Sherzer, Joel. 1976. *An areal-typological study of American Indian languages north of Mexico.* Amsterdam: North-Holland.

Shosted, Ryan K. 2006. Correlating complexity: A typological approach. *Linguistic Typology* 10(1). 1–40.

Šin, Ute. 1996. *Padežnata sistema v srednobălgarskija prevod na Manasievata hronika.* Sofia: Institut Lingva.

Sinnemäki, Kaius. 2008. Complexity trade-offs in core argument marking. In Matti Miestamo, Kaius Sinnemäki, & Fred Karlsson (eds.), *Language complexity: Typology, contact, change* (Studies in language companion series 94), 67–88. Amsterdam: Benjamins.

Sinnemäki, Kaius. 2011. *Language universals and linguistic complexity: Three case studies in core argument marking.* Helsinki: University of Helsinki doctoral dissertation.

Sinnemäki, Kaius. 2014. A typological perspective on Differential Object Marking. *Linguistics* 52(2). 281–313.

Sköld, Hannes. 1923. Linguistic gleanings: 7 The making of modern Bulgarian. In *Lunds universitets årsberättelse 1922–1923*, 18–27. Lund: Lunds universitet.

Sobolev, Andrej N. 1991. K istorii utraty balkanoslavjanskogo sklonenija. *Zbornik Matice srpske za filologiju i lingvistiku* 34(2). 7–41.

Solta, Georg Renatus. 1980. *Einführung in die Balkanlinguistik mit besonderer Berücksichtigung des Substrats und des Balkanlateinischen.* Darmstadt: Wissenschaftliche Buchgesellschaft.

Spencer, Andrew. 2007. The possessum-agreement construction or ‘Does Albanian have a genitive case?’. *Essex Research Reports in Linguistics* 53, http://repository.essex.ac.uk/114/1/ERRL53.pdf. (28 November 2013.)

Spencer, Andrew. 2008. Does Hungarian have a case system? In Greville G. Corbett & Michael Noonan (eds.), *Case and grammatical relations*, 35–56. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Steinke, Klaus. 1968. *Studien über den Verfall der bulgarischen Deklination.* Munich: Otto Sagner.

Stojkov, St[ojko]. (ed.). 1966. *Bălgarski dialekten atlas II: Severoistočna Bălgarija*. Sofia: Izdatelstvo na Bălgarska Akademija na naukite.

Stojkov, Stojko. 1968. Nominale Kasusformen in der bulgarischen Sprache. *Die Welt der Slaven* 13. 28–46.

Stolk, Joanne. 2013. A diachronic account of dative by genitive replacement in papyrological Greek. Paper presented at ICHL21: International Conference on Historical Linguistics, Oslo, 5–9 August 2013.

Stölting, Wilfried. 1970. *Beiträge zur Geschichte des Artikels im Bulgarischen.* Munich: Verlag Otto Sagner.

Stolz, Thomas, Sonja Kettler, Sonja Stroh, & Aina Urdze. 2008. *Split Possession: An areal-linguistic study of the alienability correlation and relatd phenomena in the languages of Europe.* Amsterdam: John Benjamins.

Svane, Gunnar. 1962. O sintaksičeskom primenenii bolgarskogo člena v XVIII veke. B. Dobrejšovo evangelie. *Scando-Slavica* 8. 224–238.

Svane, Gunnar. 1961. O sintaksičeskom primenenii bolgarskogo člena v XVIII veke. A. Čergedskie teksty. *Scando-Slavica* 7. 233–251.

Thomason, Olga. 2006. *Prepositional Systems in Biblical Greek, Gothic, Classical Armenian, and Old Church Slavic.* (University of Georgia PhD thesis: Athens, GA).

Thomason, Sarah Grey. 2001. *Language contact: An introduction.* Edinburgh: Edinburgh University Press.

Thomason, Sarah Grey. 2008. Social and linguistic factors as predictors of contact-induced change. *Journal of Language Contact* 2(1). 42–56.

Thomason, Sarah Grey. 2010. Contact Explanations in Linguistics. In Raymond Hickey (ed.), *The handbook of language contact*. Malden, Massachusetts: Wiley-Blackwell. https://getd.libs.uga.edu/pdfs/thomason_olga_a_200605_phd.pdf (23 April 2013).

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics.* Berkeley, CA: University of California Press.

Tiraspol'skij, G[ennadij] I[saakovič]. 1980. K voprosu ob uslovijah i pričinah utraty bolgarskogo substantivnogo sklonenija. *Sovetskoe slavjanovedenie* 4.

Topolinjska, Zuzana. 2006. Trojniot člen: Da ili ne? *Južnoslovenski filolog* 62. 7–15.

Tosco, Mauro. 2008. What to do when you are unhappy with language areas but you do not want to quit. *Journal of Language Contact* 2(1). 112–123.

Trubetzkoy, Nikolai Sergeevich. 1928. [Proposition 16]. In *Actes du premier congrès international de linguistes*, 17–18.

Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity.* Oxford: Oxford University Press.

Väänänen, Veikko. 1981. *Introduction au latin vulgaire.* 3rd edn. Paris: Klincksieck.

Vaillant, André. 1958. *Grammaire comparée des langues slaves*, vol. 2, *Morphologie. 1ère partie: Flexion nominale.* Paris: Klincksieck.

Večerka, Radoslav. 1984. *Staroslověnština.* Prague: SPN.

Velcheva, Boryana. 2014. On Bulgarian Historical Dialectology. *Scripta & e-Scripta*(13). 57–61.

Verschik, Anna. 2012. Practising receptive multilingualism: Estonian–Finnish communication in Tallinn. *International Journal of Bilingualism* 16(3). 265–286.

Vidoeski, Božidar. 2005. *Dialects of Macedonian.* Bloomington, IN: Slavica Publishers.

Zamboni, Alberto. 2000. *Alle origini dell'italiano: Dinamiche e tipologie della transizione dal latino.* Rome: Carocci.

Zegrean, Iulia-Georgiana. 2012. *Balkan Romance: Aspects on the syntax of Istro-Romanian.* http://dspace.unive.it/bitstream/handle/10579/1166/Zegrean.pdf?sequence=1 (22 April 2014).

Zymberi, Isa. 1991. *Colloquial Albanian.* London: Routledge.