

Ss. Cyril and Methodius University, Skopje
Blaže Koneski Faculty of Philology

JOURNAL OF CONTEMPORARY PHILOLOGY

Skopje, June 2025

Publisher:

Ss Cyril and Methodius University, Skopje
Blaže Koneski Faculty of Philology

Editor-in-chief at Blaže Koneski Faculty of Philology:

Vladimir Martinovski, Dean

Editorial Board

Elisaveta Popovska, editor-in-chief
Sonja Kitanovska-Kimovska
Anastazija Kirkova-Naskova
Iskra Tasevska Hadži Boškova
Kristina N. Nikolovska

International Editorial Board

Victor Friedman, University of Chicago, USA
Ewa Waniek-Klimczak, University of Łódź, Poland
Slobodanka Dimova, University of Copenhagen, Denmark
Mohammad Dabir-Moghaddam, University of Tehran, Iran
Irena Sawicka, Nicolaus Copernicus University in Toruń, Poland
Jim Hlavac, Monash University, Australia
Alice Henderson, University of Grenoble Alpes, France
Roman Krivko, University of Vienna, Austria
Alla Sheshken, Moscow State University, Russia
Motoki Nomachi, Hokkaido University, Japan
Grace Fielder, University of Arizona, USA
Christina Kramer, University of Toronto, Canada
Zlatko Kramarić, University of Osijek, Croatia
Wolfgang Motsch, University of Mannheim, Germany
Elena Ivanova, University of St Petersburg, Russia
María Isabel González-Rey, University of Santiago de Compostela, Spain
Mladen Uhlík, University of Ljubljana, Slovenia
Daniela Dinca, University of Craiova, Romania
Blake Smith, Huntington Library, California, USA
Iwona Łuczaków, University of Wrocław, Poland
Jean-Marc Defays, University of Liège, Belgium
Natalija Boronnikova, Perm State University, Russia
Mila Samardžić, University of Belgrade, Serbia
Vanna Zaccaro, University of Bari, Italy
Marjan Marković, Ss Cyril and Methodius University, R. Macedonia
Slavica Srbinovska, Ss Cyril and Methodius University, R. Macedonia
Eleni Bužarovska, Ss Cyril and Methodius University, R. Macedonia

Proofreading:

Iskra Tasevska Hadži Boškova

Printed by:

Mar-Saž, Skopje

Print run:

100

ISSN 2545-4765 (print)

ISSN 2545-4773 (electronic)

© **Blaže Koneski Faculty of Philology, Skopje, 2024**



Republic of North Macedonia

Ministry of Culture and Tourism

This issue of the journal is published with the support of
The Ministry of Culture of the Republic of North Macedonia

Универзитет „Св. Кирил и Методиј“
Филолошки факултет „Блаже Конески“ – Скопје

СОВРЕМЕНА ФИЛОЛОГИЈА

Скопје, јуни 2025

Издавач:

Универзитет „Св. Кирил и Методиј“, Скопје
Филолошки факултет „Блаже Конески“

Уредник на издавачката дејност на Филолошки факултет „Блаже Конески“:

Владимир Мартиновски, декан

Редакциски одбор

Елисавета Поповска, главен уредник
Соња Китановска-Кимовска
Анастасија Киркова-Наскова
Искра Тасевска Хаџи Бошкова
Кристина Н. Николовска

Меѓународен редакциски одбор

Виктор Фридман, Универзитет во Чикаго, САД
Ева Ваниек-Климчак, Универзитет во Лоѓ, Полска
Слободанка Димова, Универзитет во Копенхаген, Данска
Мохамад Дабир-Могадам, Универзитет во Техеран, Иран
Ирена Савицка, Универзитет „Никола Коперник“ во Торун, Полска
Џим Хлавац, Универзитет Монаш, Австралија
Алис Хендерсон, Универзитет во Гренобл, Франција
Роман Кривко, Универзитет во Виена, Австрија
Ала Шешкен, Московски државен универзитет, Русија
Мотоки Номачи, Универзитет Хокаидо, Јапонија
Грејс Филдер, Универзитет во Аризона, САД
Кристина Крамер, Универзитет во Торонто, Канада
Златко Крамарик, Универзитет во Осиек, Хрватска
Волфганг Моч, Универзитет во Манхајм, Германија
Елена Иванова, Универзитет во Санкт Петербург, Русија
Марија Изабел Гонзалес, Универзитет во Сантијаго де Компостела, Шпанија
Младен Ухлик, Универзитет во Љубљана, Словенија
Даниела Динка, Универзитет во Крајова, Романија
Блејк Смит, Библиотека Хантингтон во Калифорнија, САД
Ивона Лучков, Универзитет во Вроцлав, Полска
Жан Марк Дефаји, Универзитет во Лиеж, Белгија
Наталија Бороникова, Универзитет во Перм, Русија
Мила Самарциќ, Универзитет во Белград, Србија
Вана Сакаро, Универзитет во Бари, Италија
Марјан Марковиќ, Универзитет Св. Кирил и Методиј, Скопје, Р. Македонија
Славица Србиновска, Универзитет „Св. Кирил и Методиј“, Скопје, Р. Македонија
Елени Бужаровска, Универзитет „Св. Кирил и Методиј“, Скопје, Р. Македонија

Лектура:

Искра Тасевска Хаџи Бошкова

Компјутерска обработка и печат:

Мар-саж, Скопје

Тираж:

100

ISSN 2545-4765 (печатена верзија)

ISSN 2545-4773 (електронска верзија)

© Филолошки факултет „Блаже Конески“, Скопје, 2024



Republic of North Macedonia

Ministry of Culture and Tourism

Овој број на списанието е објавен со поддршка од
Министерството на култура на Република Северна Македонија

Table of contents

Содржина

Eleni Bužarovska, guest editor

Foreword

Елени Бужаровска

Предговор

7

Elena Ivanova

Syntax of Clitics in Bulgarian Contact Dialects

Елена Иванова

Синтаксата на клитиките во бугарските дијалекти во јазичен контакт

11

Kirill Kozhanov, Victor A. Friedman

New Research on Romani in Eastern North Macedonia: The Maleshevo Dialect

Кирил Кожанов, Виктор А. Фридман

Ново истражување за ромскиот јазик во Источна Северна Македонија: малешевскиот дијалект

25

Maxim Makartsev

Contact-induced changes in the order of modifying adjectives and nouns in Slavic dialects in Albania

Максим Макарцев

Влијанието на јазичниот контакт во редоследот на придавките и на именките во словенските дијалекти во Албанија

41

Maria Morozova, Alexander Rusakov

Labile Verbs in Modern Albanian: a Preliminary Corpus-Based Study

Марија Морозова, Александер Русаков

Лабилните глаголи во современиот албански јазик: корпусно истражување

67

Julian Rentzsch

Event modality in Balkan Turkish: Formal and semantic variation in contact

Part 1. Possibility and Necessity

Јулијан Ренч

Предикациската модалност во балканскиот турски јазик: формална и семантичка варијантност во јазичен контакт

Дел 1. Веројатност и нужност

85**Ирена Савицка**

Акцентот на зборот како балканска карактеристика

Irena Sawicka

Accent as a Balkan Feature

107**Max Wahlström, Don Killian**

The Macedonian and Balkan Slavic Three-Part Article and the Typology of Deictic Articles

Макс Валстром, Дон Килиан

Македонскиот и балканскословенскиот троен член и типологијата на деиктичките членови

115**Olivier Winistörfer, Anastasia Escher, Daria Konior**

Lost in the Balkans: Differential Place Marking in the Aromanian Varieties

Оливер Винисторфер, Анастасија Ешер, Дарија Кониор

Изгубени на Балканот: локативното диференцијално маркирање во ароманските дијалекти

129

FOREWORD

This volume presents a series of contributions that address key morphosyntactic, semantic, phonological, and contact-induced phenomena in the languages and dialects of the Balkans. The eight studies collected here examine diverse linguistic structures within a shared areal and typological context, offering new data and analyses that contribute to our understanding of the Balkan Sprachbund and its internal diversity.

Elena Ivanova focuses on clitic syntax in Bulgarian dialects spoken in Romania and in the dialect of Novo Selo in northwestern Bulgaria. The study identifies syntactic interference from Romanian and dialectal influence from Macedonian settlers. While the clitic order typical of Standard Bulgarian is generally retained, the author shows that in these dialects clitics may occur in sentence-initial position, violating the Tobler-Mussafia law. The author concludes that this deviation is contact-induced, but also notes that restrictions on clitic placement in the donor languages are often not transferred. The results contribute important insights into clitic placement in contact settings.

In their contribution, **Kirill Kozhanov** and **Victor A. Friedman** present new field data on Romani varieties spoken in the Maleshevo region of eastern North Macedonia. In addition to the description of grammatical features, the authors document and contextualize lexical borrowings from local Macedonian dialect. Referred to as Maleshevo-Pirin Romani, these dialects are shown to differ notably from other South Balkan Romani varieties, especially those spoken in Skopje, while sharing features with dialects from southwestern Bulgaria, largely due to close familial ties among local communities. Situated at the intersection of the East-West dialectal divide, these varieties exhibit influence from surrounding Turkish and Macedonian dialects.

Maxim Makartsev investigates word order variation in nominal phrases in South Slavic dialects spoken in Albania. Focusing on the placement of adjectives in relation to nouns, the study documents a shift from the original Adjective–Noun order to the Albanian-influenced Noun–Adjective order. The analysis is based on dialects from Golloborda, Korça, Prespa, and Štokavian communities in Shijak and Myzeqe. Sociolinguistic factors such as age, gender, settlement type, and community compactness have been shown to correlate with the adoption of borrowed word order. The study also considers the potential role of standard Slavic languages in reinforcing the original word order, given the competing pressures of language contact.

Maria Morozova and **Alexander Rusakov** offer a corpus-based investigation of labile verbs in Standard Albanian, using data from the Albanian National Corpus. The study focuses on eight P-labile verbs from different semantic domains. The authors find significant variation among the verbs in terms of their lability. While phasal verbs consistently alternate between transitive and intransitive forms, other verbs exhibit more restricted patterns. The study reveals important distinctions among verb types, including motion verbs, deadjectival color verbs, and verbs of distinction, each with distinct patterns of transitivity alternation. The quantified results highlight the role of verbal semantics in the distribution of lability and contribute to a deeper typological understanding of this phenomenon.

Julian Rentzch investigates the expression of modality of possibility and necessity in Balkan Turkish in Kosovo, North Macedonia, Bulgaria, and Eastern Thrace (Turkey). The modal constructions in these Turkish dialects are compared with their functional equivalents in Modern Standard Turkish and existing variants of Ottoman Turkish. The variation, which is characteristic of different syntactic positions of the examined constructions, pertains to a range of lexical, semantic, and morphosyntactic features. Both infinitival and subjunctive complementation also exhibit variation. The author argues that certain modal structures in the

Turkish dialects of the Balkans have emerged as a result of two processes: the internal development of inherited structures and the influence of contact with neighboring languages.

Irena Sawicka examines two prosodic phenomena in Macedonian dialects: the restriction of stress to the final three syllables of a word and the presence of double accentuation. While both features have been attributed to convergence within the central Balkan area, Sawicka argues that the three-syllable stress limit reflects a more general phonological tendency and is not a distinctively Balkan phenomenon. In contrast, a specific type of double accentuation found in contact zones between Slavic and Greek dialects may constitute a localized areal feature. The analysis questions previously held assumptions about the areal scope of certain prosodic traits.

The article by **Max Wahlström** and **Don Killian** examines the morphosyntactic and semantic features of the tripartite definite articles in the Macedonian language in order to determine their status within the broader typology of definiteness and deixis. Using data from the spoken corpus of Macedonian, the authors demonstrate that there are significant differences in distribution among the proximal, neutral, and distal forms. In other Balkan Slavic languages as well, one member of the tripartite system tends to dominate over the others. The authors also critically examine claims regarding the secondary functions of these articles, arguing that such functions are not essential criteria for articlehood.

The study by **Olivier Winistörfer**, **Anastasia Escher**, and **Daria Konior** addresses Differential Place Marking in Aromanian and related Balkan Romance varieties. Drawing on data from Kruševo, Ohrid, Struga, and Turia/Kranéa, the authors examine the conditions under which locative relations are zero-marked or overtly marked. They find that factors such as the referential status of the toponym (proper vs. common noun) and the perceived distance to the location (proximal vs. distal) play a significant role in determining marking patterns. The study contributes to the deeper understanding of grammatical variation in the Balkan area.

Collectively, the articles in this thematic volume have significantly enriched the field of Balkan linguistics by addressing topics related to language contact. Through typologically oriented, empirically grounded analyses of linguistic phenomena in dialects and languages in contact situations, the volume contributes to a deeper understanding of the Balkan linguistic area.

We extend our thanks to all the contributors for their in-depth research and insightful work which have made this volume possible.

Sincerely,

Eleni Bužarovska

Guest Editor

Professor of Linguistics, Ss. Cyril and Methodius University

Faculty of Philology “Blazhe Koneski”, Skopje

ПРЕДГОВОР

Во овој тематски зборник се претставени осум статии во кои се истражуваат морфосинтаксичките, семантичките и фонолошките својства во јазиците и во дијалектите на Балканот. Анализата на различните јазични структури од типолошки аспект нуди нови сознанија што придонесуваат за нашето подлабоко разбирање на Балканскиот јазичен сојуз и неговата внатрешна разновидност.

Во првата статија, **Елена Иванова** се фокусира на синтаксата на клитиките во бугарските дијалекти што се зборуваат во Романија и во Ново Село во северозападна Бугарија. Анализата покажува дека позицијата на клитиките во овие говори е под влијание на романскиот јазик и дијалектот на македонските доселеници. Во разгледуваните бугарски дијалекти, клитиките можат да заземат иницијална позиција со што се прекршува законот на Тоблер-Мусафија. Иванова заклучува дека отстапувањето во редоследот на клитиките е резултат на јазичен контакт, но исто така забележува дека влијанието на соседните говори не е секогаш доследно. Резултатите од истражувањето нудат нови сознанија за синтаксичките промени во редоследот на клитиките како резултат на јазичен контакт.

Во статијата на **Кирил Кожанов** и **Виктор А. Фридман** се претставени нови теренски податоци за ромските говори во регионот на Малешево во источниот дел на Северна Македонија. Покрај описот на граматичките својства, прикажани се и лексичките заемки од локалниот македонски говор. Овие дијалекти, наречени малешевско-пирински ромски, значително се разликуваат од другите јужнобалкански ромски варијанти, особено од оние што се зборуваат во Скопје. Од друга страна, тие истовремено имаат заеднички карактеристики со дијалектите од југозападна Бугарија, главно поради блиските семејни врски меѓу локалните заедници. Сместени на границата меѓу источната и западната дијалектна поделба, овие говори се под влијание на околните турски и македонски дијалекти.

Максим Макарецв ги истражува варијациите во редоследот на зборовите во именските фрази во јужнословенските дијалекти што се зборуваат во Албанија. Во фокусот на анализата е положбата на придавката во однос на именката во именската синтагма во дијалектите од Голобордо, Корча, Преспа и штокавските заедници во Шијак и Музакија. Резултатите покажуваат дека во овие дијалекти се менува првичниот редослед придавка–именка во именка–придавка под влијание на албанскиот јазик. Социолингвистичките фактори како што се возраста, полот, местото на живеење и компактоста на заедницата корелираат со усвојувањето на позајмениот редослед. Исто така, се разгледува можното влијание на стандардните словенски јазици во зајакнувањето на првичниот редослед и покрај спротивното влијанија на контактот.

Во својата статијата, **Марија Морозова** и **Александер Русаков** ги претставуваат резултатите од корпусната анализа на лабилните глаголи во стандардниот албански јазик. Користејќи податоци од Националниот корпус на Албанија, во фокусот на истражувањето се осум пациентивни лабилни глаголи од различни семантички домени. Авторите откриваат значителна варијација меѓу анализираните глаголи во однос на лабилноста. Утврдено е дека кај фазните глаголи преодната и непреодната форма се употребува наизменично за разлика од глаголите на движење и деадјективните глаголи, кај кои едната форма преовладува над другата. Добиените резултати ја потврдуваат улогата на глаголската семантика во дистрибуцијата на лабилноста со што се продлабочува разбирањето на оваа појава од типолошки аспект.

Јулијан Ренч го истражува изразувањето на модалните значења можност и нужност во балканскиот турски јазик што се зборува во Косово, Северна Македонија, Бугарија и Источна Тракија (Турција). Модалните конструкции во овие турски балкански дијалекти

се споредуваат со нивни функционални еквиваленти во модерниот стандарден турски јазик и постоечките варијанти на Османлискиот турски јазик. Варијантноста, која е карактеристична за разни синтаксички позиции на истражуваните конструкции, се однесува на низа лексички, семантички и морфосинтаксички особини. Инфинитивната и субјунктивната комплементација исто така се одликуваат со варијантност. Авторот смета дека одредени модални структури во турските дијалекти на Балканот настанале како резултат на два процеси: внатрешниот развој на наследените структури и влијанието на контактот со соседните јазици.

Ирена Савицка истражува два прозодиски феномени во македонските дијалекти: ограничувањето на акцентот на последните три слога од збор и појавата на двоен акцент. Иако се сметало дека и двете карактеристики се должат на конвергенција во централниот балкански простор, Савицка тврди дека ограничувањето на акцентот на три слога одразува една поопшта фонолошка тенденција и затоа не е исклучиво балкански феномен. За разлика од тоа, специфичниот тип на двоен акцент, којшто се среќава во контактните зони меѓу словенски и грчки дијалекти, може да претставува локализирана ареална карактеристика. Анализата ги доведува во прашање претходните претпоставки за ареалниот карактер на одредени прозодиски особини.

Во статијата на **Макс Валстром** и **Дон Килиан** се истражуваат морфосинтаксичките и семантичките карактеристики на тројните членови во македонскиот јазик со цел да се определи нивниот статус во пошироката типологија на определеност и деиксис. Користејќи податоци од говорниот корпус на македонскиот јазик, авторите покажуваат дека помеѓу проксималните, неутралните и дисталните форми постојат значајни разлики во дистрибуцијата. И во другите балкански словенски јазици еден член од трипартитниот систем доминира над другите. Авторите исто така критички ги разгледуваат тврдењата за секундарните функции на овие членови, тврдејќи дека тие не се главни критериуми за да се сметаат определени членови.

Статијата на **Оливије Винисторфер**, **Анастасија Ешер** и **Дарија Кониор** го испитува диференцијалното означување на локацијата во влашкиот и во сродни балканско-романски варијанти. Користејќи податоци од Крушево, Охрид, Струга и Турија/Кранеа, авторите ги анализираат условите под кои локативните релации се означуваат експлицитно и кога не се означуваат со предлог. Анализата покажува дека факторите како што се референцијалниот статус на топонимот (сопствена наспроти општа именка) и перципираното растојание до местото означено со топонимот (проксимално наспроти дистално) играат значајна улога во употребата на диференцијалното означување на локацијата во овие дијалекти.

Општо земено, статиите во овој тематски зборник придонесуваат за нашето подлабоко разбирање на јазичниот контакт и неговото влијание врз внатрешните промени во морфосинтаксата и фонолошките системи на јазиците во овој регион. Зборникот ја нагласува важноста на продлабоченото дијалектолошко истражување и компаративната типолошка анализа за изучувањето на балканскиот јазичен простор.

Им изразуваме благодарност на сите автори за нивните темелни и оригинални истражувања со што се овозможи реализацијата на овој тематски број на списанието *Современа Филологија*.

Елени Бужаровска

Гостин уредник

Редовен професор на Филолошкиот факултет „Блаже Конески“,
Универзитет „Св. Кирил и Методиј“ во Скопје

SYNTAX OF CLITICS IN BULGARIAN CONTACT DIALECTS

Elena Ivanova

St. Petersburg State University
e.y.ivanova@spbu.ru

The article analyzes interference phenomena in the syntax of clitics in Bulgarian dialects that have been in contact with the Romanian language over a long period of time. The analysis focuses on Bulgarian dialects spoken in the territory of Romania, as well as the specific dialect of Novo Selo in northwestern Bulgaria. It is assumed that both Romanian syntax and dialectal features introduced by Macedonian settlers have influenced this dialect. The principles of clitic order in the sentence, characteristic of Standard Bulgarian, are largely preserved in these dialects. This is mainly due to the fact that both the donor and recipient languages belong to the group of languages with verb-oriented clitics. The analysis shows that in the examined Bulgarian dialects, clitics can appear in initial position, which means that the Tobler-Mussafia law is violated under the influence of the contact languages. At the same time, most of the restrictions on initial clitic placement present in the donor language are shown to be irrelevant.

Keywords: contact-induced linguistic changes, Romanian, Tobler-Mussafia law.

СИНТАКСАТА НА КЛИТИКИТЕ ВО БУГАРСКИТЕ ДИЈАЛЕКТИ ВО ЈАЗИЧЕН КОНТАКТ

Елена Иванова

Државен универзитет во Санкт Петербург

e.y.ivanova@spbu.ru

Во статијата се анализираат интерференциските феномени во синтаксата на клитиките во бугарските дијалекти што се во контакт со романскиот јазик во подолг временски период. Во фокусот на анализата се бугарските дијалекти на територијата на Романија, како и специфичниот дијалект на Ново Село во Северозападна Бугарија. Се претпоставува дека романската синтакса и дијалектните особини на македонските доселеници влијаеле врз овој дијалект. Принципите на редоследот на клитиките во реченицата, карактеристични за стандардниот бугарски јазик, се во голема мера зачувани во овие дијалекти. Тоа главно се должи на фактот дека и јазикот давател и јазикот примач припаѓаат на јазици со глаголски ориентирани клитики. Анализата покажува дека во разгледуваните бугарски дијалекти, под влијание на соседните јазици, клитиките можат да заземат иницијална позиција, со што се прекршува законот на Тоблер–Мусафија. Истовремено, поголемиот дел од ограничувањата за иницијалната позиција на клитиките во јазикот давател се покажуваат како ирелевантни.

Клучни зборови: промени од јазичен контакт, романски јазик, закон на Тоблер–Мусафија

1 Introduction and general information about contact dialects

Bulgarian is known to be one of the languages in which the so-called Tobler-Mussafia law applies, as it does not allow enclitics to be placed on the left periphery of the clause (e.g., Dimitrova-Vulchanova 1999; Franks 2008; Dimitrova 2023: 54–57). However, in areas of contact with Romanian, a language in which this law does not apply, Bulgarian speakers may use initial clitics, as descriptions of the Greben dialect located in north-eastern Bulgaria (Kochev 1969) or the dialect of the endmost north-western region in Bregovo Municipality demonstrate (Marinov 2008). It is more pronounced in fully foreign-language environments, particularly in Bulgarian dialects on Romanian territory (Mladenov 1993).

The object of our interest in this paper will be two contact areas, each of them exhibiting evidence of the influence of the Romanian language on the Bulgarian dialect. Additionally, in one of these areas, certain phenomena are observed that may have emerged under the influence of Macedonian dialects. Our aim is to compare the possibilities for the realization of clitic initiality in the two regions under review, highlighting both the permissions and constraints regarding the clitic placement at the absolute beginning of the sentence in these Bulgarian contact dialects. A key question is whether the morphosyntactic constraints on the clitic initiality present in the donor language are transferred to the recipient language.

1) The first area is located in the south of Romania in the historical region of Wallachia, where to the north of the Danube numerous villages are inhabited by descendants of Bulgarian settlers who arrived primarily during the 18th and early 19th centuries.¹

In southern Romania, all groups of Bulgarian dialects found in northern Bulgaria are also represented. The language of these dialects has been described in a number of studies (Bolokan 1968; Dimchev 1974; Sugai 2015a, 2015b, among others), most comprehensively in a monograph by Mladenov – *Bulgarian Dialects in Romania* (Mladenov 1993), and is documented in the *Transdanubian Electronic Corpus* of texts and audio recordings (Mladenova and Mladenova 2001–2018), a supplement to Mladenov’s monograph. Our study uses only a portion of these materials, mainly idioms from the villages of the territory of Muntenia and around the Olt River. They are representative of Bjala-Slatina, Cibrica-Ogosta, Nikopol and Moesian dialects found within the territory of Romania. Although the idioms of each of the considered dialects have their particularities, the situation regarding initial clitics displays common features making it acceptable to apply a generalised analysis to the loci of this area.

As the texts of the *Transdanubian Electronic Corpus* show, the level of Bulgarian language proficiency among the residents varies not by locality or even by village, but at the level of families and individual informants: some informants have high degree of competence in the Slavic idiom, while others can produce only specific utterances. The linguistic behaviour of informants depends largely on factors such as age, education, descent, discourse strategies within the family and community. However, in general, the language situation in the villages under analysis can be characterised as one of non-balanced bilingualism, which is typical of the modern Balkan region as a whole (e.g., Konyor and Sobolev 2017; Morozova and Rusakov 2021). In such situations, “the speakers continue to use their L1, but the sociolinguistically dominant L2 becomes also linguistically dominant for them” (Morozova and Rusakov 2021: 1012). As early as the 1970s, researchers of Bulgarian dialects in Romania noted a high degree of linguistic integration of the Bulgarian population (see Mladenov 1993: 50–54, 364–372). For many informants, Romanian was then (and is now) the dominant language of communication both within and outside the family. Considerable interference is evidenced by a number of structural and lexico-semantic changes that emerged under the influence of

¹ For more details see Romanski (1930), Mladenov (1993), Mladenov, Nyagulov and Zhechev (1994).

Romanian (Dimchev 1974; Mladenov 1993), which we cannot delve into here. We only note that on the Thomason and Kaufman scale (1988: 74), the level of this interference can be estimated to be at least 3 (Ivanova, in press).

2) The second focus of our attention is the village of Novo Selo, which is located on the banks of the Danube River, in the northwestern part of Bulgaria (Vidin Region), a few kilometres east of the mouth of the Timok River. The inhabitants of this village have long been in contact with the Romanian-speaking population, surrounded by neighbouring Romanian-speaking villages. However, there was no active bilingualism at the time of data collection in the 1950s and 1960s (Mladenov 1969: 10; Mladenova 2003: 1), and this remains the case today. Only very elderly individuals still understand Romanian.

The dialect of Novo Selo is an autonomous, coherent and unique idiom. The genetic basis of the dialect is a subject of debate (cf. Mladenov 1969; Sobolev 1995). We will not address this issue here, as for the present work, it suffices to say that alongside the Romanian influence, there are features in this dialect introduced by Macedonian speakers. According to Miletic's hypothesis (Miletic 1901: 639–641), these could have been residents of Tetovo, Kumanovo and Kratovo regions in Macedonia, who also lived for some time in Banat (Mladenova 2003: 1–2).²

Thus, although the level of contact with the Romanian-speaking population in the history of this dialect has not been as constant and prolonged as for Bulgarians in Romania, the influence of the Macedonian language has had an effect, in particular, in that some constructions with initial clitics are widely represented in this dialect. “The impact of the two languages with the cancelled constraint on the clitic's initial position had a catalysing effect on this tendency” (Mladenov 1969: 162).

The primary source of material for this paper consists of the linguistic data available in the descriptions of the examined dialects by Maksim Mladenov (Mladenov 1969, 1993), as well as the texts from the aforementioned electronic resource by Olga Mladenova and Darina Mladenova (2001–2018). All these materials represent data from the 1950s and 1960s. Some information about the current state was retrieved from the works of Sugai (Sugai 2015a, b), which have data of 2012–2013 from the Romanian villages of Valea Dragului and Brănești, and from Ivan Iliev's interviews with residents of Novo Selo (Iliev 2018).

The article is further structured as follows: in section 2, which provides an account of the syntax of clitics in Standard Bulgarian, we also examine contact languages with the cancelled Tobler-Mussafia requirement and list their restrictions on the initial position of clitics. Section 3 addresses the paradigmatic and syntagmatic properties of initial clitics in the contact dialects under study, while section 4 presents a summary of the results.

2 Linearization of clitics in Modern Bulgarian and in contact languages: a brief survey

2.1 Modern standard Bulgarian features clitics³ from different morphological classes: pronominal clitics (short forms of the dative and accusative cases, including reflexive forms), verbal clitics (forms of the verb *съм* ‘to be’ in the present tense), and discourse particles (the interrogative particle *ли*, dative ethicus *ми*, *му*, the modal particle *си*).

The analysis of the syntactic behaviour of clitics involves at least two aspects:

1) sequence of clitics with respect to each other (clitic template); see Table 1.

² On the Macedonian influence on the dialect, see Stephan Mladenov (Mladenov 1901: 498) and Maxim Mladenov (Mladenov 1969: 71–77, 193–195).

³ We adopt a syntax-based approach to the identification of clitics: “*syntactic* or ‘special’ clitics in terms of (Zwicky 1977) are elements taking syntactic positions non-available for *phrases*, i.e. multi-word constituents with head and complement elements”. (Zimmerling 2022: 7)

2) the rules for the cluster placement in a sentence.

We will focus only in the latter aspect, as in the dialects under study the clitic template does not differ significantly from Standard Bulgarian,⁴ while the placement of a cluster of clitics in a sentence exhibits its own peculiarities.

Table 1. The order of clitics in a non-interrogative Bulgarian sentence (according to Ivanova and Gradinarova 2015: 512)

1	2	3	4	5	6
Future tense particle <i>ще</i> / negative particle <i>не</i> ⁵	Verbal enclitics of the present tense, except 3.sg.: <i>съм, си, сме, сте, са</i>	Dativus ethicus <i>ми, ти</i> / modal particle <i>си</i>	Pronominal dative enclitics <i>ми, ти, му, ѝ, му; ни, ви, им,</i> / reflexive pronominal <i>си</i>	Pronominal accusative enclitics <i>ме, го, я, го; ни, ви, ги</i> / reflexive pronominal <i>се</i>	Verbal enclitic of the present tense <i>е</i> (3.sg.)

We focus only on clause-level clitics. Phrase-level clitics, in particular those of the nominal phrase/prepositional phrase, are discussed only insofar as they can be extracted from the phrase. In Bulgarian, these include the possessive clitics (genetically derived from the dative pronouns), which are subject to the operation of possessor raising and can be inserted into a chain of sentential clitics if the position of the dative clitic is not occupied (Schürcks and Wunderlich 2003: 11; Cinque and Krapova 2009), see section 3.⁶

Clitic clustering in Modern Bulgarian is described in the literature as involving elements that are both verb-adjacent and 2P elements (Dimitrova-Vulchanova 1999; Zimmerling 2012a, b; Zimmerling 2022: 12; Dimitrova 2023). The requirement for clitic-verb adjacency implies that clitics are in the immediate vicinity of the verb (before or after it), which is both their syntactic and, usually, prosodic host (Dimitrova 2023: 55–56). This is illustrated by examples (1a) with the initial subject *мой* ‘he’, (1b) with the initial adverb *вече* ‘already’, and (2) with the initial verb.

(1) XP–CL–V:

- a. *Той ми*_{1sg.dat} *се*_{refl} *обеди вече.*
‘He has already called me.’
b. *Вече ми*_{1sg.dat} *се*_{refl} *обеди.*

(2) V–CL:

- Обеди ми*_{1sg.dat} *се*_{refl} *вече.*
‘[He] has already called me.’

⁴ The most significant divergence is not in the sequence of the clitics, but in the position of the negative particle *не*, which in some Bulgarian dialects in Romania (Bjala-Slatina, Nikopol, Cibrica-Ogosta dialects) is adjacent to the verb (Mladenov 1993). Considerable shifts in the cluster are found in the Vidin-Lom dialects (Mladenov 1993: 81), which we do not analyze here.

⁵ It would be reasonable to consider *ще* and *не* not as parts of the Bulgarian cluster, but as so-called *clitic bases* (Zimmerling 2022: 12).

⁶ For detailed rules for clustering of pronominal clitics including the possessive dative and the combinatorics of particles of pronominal origin see Petrova and Ivanova (2017: 85–91).

The initial phrase (XP) in Bulgarian can also represent a group of constituents with equal communicative status, as in (3), cf. “Bg [Bulgarian] is the only Slavic language, where main clauses with long topicalized constituents before the clitics can be communicatively neutral”. (Zimmerling and Kosta 2013: 194)

- (3) XP [XP₁+XP₂] –CL–V:
 [Вчера рано сутринта Иван] ми_{1sg.dat} се_{refl} обади.
 ‘Early yesterday morning Ivan called me’ (Ivanova and Gradinarova 2015: 531)

The second position feature (*2P clitics*), as applied to Bulgarian as a language with verb-adjacent clitics, means that clitics are oriented to the left periphery of the clause. This is realized as the 2P position under the basic word order, as in the above examples. The 2P-position of clitics in the Bulgarian sentence is not an absolute rule, but only a tendency, which, however, has been statistically confirmed: a recent study of the corpus of spoken Bulgarian (<http://www.bgspeech.net/>) by Dimitrova revealed that under basic word order, pronoun clitics occupy the second position in more than 80 percent of their occurrences (Dimitrova 2023: 52).

As in most languages with clitic clusters, Bulgarian has syntactic barriers that give rise to derived context-specific word orders moving the cluster of clitics to the right of the left boundary of the sentence⁷. Interrogative utterances have a wider set of barriers than declarative ones, to the effect that even the XP-V-Q -CL order is allowed, which is ruled out for non-interrogative sentences, as seen in (4) with a barrier (/) after the subject *мой* ‘he’:

- (4) S //V–Q – CL [CL.3SG.DAT–CL.3SG.ACC–CL.3SG.PRS]
 Той // върнал ли му го е?
 ‘Has he returned it to him?’

3) Bulgarian is one of the languages where the Tobler-Mussafia law applies, see, e. g., Franks (2008), Dimitrova-Vulchanova (1999), Dimitrova (2023: 54–57).

In the dialects discussed below, all the linear-syntactic properties of clitics are observed, except for parameter 3, i.e., they allow clitics to be placed on the clausal left edge. In (5), we can clearly see the difference between the realization of word order in the Bulgarian standard language and in the Bulgarian dialects of Romania. The first line comes from a dialectologist (A), a native speaker of the standard language, who, following the standard rules, places the pronoun clitic *му* in the postverbal position. By contrast, the informant (B) in his reply begins the clause with this clitic:

- (5) A. туряте му_{3sg.dat.m} сол
 B. Сол. да| му_{3sg.dat.m} турим сол (R, Calomfirești, M-1)⁸
 ‘You put salt into it [soured milk]’
 ‘Salt. Yes, we put salt into it.’

⁷ For more details see Dimitrova-Vulchanova (1999), Dimitrova-Vulchanova and Hellan (1999), Zimmerling (2012b: 19–20), Zimmerling (2013: 113–118).

⁸ Examples are given in the transcription of their sources and are presented in the following way: examples from the Bulgarian dialects of Romania bear the mark “R” (if known, the exact locus is specified), for the dialect of Novo Selo the mark “NS” is used. Next, the source of the example is indicated: M1 – the *Transdanubian Electronic Corpus*, M-2 – Mladenov 1993, M-3 – Mladenov 1969, or other printed source. The number after the colon (in some examples below) indicates the page number of the printed source.

The initial position of clitics in these dialects undoubtedly has a contact-induced nature. Before we address its realization in more detail, a few remarks will be made about the clitic placement in the languages that influenced the borrowing.

2.2 Romanian belongs to the group of languages *with VP-internal clitics* (Zimmerling 2022: 9), like other Romance languages, Modern Greek, and Albanian. Although the syntax of clitics in languages with VP-internal clitics and those with verb-adjacent clitics is somewhat different, in both types of languages pronominal clitics are verb-oriented.

In Romanian, pronoun clitics are usually placed preverbally, with the exception of a few morphosyntactic contexts. According to Gerstenberger (2022), Romanian weak pronouns occur in preverbal position with finite (6a), non-finite verb forms, and negated imperative forms, and in postverbal position with participles/gerunds as well as with non-negated imperative verb forms (6b).⁹

- (6) a. *Mi*_{1sg.dat} *le*_{3pl.acc.f} *dai acum*. ‘You give them to me now.’
 b. *Dă-mi*_{1sg.dat} *-le*_{3pl.acc.f} *acum!* ‘Give them to me now!’ (Gerstenberger 2022: 57)

There is also a phonological restriction for the pronoun /o/, 3.sg.acc.f. As Gerstenberger (2022: 41) notes: “Preverbally, it occurs only if there is no auxiliary starting with a vowel.”

- (7) a. *Mi*_{1sg.dat} *-o*_{3sg.acc.f} *dai*.
 You give her/it to me.’
 b. *Mi*_{1sg.dat} *-ai dat-o*_{3sg.acc.f}
 ‘You have given her/it to me.’ (Gerstenberger 2022: 57)

2.3 Macedonian belongs to languages with VP-adjacent clitics (Zimmerling 2022: 9) and to those that allow the front position before clitics to be unfilled (Dimitrova-Vulchanova 1999: 74). Restrictions on initial clitics depend not only on the finiteness/non-finiteness of the predicate, but above all on its type – verbal or non-verbal. According to a formulation by Zimmerling, “Macedonian is a language with so-called *clause shifting* <...>, where the linearization strategy and the prosodic orientation of clitics (proclitics vs strict enclitics vs universal clitics) crucially depend on the clause type” (Zimmerling 2022: 13).

This allows to place Macedonian language in an “extremely rare” typological group of CL1/CL2 languages (Zimmerling 2015: 467).¹⁰

The main types of sentences that restrict the initiality of clitics in Macedonian are those with a nominal predicate (see example 8a with a noun predicate and 8b with an adjective predicate) and those with an imperative (9a). Notably, the restrictions on the imperative apply not only to the positive, but also to the negative forms (9b). Non-finite forms also impose restrictions: clitics are postverbal when used with adverbial participles (9c), and variation in placement is allowed with other participles (9d). Sentences with finite verbal predicates do not have morphosyntactic restrictions on the initiality of clitics (10).

- (8) a. *Правник сум*_{1sg.prs} (**Сум правник*) (Mišeska Tomić 2008: 48)
 ‘I am a lawyer’
 b. *Болен ѝ*_{3sg.dat.f} *e*_{3sg.prs} *сином* (ibid: 33)
 ‘Her son is sick / She has a sick son.’

⁹ See also Niculescu (2008).

¹⁰ Cf. the analysis in Alexander (1994: 3–8) and Mišeska Tomić (2008: 9–52).

- (9) a. *Земѝ зО_{3sg.acc.m}* (ibid: 51)
‘Take it.’
- b. *Не давај му_{3sg.dat} зО_{3sg.acc.m}* (ibid: 52)
‘Don’t give it to him.’
- c. *Даважќи му_{3sg.dat} зО_{3sg.acc.m}*
‘Giving him it.’ (Dimitrova-Vulchanova 1999: 76)
- d. *Му_{3sg.dat.m} е_{3sg.prs} речено да дојде / Речено му_{3sg.dat.m} е_{3sg.prs} да дојде*
‘He is told to come.’ (Mišeska Tomić 2008: 44)
- (10) *Ми_{1sg.dat} се_{refl} истури млекото.* (Mitkovska 2011: 87)
‘The milk spilt on me.’

The extent to which these Romanian and Macedonian restrictions are reflected in the dialects under discussion will be addressed below.

3 Initial clitics in Bulgarian dialects of Romania and in the Novo-Selo dialect

In the dialects under study, initial clitics have a wide distribution. The wide range of their paradigmatic and syntagmatic possibilities becomes particularly evident when they are compared to dialects that have a weak degree of contact with Romanian, such as the Greben dialect in Bulgaria (Kochev 1969), where only residual contact-induced phenomena are present. As shown in Ivanova (2024) and Ivanova (in press), in the Greben dialect, initial clitics: a) have a restricted paradigm (3rd person auxiliary verbs do not appear at the beginning of a sentence), b) cannot function as represent copulas, c) are allowed only in declarative sentences, d) are optional.

In this context, the examined dialects of Romania and the Novo-Selo dialect demonstrate an obvious expansion of the initiality feature and display similar tendencies in the extension of clitic placement possibilities.

1. The clitics that can start a clause have no categorial constraints: they can be verbal and pronominal, including particles of pronominal origin, such as the reflexive *се* and *си*. Inflected clitics are represented in the material of the dialects by full paradigms. Below are the examples from the Bulgarian idioms of Romania (11) and from the Novo Selo dialect (12).

- (11) a. *Е_{3sg.prs} извадила пилетата* (R, Stoenеști, M-2: 383)
‘She took out the chickens.’
- b. *ЗУ_{3sg.acc.m} зеа на рџцете* (R, Băleni-Sârbi, M-1)
‘He was taken in their arms.’
- c. *си_{refl.dat} вџрви свадбџтъ нџпрет* (R, Bila, M-1)
‘The wedding is going on’
- (12) a. *Е_{3sg.prs} бил бџчварин* (NS, M-3: 304)
‘He was a barrel maker.’
- b. *Мџ_{1sg.acc} зрџбе* (NS, M-3: 117)
‘I’m itchy.’
- c. *Се_{refl} дигнџиџмо у двџнаес чџса* (NS, M-3: 302)
‘We used to get up at twelve o’clock.’

In the analyzed material, there are no contexts with initial *dativus ethicus*, which is due to the functional peculiarities of the latter – namely, its occurrence within fixed structures (such

as *Такива ми ти работи* ‘That's how it is’) and the emotional colouring of the utterance. The initial positioning in the examined dialects, on the other hand, is mainly linked to the communicatively neutral sentence opening (see below).

2. The initial placement of clitics in the dialects under study does not depend on the morphological form or finiteness of the predicate. Initial clitics occur with both verbal and nominal predicates.

- (13) a. *Съм*_{1sg,prs} *на шейсет и три годин* (R, Coteana, M-1)
‘I'm 63 years old.’
b. *Съм*_{1sg,prs} *из Руманија* (R, Băleni-Sârbi, M-1)
‘I am from Romania.’
c. *Съм*_{1sg,prs} *каснер* (R, Valea Dragului, Dimchev 1974: 256)
‘I'm a cashier.’
d. *Ми*_{1sg,dat} *ѣ*_{3sg,prs} *гланну* (R, Chiajna, M-2: 377)
‘I'm hungry.’
- (14) a. *Ми*_{1sg,dat} *е*_{3sg,prs} *стра* (NS, M-3: 163)
‘I am afraid.’
b. *Ми*_{1sg,dat} *е*_{3sg,prs} *сме* (NS, M-3: 163)
‘It's funny to me.’

It should be noted, however, that no examples with an initial copula as in (13a, b, c) have been found in the dialect of Novo Selo. Mladenov, who, importantly, was himself a native speaker of the dialect, also observed that the verb *съм* ‘in independent use’ (not as part of a verbal form, cf. the perfect form *е бил* in [12a]) does not appear at the beginning of a sentence in this dialect (Mladenov 1969: 163).

The influence of the constraints inherent in the Macedonian language cannot be ruled out here: in Macedonian, as shown in 2.3, the copula in sentences with nominal predicate cannot be positioned initially, as in examples (8a, b) above.

In both areas, the initial position can be occupied by a verbal clitic with a participial predicate. This is especially relevant in the idiom of Novo Selo, where the variations of clitic placement with participial forms, peculiar to Macedonian (as in 9d above), do not seem to apply. The preposed copula, on the contrary, is typical in this dialect:

- (15) *Съм*_{1sg,prs} *станът у комуна Ѓепуреши* (R, Iepurești, M-2: 378)
‘I was born in the municipality Iepurești.’
- (16) *Съм*_{1sg,prs} *легнут по бурту* (NS, M-3: 184)
‘I'm lying on my front.’

As for adverbial participles, the dialect texts under study do not provide sufficient data to examine their ordering in relation to pronominal clitics.

3. In both the Bulgarian dialects of Romania and the Novo-Selo dialect, possessive datives have been observed in the initial position. This occurs in constructions where the external possessor is expressed by a short dative form, typically allowing a combined possessive and argument interpretation.

- (17) a. *Му*_{1sg.dat} *и*_{3sg.prs} *имету Ёозин Жизела* (R, Cioplea, M-2: 383)
 ‘My name is Josine Gisela.’
 b. *И*_{3sg.dat.f} *умр’а чил’аку* (R, Valea Dragului, Sugai 2015a: 105)
 ‘Her husband died.’, lit. ‘to her.’
- (18) a. *Му*_{1sg.dat} *потечъ кръв из нос* (NS, M-3: 132)
 ‘I got a nose bleed.’
 b. *Му*_{3sg.dat.m} *излезъл кукуй нь чъло* (NS, M-3: 135)
 ‘He's got a lump on his forehead.’

An expansion of the options for the initial possessive dative in the Bulgarian dialects of Romania can be observed in the example below. In (19), from the Moesian dialect, the absolute initial position is occupied by the possessive clitic *му*, which doubles the prepositional phrase *на мунчето*.

- (19) *Му*_j *имиту [на мунчету туй]*_j *Кент* (R, Brănești, Sugai 2015a:105)
 lit. ‘To him, the name [of the boy this] Kent’ / ‘The name of this boy is Kent.’

The prevalence of such expressions of possessiveness in the Bulgarian dialects is a predictable linguistic outcome of contact. The basic strategy for modern Romanian involves the use of structures with an external possessor, rather than DP-internal clitics: “DP-internal clitics are nowadays used mainly for stylistic reasons, in poetry or (highly) poetic texts. They are perceived as outdated and are no longer productive” (Hill and Tasmowski 2008: 367–368; Niculescu 2008: 487–499). In Macedonian as well, the structure with external possessor is firmly established (cf. 8b), while DP-internal clitics have a limited range of use (Mišeska Tomić 2008: 23–35; Mitkovska 2011; Mitkovska 2014: 109–130). Moreover, some external datives allow only a possessive interpretation, e. g., *Tu go чув името* (Mišeska Tomić 2008: 33; see also Mitkovska 2011: 93–100).

4. In both areas, initial clitics are allowed not only in declarative sentences but also in general yes-no questions:

- (20) *Tu*_{2sg.dat} *цѣ*_{refl} *досна?* (R, Iepurești, M-2: 383)
 ‘Did you feel like sleeping?’
- (21) *Cu*_{2sg.prs} *за*_{3sg.acc.n} *узел?* (NS, M-3: 183)
 ‘Did you take this?’

The only exception in both areas is the positive imperative form, with obligatory postverbal positioning of clitics:

- (22) *питѣйте мѣ*_{1sg.acc} (R, Valea Dragului, M-1)
 ‘Ask me.’
- (23) *Узи си*_{refl.dat} *мърку леп* (NS, M-3: 73)
 ‘Get yourself some bread.’

5. In both areas, the construction with pronoun reduplication is actively used. Particularly relevant to our discussion is the widespread use – unlike in colloquial Bulgarian and most Bulgarian dialects (Krapova and Tiševa 2006) – of the construction with right dislocation (Sugai 2015b: 97–100), which allows the occurrence of initial clitics. Such a communicative

syntactic pattern reflects a broader trend towards the grammaticalization of pronominal reduplication of the object (cf. Sugai 2015b).

(24) *Им*_{3pl.dat} *цѣлувѣ ръкъ нѣ кумѣ и куматѣ* (R, M-2: 233)
‘She kisses their hands, of godfather and godmother.’

(25) *Га*_{3sg.acc.m} *убили бѣшту ни* (NS, M-3: 74)
‘He was killed, our father.’

6. Initial clitics in the informants’ usage have come to be associated with a communicatively neutral status in clause initial position (also evidenced by an increased use of right dislocation in instances of pronominal reduplication). Our observations of the texts from Bulgarian dialects in Romania show that the initial clitic order is the default option for most informants, that is, the initial positioning of clitics is used as a primary narrative strategy.

The pattern appears most consistently in narratives recounting a sequence of actions, when the speaker simply conveys their chronological order. This is illustrated in the informant’s response (26) to the question of how to cook *kachamak* (a type of maize porridge):

(26) *качамак? къ?|| гу*_{3sg.acc.m} *турии нѣ огън’ѣ| го*_{3sg.acc.m} *вариии| го*_{3sg.acc.m} *бъркъии|*
*го*_{3sg.acc.m} *вариии| го*_{3sg.acc.m} *бъркъии ду куѣи го нѣпраии къчѣмак* (R, Calomfirești, M-1)
‘Kachamak? How? You put it on fire, boil it, stir it, boil it, stir it, until you’ve made kachamak.’

The postverbal position of clitics in similar communicative-syntactic conditions (if the informant uses this option at all) typically signals a departure from the norm; that is, it correlates with the presence of information-structural triggers or occurs in special constructions involving postposition, such as lexical repetition with syntactic extension in (27).

(27) *сетне ѿ*_{3sg.acc.f} *дѣрпѣми| дѣрпѣми ѿ*_{3sg.acc.f} *дудет съ утлени*
‘Then we tug it, tug it until it comes off.’ (R, Valea Dragului, M-1)

However, the factors influencing the choice between initial and non-initial position for some informants still require further investigation.

4 Concluding remarks

The expansion of paradigmatic and syntactic possibilities for initial clitics as a contact-induced phenomenon in the dialects under consideration follows similar patterns: the involvement of different morphological classes of clitics, the expansion of constructions that permit initial clitics (including nominal constructions, constructions with external possessors, and constructions with pronominal reduplication of the right-dislocation type), and the inclusion of not only declarative but also interrogative sentences. At the same time, the restrictions on clitic initiality from the donor language become irrelevant. An exception to this is the absence of constructions with an initial copular verb in the present tense in the Novo Selo dialect, mirroring the restrictions found in Macedonian. However, other Macedonian postverbal position rules are not represented: clitics with participial predicates always take the preverbal position, (cf. [9b] and [15–16]); the postverbal position with the negative imperative is not

allowed. A common type of clause that prevents clitic initiality is the positive imperative, which reflects a relict phenomenon found both within and outside the Balkans.

The initial position of clitics appears to be quite borrowable in Bulgarian when in contact with languages with cancelled Tobler-Mussafia requirement. We cannot delve into the peculiarities of cliticization in Bulgarian that favor this type of borrowing (Ivanova in press), but this transfer is undoubtedly facilitated by the word order in the donor languages. This is related to the role of the verb complex in sentence structure and the parallelism in the pronoun-verb sequence across the Balkan languages (Friedman and Joseph in press).

Any instance of borrowing calls for the discovery of motivation, which, in turn, can account for the greater or lesser borrowability of various phenomena (Matras 2020: 173–175, 252–257).

Thus, a functional trigger of clitic initiality in contact areas is the simplification of the linear-syntagmatic structure, serving as a means to promote uniformity within the bilingual linguistic repertoire (Matras 2020: 257). This simplification is evident in the fact that a native speaker of these Bulgarian dialects does not need the additional movement required in standard Bulgarian, namely, the step of moving clitics to the postverbal position in those (quite frequent) cases where the position before the verb-clitic complex remains unfilled. Indirect evidence supporting simplification as a motivation for borrowing is the disregard, in the recipient language, of most clitic postposition rules from the donor language, as demonstrated by the reviewed dialects.

Bibliography

- Alexander, R. (1994). The balkanization of Wackernagel's Law. *Indiana Slavic Studies*, 7: 1–8.
- Bolokan, G. (1968): Bolgarskij govor sela Brenesht'. Vokalizm. *Revue roumaine de linguistique*, 2: 147–164.
- Cinque, G. and Krapova, I. (2009). The two „possessor raising” constructions of Bulgarian. In S. Franks, V. Chidambaram and B. D. Joseph (eds.). *A Linguist's Linguist: Studies in South Slavic Linguistics in Honor of E. Wayles Browne*, 123–148. Bloomington: Slavica Publishers.
- Dimchev, K. (1974). Morfologichno-sintaktichni modeli, ustanoveni pod rumänsko vliyanie v bälgarskiya govor na s. Valya Draguluj (Rumänija). *Ezikovedski izsledvaniya*, 255–259. Sofiya: BAN.
- Dimitrova, C. (2023). *Bälgarskite klitiki: istoriya i nastoyashche*. Sofiya: Institut za bälgarski ezik.
- Dimitrova-Vulchanova, M. (1999). Clitics in the Slavic languages. In van H. Riemsdijk (ed.). *Clitics in the languages of Europe*, 83–121. Berlin/New York: de Gruyter.
- Dimitrova-Vulchanova, M. and Hellan, L. (1999). Clitics and Bulgarian Clause Structure. In van H. Riemsdijk (ed.). *Clitics in the Languages of Europe*, 469–514. Berlin/New York: de Gruyter.
- Franks, S. (2008). Clitic placement, Prosody and the Bulgarian verbal complex. *Journal of Slavic linguistics*, 16 (1): 91–137.
- Friedman, V. A. and Joseph B. D. (in press). *The Balkan Languages*. Cambridge: Cambridge University Press.
- Iliev, I. (2018). Bälgarska dialektologiya: Novo selo, Vidinsko. Available from: <https://www.youtube.com/watch?v=-HmU4xQvzmE>
- Ivanova, E. Yu. (2024). Linejno-sintaksicheskie svoystva klitik v grebenskom dialekte bolgarskogo yazyka. In *LII Mezhdunarodnaya nauchnaya filologicheskaya konferenciya imeni L. A. Verbickoj 19–26 marta 2024, Sankt-Peterburg. Sbornik tezisov*, 1473–1475. SPb.: Izd-vo SPbGU.
- Ivanova, E. Yu. (in press). Inicial'nye klitiki v bolgarskih govorah v svete kontaktologii. *Indoevropskoe yazykoznanie i klassicheskaya filologiya*, XXIX.
- Ivanova, E. Yu. and Gradinarova, A. A. (2015). *Sintaksicheskaya sistema bolgarskogo yazyka na fone russkogo*. Moscow: Yazyki slavyanskoj kultury.

- Kochev, I. (1969). *Grebenskiyat govor v Silistrensko (s osoben ogled kăm leksikalnata mu sistema). Trudove po bălgarska dialektologiya*. Kn. 5. Sofiya: BAN.
- Konyor, D. V. and Sobolev, A. N. (2017). Osobennosti neravnovesnogo bilingvizma u rumynoyazychnyh karashevcev v sele Yabalcha. *Indoevropskoe yazykoznanie i klassicheskaya filologiya*, 21: 985–1001.
- Krapova, I. and Tisheva, J. (2006). Clitic reduplication structures in the Bulgarian dialects. In M. Koletnik and V. Smole (eds.). *Diahronija in sinhronija v dialektoloških raziskavah*, 415–421. Maribor: Slavistično društvo.
- Legendre, G. (2000). Optimal Romanian Clitics: A Cross-Linguistic Perspective. In V. Motapanyane (ed.). *Comparative Studies in Romanian Syntax*, 227–264. Oxford: Elsevier.
- Marinov, V. (2008). *Bilingvalna interferentsiya v krayniya bălgarski severozapad*. Veliko Tărnovo: Astarta.
- Matras, Y. (2020 [2009]). *Language Contact*, 2nd edn. Cambridge: Cambridge University Press.
- Miletič, L. (1901). “Arnautite” v Silistrensko i sledi ot nosovki v tekhnija ezik. *Periodichesko spisanie na Bălgarskoto knizhovno druzhestvo v Sofiya*, God. 12, kn. LXI, sv. 9, 623–667. Sofiya: Dărzhavna pečatnica. Available from: [Miletitch_arnautite_vo_Silistrensko.pdf](#)
- Mišeska Tomić, O. (2008). Makedonskite klitiki vo balkanskoto okruženie. Skopje: MANU.
- Mitkovska, L. (2011). Competition between Nominal Possessive Constructions and the Possessive Dative in Macedonian. In M. Nomachi (ed.). *The Grammar of Possessivity in South Slavic Languages: Synchronic and Diachronic Perspectives*, 83–110. Sapporo: Hokkaido University.
- Mitkovska, L. (2014). Izrazuvanje prisvojnost vo makedonskiot jazik. Vnatrešni i nadvorešni prisvojni konstrukcii. *Makedonistika*, 13: 9–176.
- Mladenov, M. S. (1969). *Govorăt na Novo selo Vidinsko. Prinov kăm problema za smesenite govori. Trudove po bălgarska dialektologiya*. Kn. 6. Sofiya: BAN.
- Mladenov, M. S. 1993: *Bălgarskite govori v Rumăniya*. Sofia: BAN.
- Mladenov, M., Nyagulov, B. and Zhechev, N. (1994). Bălgarite v Rumăniya XVII–XX v. Dokumenti i materialı. Sofiya: Akademichno izdatelstvo “Prof. Marin Drinov”.
- Mladenov, St. (1901). Kăm văprosa za ezika i nacionalnata prinadlezhnost na Novo selo, Vidinsko. In Iv. Shishmanov (ed). *Sbornik za narodni umotvoreniya, nauka i knizhnina*. 471–506. Kn. XVIII. Sofiya. Dărzhavna pečatnica.
- Mladenova, O. (2003). *English summary to: M. Mladenov. The Dialect of Novo Selo, Vidin Region: A Contribution to the Study of Mixed Dialects*. PRISM: University of Calgary's Digital Repository. Available from: <http://hdl.handle.net/1880/44632>
- Mladenova, O. and Mladenova, D. (2001–2008). *Transdanubian Electronic Corpus*. Available from: <http://www.corpusbdr.info>.
- Morozova, M. S. and Rusakov, A. Y. (2021). Societal multilingualism à la balkanique: the Montenegrin Velja Gorana and beyond. *International Journal of Bilingualism*, 25 (4): 999–1018.
- Niculescu, D. (2008). Romanian possessive dative – the limits of the structure. *Revue roumaine de linguistique*, 53(4): 485–515.
- Petrova, G. and Ivanova, E. (2017). Klăsterizaciya na mestoimennite klitiki v bălgarskiya ezik: văzmozhnosti i ogranicheniya. *Săvremenni tendentsii v ezikovedskite izsledvaniya*, 116–130. Sofiya: Akademichno izdatelstvo "Prof. Marin Drinov".
- Romanski, S. (1930). *Bălgarite văv Vlashko i Moldova. Dokumenti*. Sofiya: Dărzhavna pečatnica.
- Schürcks, L. and Wunderlich, D. (2003). Determiner-possessor relation in the Bulgarian DP. In: M. Coene and Y. D’Hulst (eds.). *From NP to DP. Volume II: The expression of possession in noun phrases*, 121–139. Amsterdam: John Benjamins.
- Sobolev, A. N. (1995). O nekim južnoslovenskim govornim oazama u istočnoj Srbiji, zapadnoj Bugarskoj i Rumuniji (Vratarnica, Novo selo, Svinica). *Zbornik Matice srpske za filologiju i lingvistiku*, XXXVIII(2): 183–207.
- Sugai, K. (2015a). Za nyakoi specifichni harakteristiki na mestoimennite klitiki v grebenskiya govor v Rumăniya. *Bălgarska rech*, 2: 102–109.
- Sugai, K. (2015b). Kăm văprosa za udvoyavaneto na pryakoto dopălnenie v bălgarskiya govor v Rumăniya (Analiz v săpostavka s rumănskiya ezik). In M. Mladenova (ed.). *Dvizhenie i prostranstvo v slavyanskite ezici, literaturi i kulturi*, 1: 94–102. Sofiya: Univ. izd. “Sv. Kliment Ohridski”.

- Thomason, S. G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Zimmerling, A. (2012a). Sistemy poryadka slov s klitikami v slavyanskix jazykah. *Voprosy jazykoznanija*, 5: 3–37.
- Zimmerling, A. (2012b). Sistemy poryadka slov s klitikami v tipologičeskom aspekte. *Voprosy jazykoznanija*, 4: 3–38.
- Zimmerling, A. V. (2013). *Sistemy poryadka slov slavyanskikh jazykov v tipologičeskom aspekte*. Moscow: Yazyki slavyanskoi kultury.
- Zimmerling, A. (2015). 1P orders in 2P languages. In E.A. Lyutikova, A.V. Zimmerling and M.B. Konoshenko (eds.). *Tipologiya morfosintaksicheskikh parametrov 2015*, 459–483. Moscow: Moscow Pedagogical State University.
- Zimmerling, A. (2022). Clitics. In J. Fellerer and N. Bermel (eds.). *The Oxford Guide to the Slavonic Languages*. Oxford: Oxford University Press.
- Zimmerling, A. and Kosta, P. (2013). Slavic clitics: a typology. *Language Typology and Universals (STUF)* 66 (2): 178–214.
- Zwicky, A. (1977). *On Clitics*. Bloomington: Indiana University Linguistics Club.

NEW RESEARCH ON ROMANI IN EASTERN NORTH MACEDONIA: THE MALESHEVO DIALECT

Kirill Kozhanov

Universität Potsdam
kozhanov@uni-potsdam.de

Victor A. Friedman

Chicago University
vfriedm@uchicago.edu

This article presents new data on previously undescribed Romani varieties spoken in the eastern part of the Republic of North Macedonia, specifically in the geographical region of Maleshevo. We show that this is a South Balkan dialect that differs from other South Balkan dialects spoken in North Macedonia (especially in Skopje), while sharing several features with the Romani dialects of southwestern Bulgaria. This pattern corresponds to the geographical distribution of South Slavic dialects. Following South Slavic terminology, these Romani varieties can be referred to as Maleshevo-Pirin Romani. In addition to describing typical dialectal features, this study pays special attention to borrowings from local Macedonian dialects.

Keywords: Romani, Macedonian, dialectology, language contact.

НОВО ИСТРАЖУВАЊЕ ЗА РОМСКИОТ ЈАЗИК ВО ИСТОЧНА СЕВЕРНА МАКЕДОНИЈА: МАЛЕШЕВСКИОТ ДИЈАЛЕКТ

Кирил Кожанов

Универзитет во Потсдам
kozhanov@uni-potsdam.de

Виктор А. Фридман

Универзитет во Чикаго
vfriedm@uchicago.edu

Во оваа статија се претставуваат нови податоци за претходно неописаните ромски варијанти, што се зборуваат во источниот дел на Република Северна Македонија, конкретно во географската област Малешево. Овие дијалекти, кои припаѓаат на групата јужнобалкански дијалекти на ромскиот јазик, се одликуваат со неколку карактеристики, што ги разликуваат од другите дијалекти што се зборуваат во Северна Македонија (првенствено во Скопје), но истовремено се поврзуваат и со ромските дијалекти од Југозападна Бугарија. Распределбата на овие ромски варијанти се поклопува со географската поделба на јужнословенските дијалекти и, според прифатената терминологија, можат да се наречат малешевско-пирински. Покрај опишувањето на нивните типични дијалектни карактеристики, во ова истражување посебно внимание им се посветува на заемките од локалните македонски дијалекти.

Клучни зборови: ромски, македонски, дијалектологија, јазичен контакт.

1 Introduction

Romani is an Indo-Aryan language that has been spoken in the Balkans since the Middle Ages (Matras 2002; Matras and Tenser 2020). The core of the Romani lexicon is undoubtedly of Indo-Aryan origin (see Oslon and Kožanov, in prep.), and its morphology retains several features typical of New Indo-Aryan languages (Beníšek 2020).

On their way from the Indian subcontinent to Europe, Romani speakers came into contact with Iranian and Armenian, as evidenced by loanwords (Scala 2020). However, the language was largely reshaped by contact with Medieval Greek during its development in the Byzantine Empire, where Romani speakers remained for several centuries: possibly since the 11th-12th centuries until the northern migrations, which began no later than the 15th century. During this period, Romani acquired several features typical of Balkan languages (Matras 1994; Friedman 1985, 2000), among which the use of finite subjunctive clauses instead of an infinitive serves as a characteristic example. Subsequent migrations of Roms¹ northward beyond the southern Balkans have led to the development of various dialects shaped by contact with different languages (Matras 2005).

Being spoken in southern Balkans, Romani must have been in contact with South Slavic languages prior to the 15th century, as evidenced by loanwords in those Romani dialect groups that left this area no later than the beginning of the 15th century (Boretzky 2013; Kozhanov and Oslon 2020). Historical documents mention Roms in the northern Balkans in the 14th century (Marushiakova and Popov 2001: 18–19), which is possibly the period when Roma settled in the territory of contemporary North Macedonia. However, the earliest attestations of Romani communities here are much later (for example, in Skopje, the capital of North Macedonia, in 1523; see Friedman 2017a: 30).

In North Macedonia, Romani is spoken throughout the country, with a particularly high concentration of speakers in the capital, Skopje. According to the 2021 census (https://www.stat.gov.mk/publikacii/2022/POPIS_DZS_web_MK.pdf), the Romani population of North Macedonia is approximately 49,000, or around 2.3% of the total population. Much of the Romani population resides in Skopje (approximately 20,000), particularly in the municipality of Šuto Orizari, where Romani is also recognized as an official language of administration (see Friedman [1999]).

Romani dialects spoken in North Macedonia belong to three distinct dialect groups: South Balkan, North Balkan, and South Vlax (for an overview of Romani dialectology, see Elšík and Beníšek 2020; for an overview of Romani dialects of Skopje, see Friedman 2017a). South Balkan Romani dialects (sometimes referred to as Balkan I) are primarily spoken in the southern Balkans and represent dialects that never left the Balkans. These dialects are characterized by the further contact with Greek, South Slavic, Turkish, and Albanian. There are also South Balkan dialects spoken outside the southern Balkans, such as Ursari in Romania and Crimean Romani in Ukraine and Russia. An overview of this dialect group is presented in Boretzky et al. (2008). In North Macedonia, this group is represented by Arli (the largest dialect of Skopje) and the Romani variety of Prilep (Boretzky and Cech 2016). North Balkan dialects (also referred to as Balkan II) are primarily spoken in the northern Balkans, with their center arguably in central and northern Bulgaria (for more details, see Boretzky 2000). Speakers of North Balkan dialects are usually referred to as Kovači in Skopje, where they are believed to

¹ In this article, when writing in English, we treat *Rom* as a normal English ethnonym rather than exoticizing it. Just as the English plural of *Turk* is *Turks*, not *Turkler* and that of *Magyar* is *Magyars*, not *Magyarok*, so too is the plural of *Rom* in English *Roms* (such integration into native grammar occurs in all the other languages of countries where Roms live, e.g. Macedonian *Rom-Romi*, Albanian *Rom-Romë* (or *Rrom-Rromë*), Turkish *Roman-Romanlar*). The adjective from Rom is Romani.

have migrated “from northeastern Bulgaria at some time in the distant past” (Friedman 2017a: 34). Finally, the speakers of South Vlax, referred to as Džambazi in Skopje, came to North Macedonia from Romanian-speaking territories.

Most linguistic work on Romani in North Macedonia has focused on the Romani of Skopje. The grammars and dictionaries that have been produced locally so far (Kepeski and Jusuf 1980; Demir and Demir 2009ab; Petrovski and Veličkovski 1999; Demir, Djurić, and Demir 2010ab) are all intended mainly for pedagogical purposes. There are no linguistically oriented grammars or dictionaries. Nevertheless, fieldwork on Romani by professional linguists in North Macedonia began in the 1960s, when Austrian linguist Mozes F. Heinschink started making recordings, which are now stored in the Heinschink Sammlung at the Austrian Academy of Sciences (see Fennesz-Juhász 1996). Some of the narratives were later published in the folklore collection (Cech et al. 2009). In the 2000s, several samples were recorded using the Romani morpho-syntax questionnaire, including six recordings from Skopje and Kumanovo, which are available online (<https://romani.dch.phil-fak.uni-koeln.de/>). A glossary of Macedonian Arli, based on the Arli dialects of Skopje and Kumanovo, was also included in RomLex, a lexical database on Romani dialects (<http://romani.uni-graz.at/romlex/>). Three recordings in three dialects made by Zuzana Bodnárová in 2019 are available online as part of the VLAH (Vanishing Languages and Cultural Heritage) commission of the Austrian Academy of Sciences (<https://www.oeaw.ac.at/vlach/collections/romani/>).

Further work on the description of Romani dialects in North Macedonia is needed, especially considering the language shift observed in some communities. In this article, we present some preliminary results of fieldwork on the Romani variety spoken in the easternmost part of North Macedonia, along the border with Bulgaria (Crnik, Delčevo, Vinica). To the best of our knowledge, this variety has not been previously described, although some data have been collected from neighboring locations (for details, see Section 2).

2 Data and research questions

The data for this study were collected during several work sessions with native speakers of the dialect. Upon meeting Senada Lamovska, a native speaker of this dialect, in Skopje in March 2024, Kirill Kozhanov, one of the authors of this paper, observed that although the variety in question belonged to the South Balkan dialect group, it was distinct from the Romani dialects spoken in Skopje. With the help of Senada, who remained the main consultant on the variety, a field trip to the eastern part of North Macedonia was organized in August 2024. Recordings were made in several locations, as indicated in Fig. 1. The description was then continued in Skopje and later via Zoom.

As of now, there are five hours of free narratives in the dialect, as well as four hours of elicitation of lexical and grammatical information recorded from ten speakers. In addition to the recordings, there are fieldwork notes taken during participant observation and unrecorded conversations. These data serve as the basis for the following overview. Currently, the data are stored in the authors’ personal archive and not available online, but we are considering creating an annotated corpus of this variety in the future.

Although no fieldwork has been conducted in the named locations, Mozes Heinschink made recordings in Kočani, a nearby city. Currently, the majority of the Romani community in Kočani has shifted to Turkish. Heinschink’s data from Kočani were later used in Boretzky et al. (2008), where Kočani is included as one of the datapoints. Additionally, some recordings of Romani dialects from the other side of the border, specifically in Blagoevgrad in southwestern Bulgaria, were made by Birgit Igla (two manuscripts from 1996 and 2002 are cited by Boretzky et al. 2008: 68). In 2009, Victor Friedman recorded an interview with a

Romani speaker who was born in Štip but resided in Trabotivište near Delčevo. In the following discussion, we will use the data from Boretzky et al. 2008 for the locations in southwestern Bulgaria.

The speakers of the dialect refer to themselves as *rom* (pl. *romá*) and distinguish themselves from *gadžó* (pl. *gadžé*), the non-Romani-speaking surrounding populations. The non-Romani population can be further specified as *das* (pl. *dasá*), typically Orthodox Christian Macedonian-speaking neighbors, and *gomí* (pl. *gomjá*), usually Muslim neighbors, often with Turkish identity, regardless of home language.² The Turks are also referred to as *xor(x)áj* (pl. *xor(x)ajá*).

When referring to their dialect, the Roms call it Kovački (see an extract from an interview in the Appendix), which should not be confused with the Kovački spoken in Skopje, a North Balkan dialect, nor with the Kovači of the Ohrid region, who are Tosk Albanian speakers of Romani origin. In North Macedonia in general, most Romani speakers (90%) are Muslims.

The same Romani variety is spoken in Crnik, Delčevo, and Trabotivište. In Berovo, according to the language consultants, the Romani community predominantly speaks Macedonian. However, in 2012, during his visit to Berovo, Victor Friedman met many Romani speakers, and at least one recording was in North Balkan Romani; most Roms of Berovo also spoke Turkish. In Vinica, Roms speak both Romani and Turkish. When telling about their parents and grandparents, Roms from these locations mention Krupnik, Blagoevgrad, Simitli, and Sandanski in Bulgaria as locations where they have family. However, since the introduction of the border between North Macedonia and Bulgaria (1912–1941, 1944–present), contacts have become less frequent.

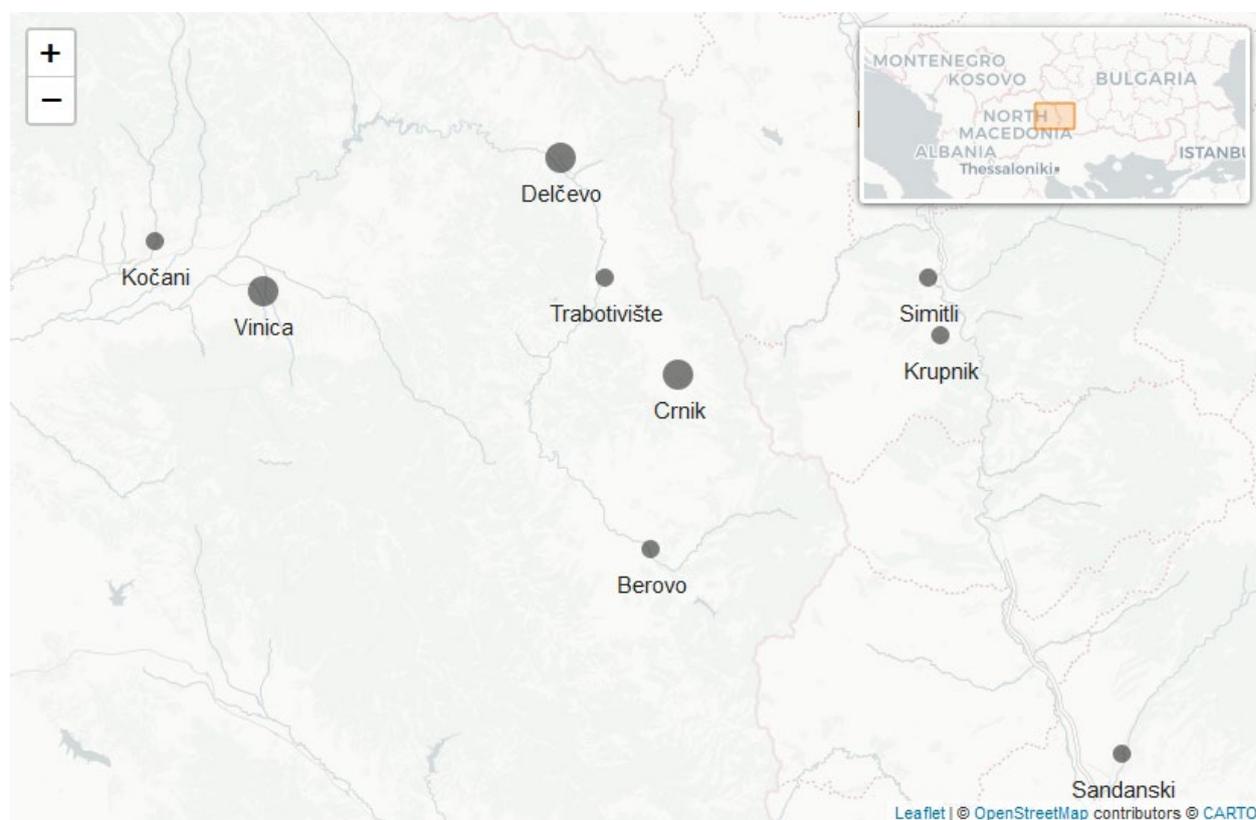


Figure 1. Map of Maleševu Romani: Locations where data were collected are marked by larger circles

² The term *gomí* in such usage is characteristic of Romani from Bosnia (see Igla 2019).

In what follows, we present several diagnostic features of these Romani varieties that help situate them within Romani dialectology. We show that this is a South Balkan dialect that shares several features with the Romani dialects of southwestern Bulgaria. Following the terminology used in South Slavic dialectology (see Friedman 1993), we refer to this variety of Romani as Maleshevo-Pirin Romani. When citing our data from eastern North Macedonia, we use the term Maleshevo Romani.

3 Preliminary results

The dialect in question undoubtedly belongs to the South Balkan dialect group. Importantly, it does not have the Romanian loanwords or sound changes typical of Vlax dialects (e.g., *čh > ś, *dl > gl, *numa(j)* ‘only’), and shows no traces of the sound change *gi-/*di- > zi-, *ki-/*ti- > ci- found in North Balkan dialect group, exemplified by Bugurdži here:

	Maleshevo	Bugurdži (Boretzky 1993)
‘song’	<i>gílí</i>	<i>zi</i>
‘day’	<i>divés</i>	<i>zis, ziés</i>
‘cheese’	<i>királ</i>	<i>cirál</i>
‘work’	<i>butí</i>	<i>bucí</i>

3.1 Phonology

The vowel system of the dialect is typical of Romani and includes the five basic phonemes: /i, e, a, o, u/. The consonant system is quite conservative, retaining the aspirated consonants /kh, ph, th, čh/ and the velars /k, g/, while introducing a few innovations, summarized below. For an overview of Romani phonology, see (Baló 2020).

3.1.1 Reflexes of Proto-Romani *ř and *nř

The reflexes of Proto-Romani *ř and *nř are always rendered as /r/ in this dialect. The development *ř > r is typical of most South Balkan dialects (Boretzky 1999: 28–29), whereas the reflex of *nř exhibits more variation in its realizations (Boretzky 1999: 28–29, map 3; Boretzky et al. 2008: 8–9, maps 3, 4). The *nř > r sound change is characteristic of Arli and Yerli dialects and contrasts with the southern part of the South Balkan dialects, exemplified in North Macedonia by the Prilep dialect, which has -nd- as a reflex of *nř (Boretzky and Cech 2016: 18).

	Maleshevo	Arli (Romlex)	Prilep (Boretzky and Cech 2016)
*ř			
‘Rom’	<i>rom</i>	<i>rom</i>	<i>rom</i>
‘stone’	<i>bar</i>	<i>bar</i>	<i>bar</i>
‘flour’	<i>varó</i>	<i>varo</i>	<i>varo</i>
*nř			
‘bread’	<i>maró</i>	<i>maro</i>	<i>mando</i>
‘egg’	<i>jaró</i>	<i>jaro</i>	<i>vando</i>

3.1.2 Retention of *s

One of the old features distinguishing Romani dialects is the alternation of *s/h* in certain morphological forms, such as the INS.SG of nominal declension, PRS.2SG/1PL in the verbal conjugation, and the initial sound in the copula forms (see Matras 1999). Maleshevo Romani consistently uses only the *-s* reflex in all these positions, which puts it in opposition to many other South Balkan or North Macedonia (Boretzky 1999: 30, map 4; Boretzky et al. 2008: 10, maps 7, 9; Friedman 2017ab). In this regard, our data align with data from Kočani and southwestern Bulgaria. Note that long present forms with *-a* are not typical for Maleshevo Romani.

		Maleshevo	Arli (Boretzky 1996: 8–9, 16–17, 25, Friedman 2017ab)
INS.SG	‘with salt’	<i>lonésa</i>	<i>-eja, -ea, -esa</i>
ACC.SG.M	‘man’	<i>manušés</i>	<i>manuše, manušes</i>
PRS.2SG	‘you do’	<i>kerés</i>	<i>kere, kereja, keres, keresa</i>
PRS.1PL	‘we do’	<i>kerás</i>	<i>kera, keraja, keras, keresa</i>
copula (PRS.3SG)		<i>si, isí</i>	<i>si, isi, hi, i</i>

3.1.3 *t' > k', *Vd' > Vg', *Cd' > Cj

The palatalized dentals *t' and *d' are typically realized as *k'* and *g'* in this dialect, a development also attested in other South Balkan dialects and typical of Macedonian, cf. for Arli (Boretzky 1996: 4) and for Prilep (Boretzky and Cech 2016: 16).

		Maleshevo	Prilep (Boretzky and Cech 2016)
*t, *t'	‘warm’	<i>tató</i>	<i>tato</i>
	‘get warm’	<i>tak'ól</i>	<i>takjol</i>
*d, *d'	‘mind’	<i>godí</i>	<i>godi</i>
	‘smart’	<i>gog'avér</i>	<i>gogja(v)er</i>

Importantly, in Maleshevo Romani, if *d' follows a consonant, it is realized as /j/. This sound change appears primarily in the past tense conjugation of the verbal paradigm with the suffix *-d-* (except for 3PL) and in the derivation of mediopassive verbs from participles, thus, appearing after sonorants. To the best of our knowledge, this change has not been previously attested in South Balkan dialects (cf. Boretzky et al. 2008: 9, map 77).

		Maleshevo	Prilep (Boretzky and Cech 2016)
‘do (PST.3SG)’		<i>kerjás</i>	<i>kerdas</i>
‘do (PST.3PL)’		<i>kerdé</i>	<i>kerde</i>
‘be done (PRS.3SG)’		<i>kerjól</i>	<i>kergjol</i>

3.1.4 *dž > dž or ž(')

The affricate *dž* can be retained but also often undergoes deaffrication and is realized as *ž* or even *ž'* such as *džal* ~ *ž(')al* ‘(s)he goes’. The existing descriptions of South Balkan dialects do not report such a change, and it could be a unique development for Romani in Eastern North Macedonia.

3.1.5 *št > šč

Another unique change in this Romani variety, not reported for other South Balkan dialects, is the optional realization of the cluster *št as /šč/. This development probably replicates the variation in the pronunciation of št ~ šč in the local Maleshevo-Pirin Macedonian dialects (cf. Friedman 1993: 302):

‘four’	<i>ščar</i>	<i>štar</i>
‘wood’	<i>kaščá</i>	<i>kaštá</i>
‘can’	<i>aščí</i>	<i>aští</i>

3.2 Morphology

The morphology of Maleshevo Romani is typical of South Balkan dialects (for an overview, see Boretzky 1999: 37–126; Boretzky et al. 2008: 11–34). In the following discussion, we present several features that form isoglosses within South Balkan dialects and are thus important for determining the position of Maleshevo Romani within this group.

3.2.1 Personal pronouns

Maleshevo Romani has the following system of personal pronouns:

	SG	PL
1	<i>me</i> (obl. <i>man</i>)	<i>amén, amé</i> (obl. <i>amén</i>)
2	<i>tu</i> (obl. <i>tut</i>)	<i>tumén</i> (obl. <i>tumén</i>)
3M	<i>ov</i> (obl. <i>les</i>)	<i>olá</i> (obl. <i>len</i>)
3F	<i>oj</i> (obl. <i>la</i>)	

One of the features that distinguishes the South Balkan dialects is the form of third-person pronouns. Maleshevo Romani represents a unique combination for the Romani dialects of North Macedonia (Boretzky 1999: 56–60, maps 22, 24; Boretzky et al. 2008: 16, maps 32–33).

3.2.2 Borrowed noun accommodation

South Balkan Romani dialects exhibit various suffixes for the accommodation of loan words, a phenomenon that shows considerable diversity across other Romani dialects as well (Elšík 2020: 168–170). In Maleshevo Romani, the following endings are primarily used for the singular and plural accommodation of borrowed nouns:

		SG	PL	
masculine nouns	‘city’	<i>-os</i>	<i>-ja</i>	↔ Mac. <i>град</i>
	‘celebration’	<i>grádos</i>	<i>grádja</i>	↔ Mac. <i>празник</i>
masculine nouns	‘phone’	<i>práznikos</i>	<i>práznikja</i>	↔ Mac. <i>телефон</i>
	‘blacksmith’	<i>-i</i>	<i>-(i)ja</i>	↔ Mac. <i>ковач</i>
	‘student’	<i>telefóni</i>	<i>telefónja</i>	↔ Mac. <i>студент</i>
feminine nouns	‘student’	<i>kováči</i>	<i>kováčja</i>	↔ Mac. <i>балтија</i> (↔ Turk. <i>balta</i>)
	‘axe’	<i>studénti</i>	<i>studéntija</i>	↔ Mac. <i>фамилија</i>
feminine nouns	‘family’	<i>-a</i>	<i>-es</i>	
	‘axe’	<i>baltíja</i>	<i>baltíjes</i>	
	‘family’	<i>famílíja</i>	<i>famílíjes</i>	

Borrowed masculine nouns receive the suffixes *-os* (pl. *-ja*) or *-i* (pl. *-(i)ja*). The distribution is phonetic: stems ending in *-v-*, *-f-*, *-r-*, *-l-*, *-m-*, *-n-*, *-t-*, *-č-* typically take the ending *-i*, while the rest employ the suffix *-os*.

When comparing the distribution of accommodation suffixes with other South Balkan dialects (cf. Boretzky 1999: 52–54, maps 19, 20; Boretzky et al. 2008: 12–13, maps 15–17), Maleshevo Romani presents a contradictory picture, highlighting its distinct position among the Romani dialects of North Macedonia. On the one hand, it differs from other Romani dialects of North Macedonia and instead aligns with the dialects of southwestern Bulgaria. This is evident in the singular suffix *-os* of masculine nouns (Boretzky et al. 2008: 12–13, map 6) and the plural suffix *-es* of feminine nouns (Boretzky et al. 2008: 12–13, map 17). This is consistent with other languages in the region, where the dialects of eastern North Macedonia align with those of Pirin Macedonia (southwestern Bulgaria). On the other hand, the accommodation suffixes *-i* (pl. *-(i)ja*), found in this dialect, is typical of other Romani dialects in North Macedonia, but is also present in the bordering Romani dialects of western Bulgaria (see Boretzky et al. 2008: 13, map 16).

3.2.3 PRS.3SG suffix **-el > -ol*

Maleshevo Romani has typically the PRS.3SG ending *-ol* (vs. Common Romani *-el*), e.g., *dikhól* ‘see’, *khelól* ‘dance’, *sikavól* ‘teach’, *phúdol* ‘blow’ etc. This variant *-ol* is present alongside *-el* in Arli (Boretzky 1996: 18), and in Prilep, *-ol* seems to be a dominant variant (Boretzky 1999: 81, map 42; Boretzky and Cech 2016: 45). According to Boretzky (1999: map 42), the *-ol* change in PRS.3SG is attested only in the South Balkan dialects of North Macedonia and not in those of Bulgaria or Greece.

Since the change **-el > -ol* is not observed in *khél* ‘dance (IMP.2SG)’ (another diagnostic word, **šél* ‘100’, is obsolete in Maleshevo Romani), this innovation must be morphological—the suffix *-ol* spreads from the paradigm of intransitive or passive verbs, as in *susl’ól* ‘get wet’, *kerjól* ‘get done’ etc., to other verbs. On the other hand, a similar change is attested in the noun *dovól* ‘God’ (vs. Common Romani *devél*), but it is probably an unrelated phonetic innovation triggered by the unique combination **-eve-*.

In an earlier description of South Balkan dialects, Boretzky (1999: 26) analyzes, most likely incorrectly, cases such as *sovol* ‘sleep’, *dol* ‘God’ in Gnjilane Arli as a result of Vlax influence.

3.2.4 Verb conjugation in the past tense

Maleshevo Romani has the following past tense endings:

	SG	PL
1	<i>-om</i>	<i>-am</i>
2	<i>-an</i>	<i>-en</i>
3	<i>-as</i>	<i>-e</i>

Two past tense suffixes are relevant to the variation within South Balkan dialects: the 1SG and 3SG suffixes. Maleshevo Romani PST.1SG suffix *-óm*, as in *dikh’óm* ‘I saw’, *kerjóm* ‘I did’ etc., is also attested in the Romani dialects of southwestern Bulgaria. In contrast, other South Balkan dialects of North Macedonia commonly have *-um* (Boretzky 1999: map 43; Boretzky et al. 2008: 27, map 79).

The PST.3SG suffix forms an East-West divide within South Balkan dialects. In the South Balkan dialects of North Macedonia, the typical suffix is *-a*, whereas *-ás* is present in the

Romani dialects of Bulgaria, including its southwestern part (Boretzky et al. 2008: 28, map 8, 80).

In Romani, past tense formation involves not only personal endings but also different past stem suffixes, which are distributed across verbal paradigms (for an overview of Romani verbal inflection, see Elšík 2020: 159–163). South Balkan dialects exhibit considerable variation in past tense stems and suffixes used by particular verbs (Boretzky et al. 2008: 63, maps 63–76). Some relevant verbs are exemplified in Maleshevo Romani in contrast to the Prilep dialect:

	Maleshevo	Prilep (Boretzky and Cech 2016)
‘give’	<i>dinj-</i> (3PL <i>dindé</i>)	<i>den-</i> , <i>dend-</i>
‘rise’	<i>uštinj-</i> (3PL <i>uštindé</i>) <i>uštindil’-</i> (3PL <i>uštindile</i>)	<i>uštind-</i>
‘go out’	<i>iklístil’</i> (3PL <i>uklístile</i>)	<i>iklist-</i>
‘laugh’	<i>asanj-</i> (3PL <i>asandé</i>) <i>asándil’-</i> (3PL <i>asándile</i>)	<i>asand-</i> , <i>asandil-</i>

Although there are similarities in the past stem suffixes found in the South Balkan dialects of North Macedonia, Maleshevo Romani occupies again a unique position. It differs not only from other Romani dialects of North Macedonia but also from the available data on Romani dialects in southwestern Bulgaria (see maps 63, 69, 70, 74 in Boretzky et al. 2008 for the verbs ‘give’, ‘rise’, ‘go out’, and ‘laugh’ respectively).

3.3 Lexis

Several lexical isoglosses (including the choice of lexemes but also phonetic changes typical of specific lexemes) create a West-East divide among South Balkan dialects.

	Maleshevo	Arli (Romlex)	Prilep (Boretzky, Cech 2016)
‘can’	<i>ašti</i> , <i>aščí</i>	<i>šaj</i>	<i>možin-</i>
‘shoe’	<i>tiráx</i>	<i>kundra</i>	<i>kundura</i>
‘iron’	<i>sas</i> (obl. <i>sases-</i>)	<i>sastrn</i>	<i>sastrîn</i> , <i>sastêrn</i>
‘name’	<i>aláv</i>	<i>anav</i>	<i>anav</i>
‘chicken’	<i>khajní</i>	<i>khani</i>	<i>khajni</i>
‘small’	<i>tiknó</i>	<i>tikno</i>	<i>tikno</i>

As shown, the lexemes ‘can’, ‘shoe’, ‘iron’, and ‘name’ differ from those found in the South Balkan dialects of North Macedonia. However, according to Boretzky et al. (2008: maps 92, 120, 118, 140), the same variants are attested in the Romani dialects of southwestern Bulgaria.

Conversely, when comparing maps 122 and 153 in Boretzky et al. (2008), the Eastern Macedonian data align with other Romani dialects of North Macedonia and differ from those on the Bulgarian side, which have *kaxni* and *cikno*, respectively.

3.4 Borrowing

In addition to the lexicon shared by most varieties of Romani, this dialect features several possibly late Greek loanwords typical of South Balkan dialects but absent in other dialect groups (cf. *prandél* ‘marry’, *naborénol* ‘become sick’). It also includes more recent borrowings from Turkish and dialectal Macedonian (specifically the Maleshevo-Pirin dialect).

The influence of local Macedonian is evident in several loanwords. For instance, *bóbos* ‘beans’ is likely borrowed from the local Macedonian dialect, which has *боб*, in contrast to Standard Macedonian *зрав* (MDA 2024: map 46). Similarly, *dúma* ‘word’ corresponds to Eastern Macedonian *дума*, whereas Standard Macedonian uses *збор*.

In the following discussion, we present several examples of further borrowings from Macedonian.

3.4.1 Borrowing of pronouns

The inherited Romani interrogative pronoun *kon* ‘who’ is replaced in the direct form by the Macedonian *koj*, while the Romani declension is retained, as seen in *kas* ‘whom (ACC)’, *káske* ‘(DAT)’ etc. This replacement is also attested in the Arli and Prilep dialects of Romani (Boretzky 1999: 67).

In Maleshevo Romani, as elsewhere, the system of indefinite and negative pronouns is fully borrowed from Macedonian, cf. *něšto* ‘something’, *níšto* ‘nothing’, *nikade* ‘nowhere’ etc.

3.4.2 Borrowing of prepositions

Maleshevo Romani employs several prepositions, the most frequent being *ki* (*ko* before masculine nouns) ‘at, in, on’ and *tari* (*taro* before masculine nouns) ‘from’. These prepositions can also be used to mark dative/locative and ablative with nouns, which is a general tendency in the Romani of North Macedonia.

Additionally, this dialect has borrowed several Macedonian prepositions. The prepositions *u* ‘in’ (again, from the local Macedonian dialect) and *od* ‘from’ appear only with place names, as in (1).

- (1) *u Skópje da phir-j-án pal ko rom-á?*
 in Skopje also walk-PST-2SG across DAT Rom-PL
 ‘did you also go to Roma in Skopje?’

Other Macedonian prepositions commonly used in this dialect include *za* ‘for’, *prez* ‘during’, *pred* ‘before’, *protiv* ‘against’, cf. (2–3), where the Macedonian preposition *za* governs the dative form of the Romani personal pronoun and a singular noun, respectively.

- (2) *nanáj bút vréme ili za túke*
 NEG.be.PRS.3SG much time and for 2SG.DAT
 ‘there is not much time [left] for you too’
- (3) *prandesál-i za mi papús-ke*
 get_married.PST-3SG.F for my grandfather-DAT.SG
 ‘she got married to my grandfather’

3.4.3 Borrowing of Macedonian preverbs

Several Macedonian preverbs, such as *iz-*, *po-*, *pre-*, *za-*, are borrowed and used with inherited Romani verbs, cf. (4), where the Macedonian preverb *iz-* is used with the Romani verb *naš-* ‘run’ (cf. the Macedonian verb *узбегза* as the source of this derivation).

- (4) *lésker-i čháj iz-naš-l-ás u avstrálija*
 his-SG.F daughter.NOM.SG PVB-run-PST-3SG in Australia
 ‘his daughter ran away to Australia’

3.4.4 Borrowing of the Macedonian comparative prefix *po*

A typical Balkan Slavic influence, found in many South Balkan dialects of Romani but absent in other Romani dialects outside the Balkans, is the use of the prefix *po-* to derive comparative adjectives and adverbs (Boretzky 1999: 55), as illustrated in (5).

- (5) *garáv-tut po-telé*
 hide.IMP.SG-2SG.RFL COMP-low
 ‘hide lower’

3.4.5 Borrowing of object indexes

An interesting feature of this dialect, which, to the best of our knowledge, has not been previously reported, is the sporadic use of Macedonian object indexes with Romani verbs, as in (6), where *-go* is an object index on the verb (traditionally referred to as an object clitic).

- (6) *dikhl’óm-go*
 see-PST.1SG-OBJ.3SG.M
 ‘I saw him.’

According to our data, the borrowed object indexes *-go* and *-gi* are commonly used in alternation with the Romani personal pronoun forms *les* and *len*, respectively, whereas the feminine singular pronominal object is represented only by the Romani pronoun *la*. As far as our data can tell, the borrowed index is not used co-referentially with another noun in the same noun phrase (i.e. as “clitic doubling”).

4 Discussion

The Romani dialect of eastern North Macedonia (the Maleshevo region) belongs to the South Balkan dialect group and exhibits several distinguishing features that set it apart from other South Balkan Romani dialects spoken in North Macedonia—the Arli dialects of Skopje and Kumanovo, as well as the Prilep variety. As shown in this study, several features of this dialect are shared with the Romani dialects spoken in the neighboring region of southwestern Bulgaria.

The common development of Romani dialects in eastern North Macedonia and southwestern Bulgaria is further supported by close family ties between Romani communities in this area. The special position of these Romani dialects corresponds to the dialect division of South Slavic (the Maleshevo-Pirin dialects). At the same time, other features of this Eastern Macedonian Romani are either unique or shared with the South Balkan dialects of North Macedonia. This places it in a special position, situated on the border of the East-West divide within South Balkan Romani dialects (Boretzky et al. 2008: 47–48; cf. also Sechidou 2011: 89–95).

The dialect is in close contact with the local Turkish and Macedonian dialects. Further documentation is needed, as language shift to Turkish or Macedonian is currently observed in Romani communities of this region.

Text in Maleshevo Romani

Recorded by Kirill Kozhanov in Crnik, North Macedonia, from a male Romani speaker, born in Crnik in 1957.

amé sijám akaná od crnik, ama mir-ó páp-us,
 1PL.NOM be.PRS.1PL now from Crnik but my-NOM.SG.M grandfather-NOM.SG
amé, amar-ó potékl-os sí od délčevo
 1PL.NOM our-NOM.SG.M origin-NOM.SG be.PRS.3 from Delčevo
 ‘we are now from Crnik, but my grandfather, us, our origin is from Delčevo’

[KK: *soske ale athe?* ‘why did they come here?’]

athé al-ó za me nen-áke.
 here come.PST-3SG.M for my grandmother-DAT.SG
 ‘he came here because of my grandmother’
a inače aménde, amar-ó potékl-os sí taro délčevo,
 but otherwise 1PL.LOC our-NOM.SG.M origin-NOM.SG be.PRS.3 from Delčevo
cárevo sélo vaker-ól-as-pes pó-rano, xoraj-á vaker-én-as les, e.
 Carevo Selo call-3SG-IMP-F RFL COMP-early Turk-NOM.PL call-3PL-IMPF 3SG.ACC.M, yeah
 ‘But otherwise, our origin is from Delčevo. It used to be called Carevo Selo—that’s what the Turks called it, yeah’

[KK: *a ko delčevo but roma živinenas?* ‘Did many Roma live in Delčevo?’]

pa i akaté sí, enjavardeš kherá isí, romá.
 so and here be.PRS.3 ninety house-NOM.PL be.PRS.3 Rom-NOM.PL
sámo o terné gelé po... germánija ker-én but-í
 only ART young-NOM.PL go.PST-3PL to Germany do-PRS.3PL work-ACC.SG
othé za-astar-d-é, isí búť mir-í familij-a,
 there PVB-catch-PST-3PL be.PRS.3 much my-NOM.SG.F family-NOM.SG
amén sijám lésa škól’sk-a drugár-ja
 1PL.NOM be.PRS.1PL 3SG.INS.M school-NOM.PL friend-NOM.PL
othé amar-ó dialékt-i povéče čalav-ól ko dasikan-ó,
 there our-NOM.SG.M dialect-NOM.SG more hit-PRS.3SG to Macedonian-NOM.SG.M
ko gadž..., gadžikan-ó dialékt-i.
 to non-Romani-NOM.SG.M dialect-NOM.SG
 ‘Well, there are 90 Roma houses here. Only the young ones went to Germany to work and settled there. I have a lot of family there. He and I are school friends. Our dialect is closer to Macedonian, to the non-Roma, non-Romani dialect’

[KK: *athe ili othe?* ‘here or there?’]

akaté.
 ‘Here’.

[KK: *a ko delčevo sar?* ‘and how is it in Delčevo?’]

ísto, amar-ó dialékt-i. Máľku sí, máľku isí rázlik-a
 same our-NOM.SG.M dialect-NOM.SG little be.PRS.3 little be.PRS.3 difference-NOM.SG
 ‘It’s the same, it’s our dialect. There is just a small difference’

[KK: *naprimer so?* ‘what, for example?’]

pa naští mé te vaker-áv tú-ke só... avká si xári neprimétlivo.
 so cannot.PRS 1SG.NOM SBJ tell-1SG 2SG-DAT what this be.PRS.3 little inconspicuous
athé, bérovo samo isí, avér čhib mátku ili othé, xári,
 here Berovo only be.PRS.3 other language.NOM.SG little or there little
ama isto amar-ó kováčk-o dialékt-i si,
 but same our-NOM.SG.M Kovačko-NOM.SG dialect-NOM.SG be.PRS.3
čalav-ól ko gadžikan-ó dóka.
 hit-PRS.3SG to non-Romani-SG.M that
amar-é phur-é ačká sikav-d-é amén, i ačká al-ó.
 our-NOM.PL old-NOM.PL this.way teach-PST-3PL 1PL.ACC and this_way come.PST-3SG.M
 ‘I can’t tell you what... It’s somewhat inconspicuous. Here, only in Berovo the language is a
 bit different, but it’s the same our Kovačko dialect, it goes to the non-Romani. Our old ones
 taught us so, and this is how it came’

[KK: *a soske vakeren kovačko dialekti?* ‘and why is it called Kovačko dialect?’]

soske? naprimer, athé sá siné kováč-ja, e rom-á.
 why for_example here all be.PST3 blacksmith-NOM.PL ART Rom-NOM.PL
léskor-o dád, mir-ó pápus,
 his-NOM.SG.M father.NOM.SG my-NOM.SG.M grandfather.NOM.SG
amén sijám da bašaln-é, bašal-ás-as,
 1PL.NOM be.PRS.1PL also musician-NOM.PL play.music-1PL-IMPF
čalgidžija, m-o dád siné čalgidžij-es, mi familij-a
 musician my-SG.M father.NOM.SG be.PST.3 musician-NOM.PL my-SG.F family-NOM.SG
cél-o.
 whole-NOM.SG

‘Why? For example, all the Roma here were blacksmiths. His father, my grandfather. And we
 are also musicians. We played music, my father was a musician, and my whole family.’

Bibliography

- Baló, M. A. (2020). Romani Phonology. In Y. Matras and A. Tenser (eds.). *The Palgrave handbook of Romani language and linguistics*, 119–153. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28105-2_5
- Beníšek, M. (2020). The historical origins of Romani. In Y. Matras and A. Tenser (eds.). *The Palgrave handbook of Romani language and linguistics*, 13–47. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28105-2_2
- Boretzky, N. (1993). *Bugurdži: Deskriptiver und historischer Abriss eines Romani-Dialekts*. Berlin: Harrassowitz.
- Boretzky, N. (1996). Arli: Materialien zu einem südbalkanischen Romani-Dialekt. *Grazer Linguistische Studien*, 46: 1–31.
- Boretzky, N. (1999). *Die Verwandtschaftsbeziehungen zwischen den südbalkanischen Romani-Dialekten: Mit einem Kartenanhang*. Frankfurt am Main: Lang.
- Boretzky, N. (2000). South Balkan II as a Romani dialect branch: Bugurdži, Drindari, and Kalajdži. *Romani Studies* 10(2): 105–183. <https://doi.org/10.3828/rs.2000.5>
- Boretzky, N. (2013). Lexikalische Slavismen im Romani. *Zeitschrift für Balkanologie*, 49: 10–46.
- Boretzky, N., Cech, P., and Iгла, B. (2008). *Die südbalkanischen Dialekte (SB I) und ihre innere Gliederung: Analyse und Karten*. Graz: Institut für Sprachwissenschaft, Universität Graz.

- Boretzky, N. and Cech, P. (2016). *Der Romani-Dialekt von Prilep/Makedonien*. Graz: Karl-Franzens-Universität Graz.
- Cech, P., Heinschink, M. F., and Halwachs, D. W. (eds.). (2009). *Kerzen und Limonen. Momelja hem limonja. Märchen der Arlije*. Klagenfurt/Celovec: Drava Verlag.
- Demir, L. M. and Demir, F. (2009a). *Gramatika e romane čhibakiri*. Skopje: Romano kulturakoro thaj edukativakoro centro “Darhia”.
- Demir, L. M. and Demir, F. (2009b). *Gramatika na romskiot jazik*. Skopje: Romski kulturen i edukativen centar “Darhia”.
- Demir, L. M., Djurić, R. and Demir, F. (2010a). *Lavustik romani-makedonikani*. Skopje: Romano kulturakoro thaj edukativakoro centro “Darhia”.
- Demir, L. M., Djurić, R. and Demir, F. (2010b). *Rečnik makedonsko-Romani*. Skopje: Romski kulturen i edukativen centar “Darhia”.
- Elšík, V. (2020). Romani morphology. In Y. Matras and A. Tenser (eds.). *The Palgrave handbook of Romani language and linguistics*, 155–186. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28105-2_6
- Elšík, V. and M. Beníšek. (2020). Romani dialectology. In Y. Matras and A. Tenser (eds.). *The Palgrave handbook of Romani language and linguistics*, 389–427. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28105-2_13
- Fennesz-Juhász, C. (1996). Tondokumente europäischer Roma. Die „Sammlung Heinschink“ im Phonogrammarchiv der Österreichischen Akademie der Wissenschaften. In J. S. Hohmann (ed.), *Handbuch zur Tsiganologie*, 311–320. Frankfurt: Lang.
- Friedman, V. A. (1985). Balkan Romani modality and other Balkan languages. *Folia Slavica*, 7: 381–389.
- Friedman, V. A. (1993). Macedonian. In B. Comrie and G. G. Corbett (eds.). *The Slavonic languages*, 249–305. London/New York: Routledge.
- Friedman, V. A. (1999). The Romani language in the Republic of Macedonia: Status, usage, and sociolinguistic perspectives. *Acta Linguistica Hungarica*, 46(3–4): 317–339.
- Friedman, V. A. (2000). Romani in the Balkan Linguistic League—Synchrony and diachrony. In Ch. Tzitzilis and Ch. Symeonides (eds.). *Valkanikē Glossologia: Synkhronia kai diakhronia. Balkanlinguistik: Synchronie und Diachronie*, 95–105. Thessaloniki: University of Thessaloniki.
- Friedman, V. A. (2017a). Seven varieties of Arli: Skopje as a center of convergence and divergence of Romani dialects. *Romani Studies*, 5(27): 29–45. <https://doi.org/10.3828/rs.20172>
- Friedman, V. A. (2017b). The Arli of Skopje: A perceptual dialectological approach. In K. Kozhanov, M. Oslon and D. W. Halwachs (eds.). *Das amen godi pala Lev Čerenkov. Romani historija, čhib taj kultura*, 210–220. Graz: treffpunkt sprachen – Forschungsbereich Plurilingualismus.
- Igla, B. (2019). Balkanski paraleli pri zaemane ot romski: *zǎdže* < *gadžo* i *γκόμεινα* < *gomí*. *Sāpostavitelno Ezikoznanie*, 44(1): 37–44.
- Kepeski, K. and Jusuf, Š. (1980). *Romani gramatika – Romska gramatika*. Skopje: Naša knjiga.
- Kozhanov, K. and Oslon, M. (2020). Romani and Slavic contact. In M. L. Greenberg (ed.). *Encyclopedia of Slavic languages and linguistics online*. Brill. https://doi.org/10.1163/2589-6229_ESLO_COM_032594
- Marushiakova, E. and Popov, V. (2001). *Gypsies in the Ottoman Empire: A contribution to the history of the Balkans*. Hatfield: University of Hertfordshire Press.
- Matras, Y. (1994). Structural Balkanisms in Romani. In N. Reiter (ed.). *Sprachlicher Standard und Substandard in Südosteuropa und Osteuropa*, 195–210. Wiesbaden: Harrassowitz.
- Matras, Y. (1999). *s/h* alternation in Romani: An historical and functional interpretation. *Grazer Linguistische Studien*, 51: 99–129.
- Matras, Y. (2002). *Romani: A linguistic introduction*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486791>
- Matras, Y. (2005). The classification of Romani dialects: A geographic-historical perspective. In D. W. Halwachs, B. Schrammel and G. Ambrosch (eds.), *General and applied Romani linguistics*, 7–26. Munich: Lincom Europa.
- Matras, Y., and A. Tenser (eds.). (2020). *The Palgrave handbook of Romani language and linguistics*. Cham: Palgrave Macmillan.

- Oslon, M. and Kožanov, K. (in prep.). *Ėtimologičeskij slovar' cyganskogo jazyka (obščecyganskaja leksika)*. <https://rromanes.org/pub/ЭЦЦЯ.pdf> (last access on 25 March 2025).
- Petrovski, T. and Veličkovski, B. (1999). *Makedonsko-romski i romsko-makedonski rečnik*. Skopje: Vorldbuk.
- Scala, A. (2020). Romani Lexicon. In Y. Matras and A. Tenser (eds.). *The Palgrave Handbook of Romani Language and Linguistics*. Palgrave Macmillan: Cham. https://doi.org/10.1007/978-3-030-28105-2_4
- Sehidou, I. (2011). *Balkan Romani: The dialect of Ajos Athanasios/Greece*. Munich: Lincom Europa.

CONTACT-INDUCED CHANGES IN THE ORDER OF MODIFYING ADJECTIVES AND NOUNS IN SLAVIC DIALECTS IN ALBANIA¹

Maxim Makartsev

Carl von Ossietzky University Oldenburg
maxim.makartsev@uni-oldenburg.de

In this article, I investigate the sociolinguistic factors that influenced the word order in nominal phrases in the South Slavic dialects that are spoken in Albania, with a specific focus on the impact of Albanian. Utilizing data from the Corpus of Slavic Dialects in Albania, I examined variations in adjective placement (pre- or post-noun) across several dialects – Golloborda Macedonian (GM), Korça Macedonian (KM), Prespa Macedonian (PM), and migrational Štokavian dialects from Shijak (SŠ) and Myzeqe (MŠ). The analysis revealed that the Albanian-influenced Noun-Adjective order (NADJ) was increasing in these dialects, and was mainly influenced by age, gender, residence type (rural versus urban), and type of community dwelling (compact versus dispersed). The stability of the conservative Adjective-Noun (ADJN) order in compact communities and older generations in contrast to the increased Albanian-influenced NADJ usage among younger speakers and by those in dispersed communities, was particularly notable. The influence of standard Slavic languages may reverse the process by supporting the conservative ADJN word order. These findings contribute to the research on the areal spread of word order patterns.

Keywords: word order variation, sociolinguistics, corpus research, Balkan languages.

¹ The article is based on the Corpus of Slavic Dialects in Albania (Makartsev and Arkhangelskiy 2024), which was created within the project “Contact-induced language change in situations of non-stable bilingualism—Its limits and modelling: Slavic (social) dialects in Albania” funded by the DFG (German Research Foundation), project number 8750/1-1 (October 16th, 2019–April 30th, 2024).

ВЛИЈАНИЕТО НА ЈАЗИЧНИОТ КОНТАКТ ВО РЕДОСЛЕДОТ НА ПРИДАВКИТЕ И НА ИМЕНКИТЕ ВО СЛОВЕНСКИТЕ ДИЈАЛЕКТИ ВО АЛБАНИЈА

Максим Макарцев

Универзитет „Карл фон Осиецки“ во Олденберг
maxim.makartsev@uni-oldenburg.de

Во статијата се истражуваат социолингвистичките фактори што влијаат врз редоследот на зборовите во именските фрази во јужнословенските дијалекти во Албанија, со посебен фокус на влијанието на албанскиот јазик. Користејќи ги податоците од Корпусот на словенските дијалекти во Албанија, во истражувањето се анализира варијантноста на позицијата на придавките (пред или по именката) во неколку дијалекти – голобрдски македонски, корчански македонски, преспански македонски и миграциони штокавски дијалекти од Шијак и од Музекија. Анализата покажува дека албанскиот јазик влијае врз зголемената употреба на редоследот именка–придавка (ИП) во овие дијалекти, зависно од возраста, полот, видот на населеното место (рурално или урбано) и од типот на заедницата (компактна или дисперзирана). Особено е стабилен конзервативниот редослед придавка–именка (ПИ) во компактните заедници и кај постарите генерации, за разлика од зголемената употреба на албанскиот редослед ИП кај помладите говорители и кај оние во дисперзираните заедници. Влијанието на стандардните словенски јазици се рефлектира врз тенденцијата за зачувување на конзервативниот редослед. Резултатите од истражувањето придонесуваат за подлабоко разбирање на факторите што влијаат на промените во редоследот на зборовите во балканскиот јазичен контекст.

Клучни зборови: варијација во редоследот на зборовите, корпусно истражување, социолингвистика, балкански јазици.

1 Introduction

In this article, I examine the influence of sociolinguistic variables on the outcomes of language contact between Albanian and South Slavic, specifically in the domain of word order within nominal phrases (NPs). The data for this study were drawn from several South Slavic dialects that are spoken in Albania, which exhibit sociolinguistic variation due to different types of contact situations concerning both standard and dialectal varieties of Albanian, and which are regularly encountered by their speakers.

1.1 Problem

Since the dominant word order in Albanian is Noun-Adjective (hereafter NADJ) and is Adjective-Noun (ADJN) in pre-contact South Slavic, this contrast provides a clear, quantifiable parameter that is suitable for investigation via quantitative methods. At the same time, the word order within the NP (as well as potential contact-induced changes in this domain) plays a crucial role in several morphosyntactic phenomena, including the distribution of clitics and the placement of articles (Friedman and Joseph 2025, 793–794). It also contributes to broader morphological restructuring in certain contact settings, such as the loss of gender distinctions in adjectives in the Albanian dialect of Mandrica, Bulgaria (Asenova 2018: 33–34, see below). Consequently, the study of contact-induced changes in NP word order is relevant for areal and comparative research on the morphosyntax of Balkan languages.

Moreover, this analysis serves as a case study of the areal diffusion of contact-induced innovations in word order. Existing studies have often approached such diffusion from the macro-level, treating distinct Abstand languages as single points without accounting for the territorial or social variations within them.² However, integrating micro-level variation into this framework enhances our understanding of how such features spread among contact varieties.

In typology, dominant word order can be defined either in a strong sense as “the only order possible”, or in a weak sense as “the order that is more frequently used” (Dryer 2013b). In this study, I adopt the weak interpretation, as word order within NPs in both Albanian and South Slavic may be influenced by pragmatic factors. Given the absence of corpus-based grammars for the relevant standard languages and contact varieties, the estimation of dominant word orders in this study relies on claims made in the existing descriptive grammars. The following examples illustrate the word order in adjectival NPs in standard Albanian and standard Macedonian:

(1) Albanian, dominant NADJ

<i>djal-i</i>	<i>i</i>	<i>Mire</i>
boy(M)-M.NOM.DEF	L.M.NOM.SG	Good
‘the good boy’ ³		

² See Dryer (2013a) for word order areality in the languages surveyed in the *World Atlas of Language Structures*.

³ I followed the Leipzig glossing rules in the examples. If no source is provided, the source was based on my personal knowledge and was verified by native speakers.

(2) Albanian, marked ADJN

i mir-i Djalë
 L.NOM.M.SG good-M.NOM.DEF boy(M)
 ‘the good boy’

(3) Macedonian, dominant ADJN

sin-o-to ezero
 blue-N.SG-DEF.N.SG lake(N).SG
 ‘the blue lake’

(4) Macedonian, marked NADJ

ezero-to sino
 lake(N).SG-DEF.N.SG blue-N.SG
 ‘the blue lake’

Examples such as (1) and (3) are described in the grammars of the standard languages as (pragmatically) neutral, non-emphasized, most frequent, and so on, and are therefore considered to represent the dominant word order for the respective languages. By contrast, examples such as (2) or (4) are characterized as being expressive, pragmatically marked, and less frequent; thus, they are regarded as being non-dominant.⁴

Descriptions of Macedonian and Albanian dialects within RN Macedonia and Albania typically provide very limited, if any, information about syntax and apply a differential approach based on the respective standard languages. They rarely have focused on NP syntax, except in cases in which it deviates significantly from the standard.

The NADJ word order is often reported in the Slavic dialects that are in contact with Albanian (see details in §1.2) as being more frequent than in the respective standard languages, although this assessment appears to be based primarily on researchers’ perceptions. However, I found no mention of these frequencies having been measured systematically in the literature. The conservative ADJN word order, see (5) for Golloborda Macedonian (GM) and (3) for standard Macedonian, remains widespread in the contact varieties, although the contact-supported NADJ, see (6) for GM and (4) for standard Macedonian, appears to be on the rise.

(5) GM,⁵ ADJN (conservative)

E pe'nuška e'sti sta'r-o dre'o
 Uhm log(F).SG be.PRS.3SG old-N tree(N).SG
me'tn-at-o vo vo'da-na, e'ne go.
 throw-NTPTCP-N in water(F).SG-DIST.F.SG there it.ACC
 ‘It’s a log, it’s an old tree thrown into water, that’s it.’ (Tre30).⁶

⁴ See the comments in Domi (2002: 103–104) and Minova Gjurkova (1994: 92, 139) as well as the details regarding the NP syntax in standard Macedonian in Topolinjska (2008), Marković (2008).

⁵ The nomenclature for the Slavic dialects used in this article does not imply any claims regarding the ethnic, national, or political identity of the speakers. It is solely employed for orientation within the Macedonian and Štokavian-based dialectological traditions.

⁶ The list of abbreviations for speakers with the relevant sociolinguistic information is included in the Corpus of Slavic Dialects in Albania (Makartsev and Arkhangelskiy 2024). They can also be found in the online supplements to this article.

(6) GM, NADJ (contact-supported by Albanian)

A'ma e zna'j-š, e'no dre'o sta'r-o

But she.ACC know-PRS.2SG one.N tree(N).SG old.N

me'tn-at-o ka'ko se' vel-it deme'k.

throw-NTPTCP-N how REFL say-PRS.3SG that.is

‘But you know it, an old thrown tree, that is, how do you call it.’ (Tre30).

The focus in this article is exclusively on the word order of adjectives and common nouns within NPs, excluding proper nouns. In addition, non-adjectival word classes that exhibit adjectival morphology in South Slavic, such as certain types of pronouns, ordinal numerals, and participles with an attributive function, have been excluded.

In the language varieties that are considered in this article, the adjective agrees with the nominal nucleus in terms of gender and number in Albanian and Macedonian, as well as in case in Štokavian. Some of the Macedonian dialects that were analyzed retain a rudimentary case system. Both Albanian and Macedonian feature a postposed definite article, which functions as an enclitic and attaches to the first constituent of the NP, as exemplified by DEF in the examples (1)–(4).

In Albanian, case marking is either integrated with the definite article within the same inflection – the so-called “definite declension” – which can be cliticized when the adjective is fronted – see the postposed adjective in the dominant word order (1) and the fronted adjective in (2) – or, in the case of non-definite forms, is expressed as an ending on the nominal nucleus (in the latter case, the fronting of the adjective is impossible in Albanian).

In the sampled Štokavian dialects, definiteness is expressed solely on the adjective through a distinct declension type.⁷ In addition, Albanian features the so-called linking article (L), a proclitic that attaches to certain adjectives. Despite its formal simplicity – it only occurs in four forms, *i*, *e*, *të*, and *së*, it imposes various morphological and syntactic constraints – see examples (1) and (2) for its positioning. The parameters that are relevant to the analysis will be introduced when needed throughout the discussion.

1.2 Previous research

Previous research on word order in NPs has highlighted areality as a potential factor in its diffusion (Dryer 2013a), although little is known about the specific conditions that facilitate or restrict this process.

In the absence of a comprehensive study of contact-induced changes in word order in the languages of the Balkans, I present the following preliminary list of contact situations in which changes in NP word order are mentioned in the academic literature, with a particular focus on the extended multilingual zone of Albanian-Romance-(South) Slavic language contact.

NADJ > ADJN

1) Albanian dialects in Bulgaria (Mandrica) and Ukraine under the Slavic influence (Morozova 2016: 461; Asenova 2018: 33–34).

2) Meglenoromanian under the Macedonian influence (Friedman and Joseph 2025, 793–794, see also the references therein).

⁷ One of the speakers (Rre04) exhibited article-like uses of postposed demonstrative pronouns, presumably due to the influence of Aegean Macedonian. However, these uses had not reached the threshold of grammaticalization as a definite article (Makartsev in press [a]).

3) Aromanian under the Greek influence: specifically, based on adjectives that have been borrowed from Greek (Bara et al. 2005: 230). The ADJN word order is increasing in Ohrid Aromanian (Marković 2007: 87).

ADJN > NADJ⁸

4) Bulgarian and other South Slavic dialects that have long been in contact with Romanian (Mladenov 1993: 385–386; Asenova 2018: 33); specifically, Banat Bulgarian (Tiševa 2007: 159); the postposition of the attribute would be a distinctive feature of the so-called “Banat sprachbund” proposed by Stojkov in 1968), in the Svinica (Tomić 1984: 71) and Krashovani (Konior 2021: 166, 173) dialects.

5) Some Western Macedonian dialects in the contemporary territory of North Macedonia are reported to exhibit an increased use of the NADJ word order. Koneski (1986: 126–127) and Jashar-Nasteva (1998: 32) suggested that this may be the result of contact with Albanian and Aromanian (Koneski also considered a possible influence of Greek), although they did not specify the exact dialects in which this phenomenon occurred. Mentions of such usage can be found, at least in the Gorni Polog (Gostivar) dialect⁹ (Popovski 1970: 90) and in the Debar town dialect (Mihajlov 1954: 27), both of which are spoken in the broader Albanian-Macedonian border region in which various strong and weak ties between Macedonian and Albanian speakers exist.

6) In their monograph on the Slavic dialects in Albania, Steinke and Ylli (2007) did not focus explicitly on syntax, but noted an increase in the use of the Albanian-influenced NADJ word order in several dialects: Prespa Macedonian¹⁰ (Steinke and Ylli 2007: 86; Cvetanovski 2010: 128–129), Golloborda Macedonian (Steinke and Ylli 2008: 82–83; Sobolev and Novik 2013: 58), Gora Macedonian (Steinke and Ylli 2010: 81),¹¹ Vraca Štokavian (Steinke and Ylli 2013: 57), and Shijak Štokavian (Steinke and Ylli 2013: 154). The phenomenon can also be observed in Vërnik Macedonian, albeit to a lesser extent (Steinke and Ylli 2007: 265), although Hristova (2003: 131) claimed that NADJ occurred there with “almost the same frequency” as ADJN.¹² With regard to Korça Macedonian, Steinke and Ylli (2007) – who only had access to contemporary spoken data from Boboshtica, not from Drenova – reported that ADJN remained more frequent than it was in other Slavic dialects in Albania (Steinke and Ylli 2007: 316).

1.3 Data

My research is based on the Corpus of Slavic Dialects in Albania (Makartsev and Arkhangelskiy 2024), which includes a sample of the following dialects:

⁸ Of particular relevance is the recent study of the change in the word order in Romani dialects by Adamou et al. (2021). The authors examined Romani speakers (whose inherited word order is the unmarked ADJN) in contact with Romanian (NADJ) and the role of semantic priming from the dominant language in the adoption of its word order. Although methodologically different from the present study, the mentioned article addresses the psycholinguistic processes in the adoption of syntactic models based on language contact by showing how the pattern transfer occurs.

⁹ Koneski further noted that, in the Western dialects, possessive pronouns (which display adjectival morphology in Macedonian), obligatorily follow the noun, although he did not specify the dialects to which this applies.

¹⁰ I provide additional literature in addition to Steinke and Ylli’s monograph where relevant.

¹¹ The same is reported for the Gora dialect that is spoken in Kosovo, where the authors state that possessive pronouns are obligatorily postposed.

¹² Needless to say, without descriptive statistics, any comparison of the provided estimates is impossible.

Table 1. Dialects sampled for analysis

Dialect	Abbreviation	Dialectal zone	Selected references	Sample size, in word forms
Golloborda Macedonian	GM	West Macedonian	(Steinke and Ylli 2008; Sobolev and Novik 2013)	~ 239.700
Korça Macedonian	KM	Southeast Macedonian	(Steinke and Ylli 2007)	~ 34.000
Prespa Macedonian	PM	West Macedonian	(Steinke and Ylli 2007; Cvetanovski 2010)	~ 171.300
Myzeqe Štokavian	MŠ	Novi Pazar-Sjenica (migrational, since the 1920s)	(Makartsev and Kikilo 2022)	~ 58.800
Shijak Štokavian	SŠ	Central Herzegovinian (migrational, since the 1880s)	(Steinke and Ylli 2013)	~ 68.800

The main descriptive statistics and the descriptions of the data types are provided in the referenced corpus. In this article, I focus on the impact of the following variables on word order preferences: age, gender, place of residence (rural/urban), type of dwelling (compact/dispersed), and the method of data collection.

I sampled the corpus for sequences of an adjective and a noun in any order, with a distance of one to three words, accounting for both preposed and postposed adjectives. The resulting sample was edited manually to exclude non-NPs and fixed expressions, and was subsequently saved as a dataset.¹³ This dataset consisted of 3,710 observations of adjectival NPs, each with a corresponding context sentence. The dataset itself, the R scripts used for graphical representations and statistical tests, as well as a supplementary file containing the results of statistical tests that were not included in this article, are described and referenced at the end of my text.

The main variable (encoded as the binary nominal variable *word_order* with two levels, ADJN/NADJ) represents the word order of a noun and an adjective within an NP, analyzed speaker-wise. The two possible values reflect the conservative ADJN and the contact-supported NADJ.¹⁴ Due to variations in the amount of speech produced by different speakers – and consequently in the number of ADJN and NADJ observations per speaker – I recalculated this variable as percentages and excluded all the speakers who produced fewer than five NPs to mitigate the impact of minor contributors, which could have significantly skewed the data when expressed as percentages.

The dependent variable was the *percentage of NAdj* per speaker (numeric continuous, rounded to the first decimal place). After applying all the filters, my dataset for the analysis consisted of 3,640 adjectival NP contexts produced by 107 speakers. All the contexts are included in the dataset (*sentence_text* variable); thus, I have minimized the number of examples presented in the body of the article.

¹³ Fixed expressions, as my recent study (Makartsev 2025) suggested, attest the conservative word order ADJN compared to non-fixed expressions.

¹⁴ Other word orders, such as ADJNADJ, are occasionally attested in my data, but were not taken into consideration. In absolute counts, they were extremely peripheral.

Age provided a relative timeframe for observing the direction of potential language change: All other factors being equal, older speakers would be expected to exhibit less Albanian influence compared to younger speakers. This pattern is linked to the gradual integration of Slavic-speaking communities into Albanian society, the expansion of education and Albanian-language mass media, increased mobility,¹⁵ and rising exogamy.

Age is treated as an independent speaker-level numeric variable. For the purposes of the statistical methods applied in this article, I categorized speakers into age cohorts with a 20-year interval, which I refer to as generations. This variable is encoded as an ordinal variable, *generation*, with five levels: 1920–1939, 1940–1959, 1960–1979, 1980–1999, and 2000–2019.

Gender (in terms of my data, encoded as the binary nominal variable *gender*: f/m) in the researched communities is often associated with social and spatial mobility. Until recently, females were more likely to have lower social mobility and stronger ties to their own community, whereas males tended to be more mobile.

Place of residence (encoded as the binary nominal variable *residence*: rural/urban) is defined as a distinction between rural and urban settings. The population distribution in Albania underwent drastic changes after World War II due to industrialization and urbanization (Bërxfholi et al. 2003: 75–83), which prompted many former rural residents to move from their compact, often monolingual settlements, to multilingual, Albanian-dominated environments. In the Albanian context, urban culture entails more frequent interactions in Albanian in daily life, given that all Albanian cities and towns are linguistically dominated by Albanian.

A related variable is the type of dwelling (encoded as the binary nominal variable *dwelling_type*: compact/dispersed). Compact communities have a distinct center—one or more settlements in which the respective community forms an absolute majority, making the Slavic variety the dominant language of everyday communication. By contrast, dispersed communities do not have such a center.

In my dataset, place of residence was a speaker-level variable, as this distinction could be made within certain communities (GM, PM). By contrast, the type of dwelling was a variety-level variable, as it only applied at a broader level: The SŠ community was compact, while the MŠ community was dispersed (see below). This distinction cannot be applied in individual territorial varieties.

The opposition based on residence could not be established for SŠ and MŠ. SŠ is geographically too close to neighboring (semi)urbanized settlements, such as the towns of Sukth and Shijak, and particularly the city of Durrës, allowing for regular daily connections to its central villages, Borake and Koxhas. The number of MŠ speakers in my sample was too small to construct a comparable opposition. Since the place of residence and the type of dwelling are complementary – reflecting less exposure to Albanian in daily interactions (rural residence or compact dwelling type) versus greater exposure (urban residence or dispersed dwelling type) – they will be analyzed in tandem wherever possible.

Age and gender could not be analyzed for MŠ for the same reason, although they were considered for SŠ. Despite this limitation, the sociolinguistic contexts of SŠ and MŠ remain distinct. SŠ represents a compact community, mainly concentrated in a single settlement in which it forms an absolute majority (Borake; Koxhas is effectively a satellite village within a short walking distance). By contrast, MŠ is a dispersed community: Its speakers are scattered

¹⁵ Even if parts of the population still reside in villages, communication with family members living in cities continues uninterrupted, particularly given the widespread use of real-time communication tools such as messenger and voice and video calls.

across the city of Fier and the villages of Rreth Libofsha, Hamil, Petova, and others, without forming an absolute majority in any of them.

In my sample, education was correlated with age (the older speakers tended to have the lowest education, often only elementary school, sometimes incomplete) and with place of residence (the rural population generally had lower levels of education); therefore, education could not be included as a variable. At the same time, since only PM has compulsory school education in the respective standard language (Macedonian) among all the Slavic dialects, this will be considered as one of the factors in explaining possible differences between PM and the other dialects.

The way of collecting data (binary nominal *researcher_attending*: yes/no) is a metavariable that does not pertain to the sociolinguistic situation of a given community, but to the organization of the speech sample recordings. The corpus contains two main types of spoken data: interviews conducted by researchers (who, by definition, are outsiders in the given communities) and family conversations or simple narratives based on a graphic questionnaire, recorded by trained local assistants.¹⁶

Previous research has shown that speech samples collected by researchers and those recorded by trained local assistants in the researchers' absence sometimes exhibit observable quantitative and qualitative differences due to language management (Makartsev in press [b]). In recordings made without researchers present, speakers are less likely to adapt their language use to norms external to their community (e.g., the standard language).

The way of collecting data is an example-level variable, as some speakers had speech samples recorded both in the presence and in the absence of researchers.¹⁷

2 Analysis

In order to examine the distribution of my data, I grouped the data by dialect, residence, and generation. Gender was excluded at this stage to reduce the number of potential predictors and to prevent further data fragmentation; gender will be addressed separately in §2.4. The boxplots in Figure 1 and the descriptive statistics in Table 2 illustrate the distribution:

¹⁶ The training of the local assistants included technical instruction on how to organize the recordings and did not involve any linguistic training or discussion of possible research topics. None of my trained local assistants had any background in linguistics above the school level.

¹⁷ There were several additional parameters that could not be quantified and were mostly individual in nature, ultimately forming part of the linguistic biographies of specific speakers. Accordingly, they could not be taken into account in this study. Among these factors was the accessibility of television and radio signals from the Yugoslav side of the border during the Socialist period (only relevant for GM and PM). People who were children or young in the Prespa region during the 1970s and 1980s recalled being able to hear music from discos on the then-Yugoslav shore of Lake Prespa, although whether they could understand the lyrics is questionable. Both activities—possessing technologies capable of receiving broadcasts from Yugoslavia and approaching the border area without permission—were extremely risky and could lead to persecution; see Makartsev et al. (2016) for more details. How systematic or widespread access to Yugoslav spoken media was among the population at the time remains unclear.

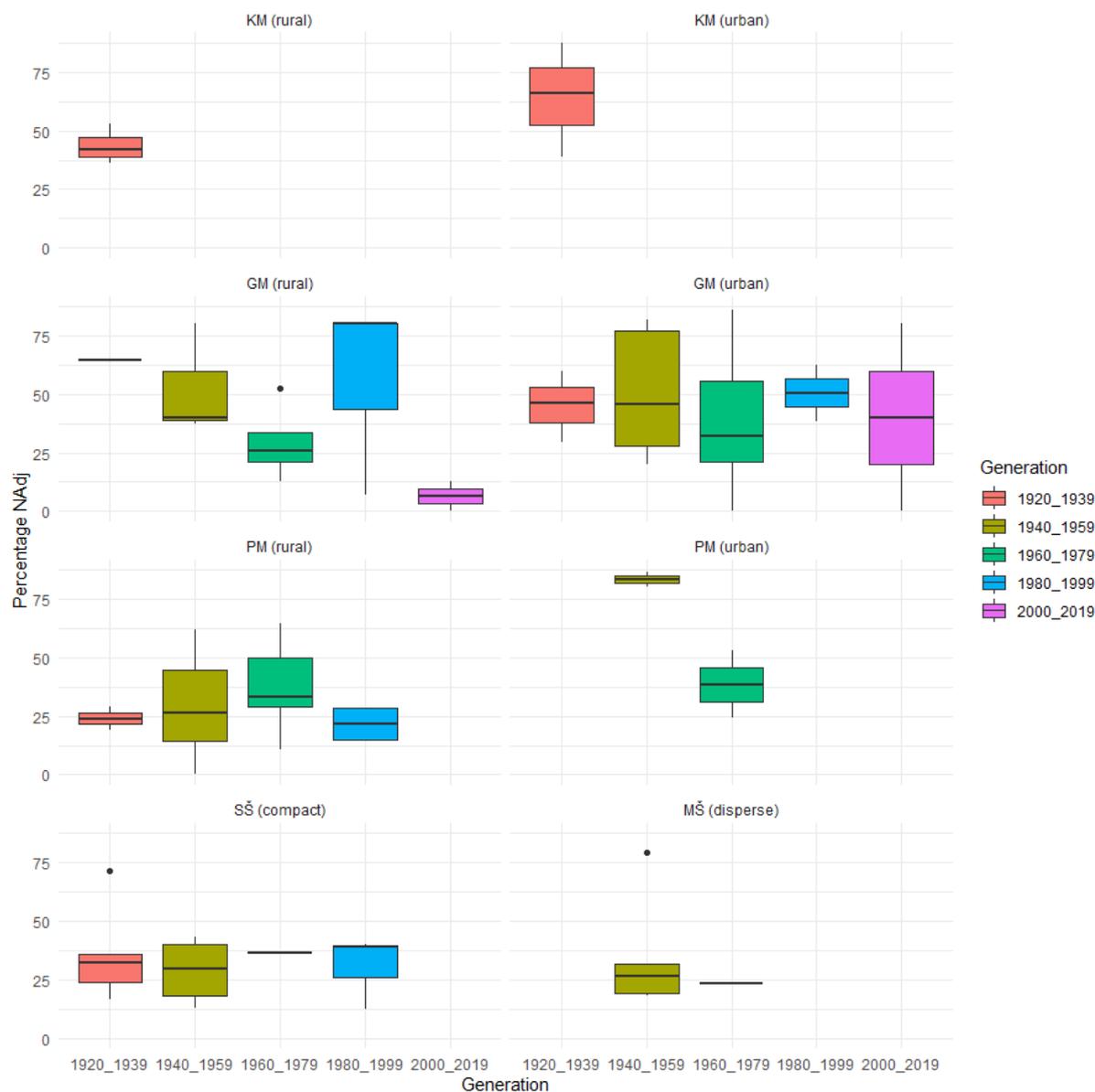


Figure 1. Distribution of NADJ percentage, speaker-wise, by generation, dialect, and residence type

Table 2. Descriptive statistics for NADJ percentage, speaker-wise, by generation, dialect, and residence type

grouping	generation	n	Mean	Median	Standard deviation (SD)	Min	Max
KM (rural)	1920_1939	3	43.67	41.7	8.42	36.4	52.9
KM (urban)	1920_1939	3	64.17	66	24.3	39	87.5
GM (rural)	1920_1939	1	64.5	64.5	NA	64.5	64.5
GM (rural)	1940_1959	3	52.5	40	23.85	37.5	80
GM (rural)	1960_1979	4	29.05	25.55	16.92	12.5	52.6
GM (rural)	1980_1999	3	55.7	80	42.09	7.1	80
GM (rural)	2000_2019	2	6.25	6.25	8.84	0	12.5

GM (urban)	1920_1939	3	45.23	46.2	15.27	29.5	60
GM (urban)	1940_1959	9	49.9	45.5	24.61	20	81.8
GM (urban)	1960_1979	15	38.45	32.3	26.5	0	85.7
GM (urban)	1980_1999	2	50.55	50.55	16.9	38.6	62.5
GM (urban)	2000_2019	2	40	40	56.57	0	80
PM (rural)	1920_1939	2	24	24	6.93	19.1	28.9
PM (rural)	1940_1959	11	29.75	26.4	21.04	0	61.9
PM (rural)	1960_1979	13	36.88	33.3	15.45	10.7	64.7
PM (rural)	1980_1999	4	21.65	21.85	8.03	14.3	28.6
PM (urban)	1940_1959	2	83.35	83.35	4.74	80	86.7
PM (urban)	1960_1979	2	38.55	38.55	20.44	24.1	53
SŠ (compact)	1920_1939	6	35.1	32.1	19.25	16.7	71.4
SŠ (compact)	1940_1959	6	28.7	29.4	13.71	12.7	43.1
SŠ (compact)	1960_1979	1	36.4	36.4	NA	36.4	36.4
SŠ (compact)	1980_1999	3	30.47	38.9	15.57	12.5	40
MŠ (dispersed)	1940_1959	5	34.86	26.7	25.42	17.9	79.2
MŠ (dispersed)	1960_1979	1	23.3	23.3	NA	23.3	23.3
Total:		106					

Unfortunately, the data are sparse and unevenly distributed across the possible groupings, which limited the potential for a detailed analysis and precluded the application of statistical models that account for multiple factors (such as regression models, conditional trees, or mixed-effects models). However, the way in which the data were attested allowed for certain comparisons and observations, although these required some manual adjustments.

1) KM stands apart from the other dialects, as all its speakers belonged to the generation that was born between 1920 and 1939. The division between rural and urban residence for KM speakers is not expected to be relevant for several reasons. The urban KM speakers all belonged to one family that originated from Drenova, while the rural KM speakers all lived in Boboshtica. Moreover, the speakers from Boboshtica were highly mobile throughout their lives, with daily activities extending beyond the village and involving frequent travel across the region and extended stays in cities. Both Boboshtica and Drenova are within walking distance of the city of Korça, which often resulted in shuttle work and other engagements in the city.

In addition, for the oldest generation speaking the other dialects, a meaningful division between urban and rural residence was not feasible due to the small number of speakers (GM: one rural and two urban speakers; PM: two rural speakers; SŠ: five rural speakers; MŠ: no speakers from the oldest generation). Consequently, the oldest generation will be analyzed separately in §2.1.

2) MŠ speakers belonged to two generations: those born between 1940 and 1959 (five speakers) and between 1960 and 1979 (one speaker). Given that, in the case of MŠ, both generations experienced comparable sociolinguistic conditions and completed their language socialization before the sociopolitical changes in Albania, I have merged them into a single

age cohort, 1940–1979. This cohort is then compared cross-dialectally in §2.2 to determine whether significant differences between MŠ speakers and speakers of other dialects exist. Since GM, PM, and SŠ allow for the observation of further variations among generations, they will be analyzed separately in §2.3.

3) Gender will be examined as a possible predictor based on the GM, PM, and SŠ data in §2.4.

4) GM, PM, and SŠ speech samples recorded in the presence versus the absence of researchers will be compared in §2.5.

2.1 KM speakers versus everyone else: The oldest generation

Since all the KM speakers belonged to the generation born between 1920 and 1939, I compared the oldest generation of speakers of all dialects in order to identify any observable trends in this subsection.

As expected, the subsample of the oldest speakers in my corpus was very small. MŠ was excluded since the youngest MŠ speaker was born in 1940, leaving four dialects for analysis: GM (four speakers), KM (six speakers), PM (two speakers), and SŠ (six speakers). While the small number of speakers may reduce the prognostic value of any analysis, it is important to note that the KM speakers included in the corpus represented the entirety of the remaining speakers at the time of data collection. Accordingly, observations based on their data are definitive.

With regard to the method of analysis, I first plotted the distribution of percentages of observations with NADJ word order, speaker-wise, and grouped by dialects. The plot is shown in Figure 1:

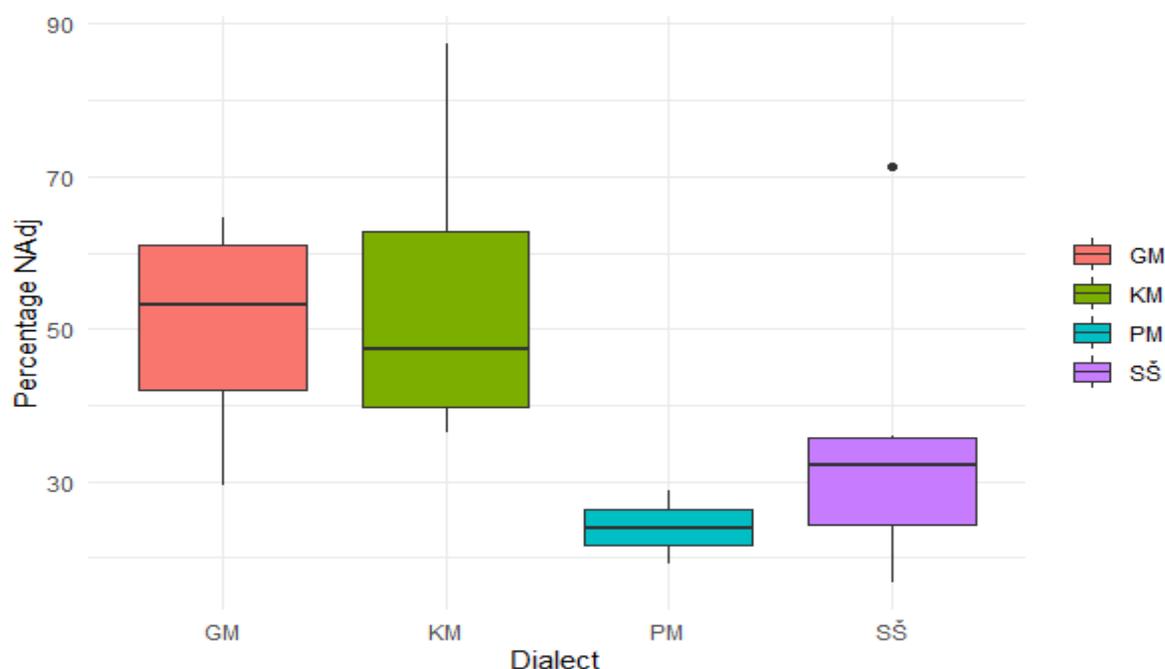


Figure 2. Distribution of NADJ percentage, speaker-wise (generation 1920–1939)

It is interesting that the dialects appeared to group as follows: GM and KM showed a similar distribution with a median of around 50%, indicating an increased use of NADJ to the extent that no dominant word order could clearly be established. PM and SŠ displayed a more conservative pattern, favoring the ADJN word order. Given the small size of the dialectal

groups in the subsample, and based on the visualization, I grouped GM and KM together and PM and SŠ together to compare the two groups and to assess whether the differences between them were statistically significant.

A Mann-Whitney U test (Wilcoxon rank-sum test) was conducted to compare the percentage of NADJ between the groups. The results indicated a statistically significant difference, $W = 69$, $p = .009$, $r = .61$. The effect size was large (Cohen 2013), suggesting that speakers of GM and KM dialects exhibited a substantially different percentage of NADJ compared to the speakers of the other dialects.

The increased use of NADJ word order was expected in KM, as this dialect has the most evident signs of language attrition: The speakers had not communicated with each other in this dialect for decades. Of note, despite the attrition, KM speakers did not differ significantly from the oldest GM speakers, suggesting that early language socialization may be more influential than the drastic shrinkage of the language community in the later stages of the individuals' lives. The conservative character of PM and SŠ was presumably linked to the compact rural type of dwelling, for which speakers of the same dialect formed the dominant population (it should be noted that the PM data are less reliable due to the small number of speakers). At the same time, among all the dialects, only PM has institutional education in the respective standard Slavic language, although its direct impact on the oldest generation may be limited (of the two PM speakers in the subsample, Tum06 received school instruction in Macedonian, while Tum18 had completed her education before Macedonian was introduced).

2.2 MŠ speakers versus everyone else: Effects of *compact* versus *dispersed* dwelling type

The MŠ speakers did not constitute a majority in any of the settlements in which they resided (dispersed dwelling), whereas the SŠ speakers lived in a compact settlement. As mentioned previously, since the type of dwelling (compact versus dispersed) and the type of residence (rural versus urban) both reflect the presumed degree of language contact with Albanian – with compact rural groups presumably experiencing less contact and dispersed urban groups experiencing more – MŠ speakers could be compared to the rest of the dataset.

I only have speech samples from six MŠ speakers, all of whom belonged to the 1940–1959 and 1960–1979 generations. I selected speakers of the same generations from the other dialects and grouped them by residence (rural versus urban). Gender variation could not be considered due to the small size of the MŠ sample. The distribution of NADJ percentage, speaker-wise, is presented in Figure 1 and Table 3:

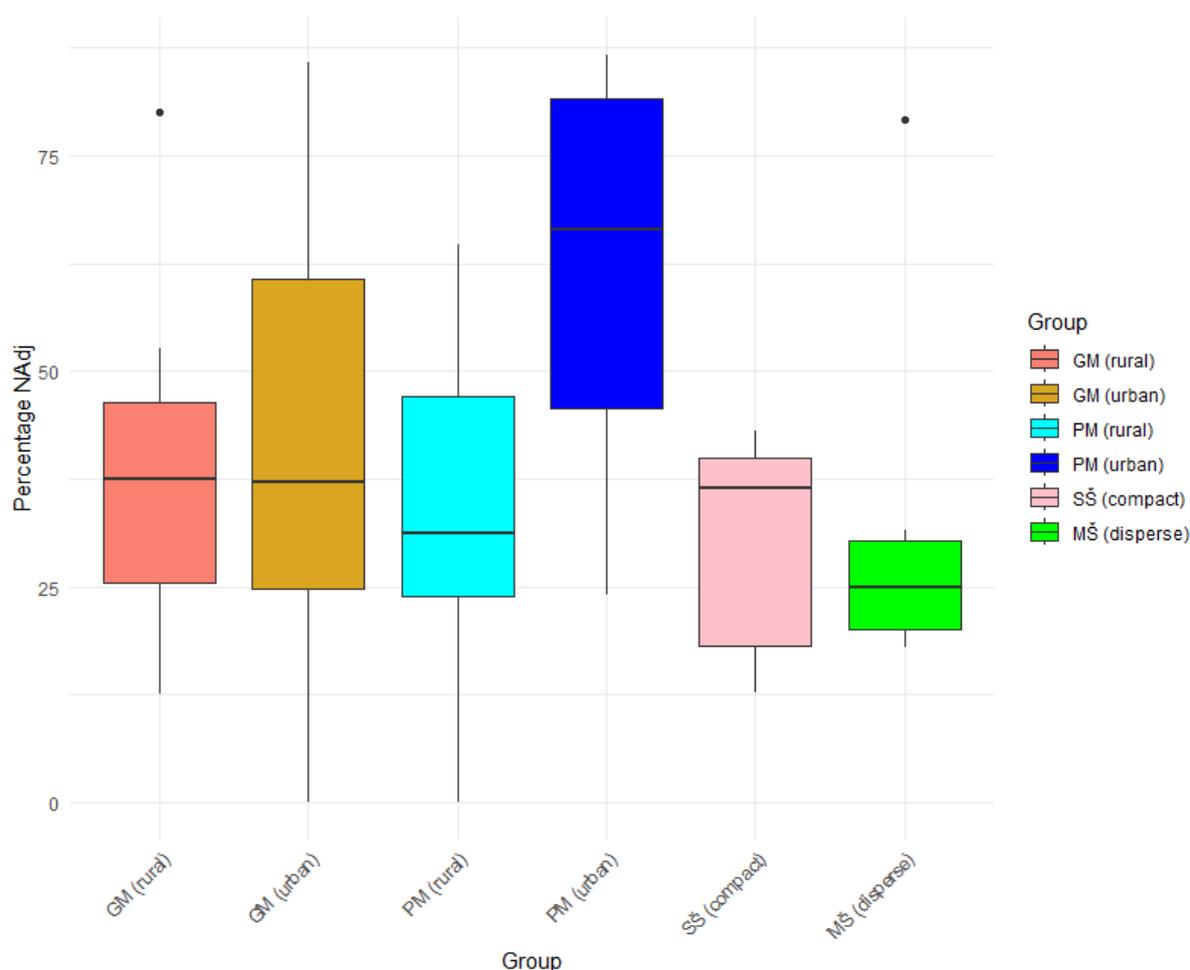


Figure 3. Distribution of NADJ percentages, speaker-wise (generations 1940–1979)

Table 3. Distribution of NADJ percentages, speaker-wise (generations 1940–1979)

grouping	<i>n</i>	Mean	Median	<i>SD</i>	Min	Max	<i>IQR</i>
PM (rural)	24	33.62	31.3	18.17	0	64.7	23.33
GM (urban)	24	42.74	37.1	25.89	0	85.7	35.98
SŠ (compact)	7	29.8	36.4	12.85	12.7	43.1	21.8
GM (rural)	7	39.1	37.5	22.13	12.5	80	20.75
MŠ (dispersed)	6	32.93	25	23.22	17.9	79.2	10.38
PM (urban)	4	60.95	66.5	28.56	24.1	86.7	35.9

Given the small size of some of the groups ($n < 5$ for PM (urban)), a non-normal distribution was assumed, limiting the analysis to non-parametric tests. A series of Brunner-Munzel tests was conducted to compare the MŠ (dispersed) group to each of the other groups in terms of the percentage of NADJ. The results indicated no statistically significant differences after applying Holm's correction for multiple comparisons. The comparison between MŠ (dispersed) and PM (urban) approached significance ($W = 2.31$, $p = .061$), but did not reach the conventional threshold after multiple-testing adjustment ($p = .306$, Holm-

corrected). All the other comparisons yielded non-significant results (see Table 2 in the supplement and the corresponding section of the Rscript). These results show that the distribution of NADJ word order was statistically similar across the different groupings.

These results have different prognostic value across the groups due to their markedly different sizes. Since all my MŠ speakers ($n = 6$) were included, it can be concluded that there were no significant differences in the preferences for the NADJ word order between MŠ speakers and speakers of other dialects from the corresponding generations.

2.3 *Urban versus rural residence as a predictor*

Figure 1 and Table 2 allow for the assessment of possible intergenerational changes by speakers of other dialects and the comparison thereof.

SŠ showed cross-generational stability in the moderate use of the contact-supported NADJ word order, as illustrated in Figure 1.¹⁸

The rural PM speakers did not show significant differences among the generations 1920–1939, 1940–1959, and 1960–1979. The visualization suggests a gradual increase in the percentage of the contact-supported NADJ word order across these generations, consistent with my initial expectations. However, the 1980–1999 generation showed a significant drop compared to the preceding cohort, as seen in Figure 1 and supported by a Mann-Whitney test ($W = 44$, $p = .047$).¹⁹ Due to the limited number of urban PM speakers, it was not possible to analyze potential intergenerational changes.

Similar processes could be observed in the visualization for rural GM speakers, although shifted by one generation: those born between 1980–1999 showed an increase in the use of NADJ compared to the previous generation, followed by a decrease among speakers who were born between 2000 and 2019. Since the number of rural GM speakers per generation never exceeded four individuals and they differed significantly in the absolute number of word order observations, I refrained from conducting statistical tests, as the results would have been inconclusive. The large fluctuations that were observed in the earlier rural GM generations may also be attributed to this factor. No differences were observed among the generations 1960–1979, 1980–1999, and 2000–2019 of urban GM speakers; the generations 1940–1959 (nine speakers) and 1960–1979 (fifteen speakers) provided a solid baseline for this comparison.

Although the absolute numbers of speakers in all the considered groups were insufficient to provide a fully statistically driven account of generational changes, some observations based on the data distribution and the statistical tests, where applicable, can still be interpreted in a broader context:

1) MŠ speakers did not show significant differences from other groups of the same age cohort, suggesting that their dispersed dwelling type had not had a substantial impact on the distribution of word order patterns. Of particular note was the lack of difference between MŠ and SŠ, which—both being Štokavian dialects—are structurally closer to each other than they are to the other dialects in my sample.²⁰ For MŠ, this comparison involved approximately the

¹⁸ Mann-Whitney U tests revealed no significant generational differences (see Table 2 in the supplement for details). Generation 1960–1979 had to be excluded due to only having one observation, which, however, as Figure 1 suggests, aligned with the rest of the SŠ speakers. As generation 1980–1999 only included three speakers in the subsample, the results should be interpreted with caution.

¹⁹ Given that there were only four speakers in the generation 1980–1999, the interpretation of the test results should be approached with caution. See Table 3 in the supplement for the results of all the tests that were conducted for this grouping.

²⁰ It could be speculated that the structure of the adjectival NP in Štokavian, with its compulsory case marking on both the adjective and the noun, is more resistant to Albanian influence than is Balkan Slavic. In Albanian,

second and third generations born in Albania following the migration in the mid-1920s. SŠ speakers migrated about forty years earlier, but have mainly remained a compact group since then, which is likely to have contributed to the preservation of word order patterns. In the case of SŠ, endogamy, as the exclusive marriage strategy (confirmed by field interviews) until the second half of the 20th century, is likely to have played a crucial role in maintaining the conservative character of the word order. Possible changes in word order among later generations of MŠ speakers could not be assessed due to the lack of data.

2) The apparent time lag in the observed decrease in NADJ word order among PM and GM speakers (for rural PM, in the 1980–1999 generation, supported by statistical test results; for rural GM, in the 2000–2019 generation, albeit based on a small sample) may be linked to differing patterns of regional accessibility from within Albania and North Macedonia. Since the early 1990s, following sociopolitical changes in Albania, the PM community has become a well-established point of contact between Albania (particularly the Korça region) and North Macedonia. Trade links developed significantly, particularly after the opening of the Gorica-Stenje border crossing in the Prespa region. By contrast, Golloborda remained difficult to access until recently: The Trebisht-Džepišta border crossing was only opened in 2013; before that, the region was only accessible via low-quality roads from within Albania, with distances of approximately 50–60 km to the nearest connected town. In fact, since the border crossing was still limited to local residents at the time of the study, non-local researchers have to rely on the same poor-quality roads to reach Golloborda from elsewhere.

3) The visualizations in Figure 1 suggest that the PM speakers in the comparative groupings used the NADJ word order somewhat less frequently compared to their GM counterparts: Their medians were lower and the central quartiles were generally more compact. This was expected, as the PM speakers had continuous schooling in the Macedonian language since 1945 (Duma 2007: 69), whereas the GM speakers only had schooling in Macedonian between 1946 and 1948 (Steinke and Ylli 2008: 31). Unfortunately, the distribution of the data did not allow for statistical comparisons between the PM and GM counterparts.

2.4 Gender as a predictor

The distribution of my data did not allow for the simultaneous analysis of multiple predictors. Some generations contained too few data points, leading to overfitting, with the sample sizes varying dramatically across generations, making certain comparisons unreliable. Furthermore, some generations did not exhibit consistent trends in the descriptive statistics and could therefore be collapsed into a single factor.

To determine how best to structure my data in a meaningful way while still exploring possible gender-related effects in the preference for the Albanian-supported NADJ word order, I first included gender as one of the predictors, and created boxplots and a table with descriptive statistics (as shown in Figure 1 and Table 2 in the introductory section to §2). These are not reported here due to their size and the large number of overly atomized groupings that cannot be meaningfully analyzed. Please refer to Figure 1 (suppl.) and Table 4 (suppl.) for these materials.

case is only marked on one of the constituents and the declension system is much less complex in terms of morphological alternations and the total number of distinct case forms. In Balkan Slavic, non-pronominal case marking is minimal. Unfortunately, the data that were available for this article were insufficient to investigate this further.

The KM and MŠ dialects needed to be excluded due to their gender imbalance, which was particularly problematic given the small sample sizes.²¹ In the remaining sample, the following groups could be compared in terms of gender:

1) Since my analysis in §2.3 indicated that GM (rural) did not show abrupt intergenerational changes before 2000, I combined generational cohorts for speakers who were born between 1920 and 1999 into a single subsample to examine possible gender effects.

2) GM (urban) speakers permitted the analysis of potential gender effects within the 1940–1959 and 1960–1979 generations, as there were sufficient speakers in these cohorts.

3) PM (rural) speakers also allowed for a gender comparison within the 1940–1959 and 1960–1979 generations.

4) The previous analysis of SŠ suggested no significant intergenerational effects, which was likely due to the compact community structure. This permitted the comparison of male and female speakers' speech production without accounting for generation.

Due to the small sample sizes in these datasets (ranging from three to 12 speakers per gender subgroup), the Mann-Whitney U test was used for all the gender comparisons. This non-parametric test does not assume normality and is appropriate for comparing the ranked distributions of the percentage of NADJ word order between male and female speakers. Applying this test uniformly across the datasets avoided unreliable normality assumptions while enabling valid comparisons of NADJ distributions. The plots in Figure 4 illustrate the distribution of the data (violin shape); wider sections indicate higher density and are combined with standard box plots. The left and right sides of the central line are symmetrical, with the points representing individual speakers:

²¹ Of the two KM female speakers, one —Dre02— was an outlier with 87.5% NADJ word order, which was also the highest rank value in my entire sample; similarly, of the two MŠ female speakers, one —Pet01— was an outlier with 79.2% NADJ.

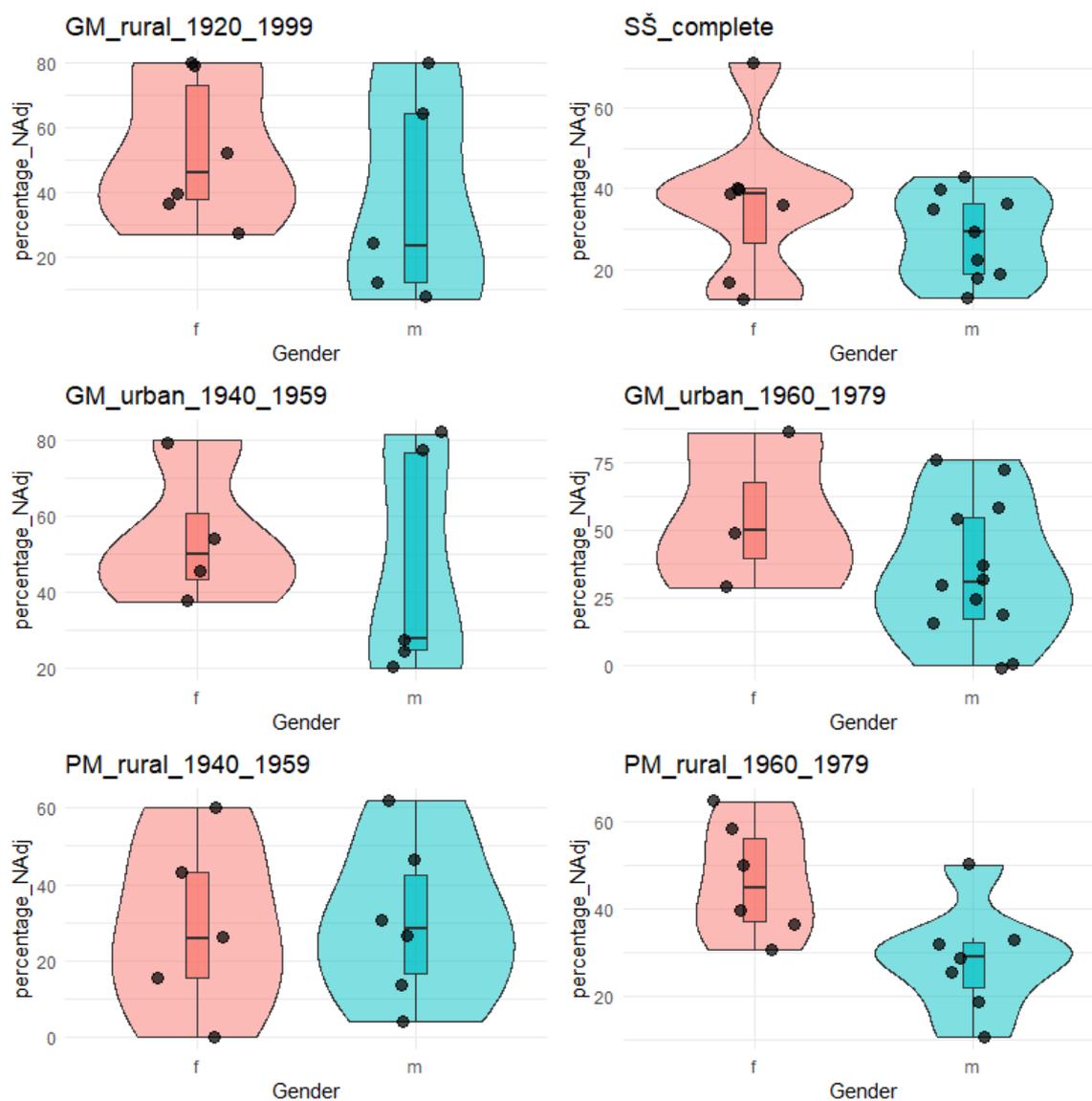


Figure 4. Distribution of NADJ percentage per gender, speaker-wise (selected groupings)

Table 4. Distribution of NADJ percentage per gender, speaker-wise, with Mann-Whitney U test results (selected groupings)

gender	Median	Mean	<i>SD</i>	<i>IQR</i>	Min	Max	Count	Group	<i>W</i>	<i>p</i> -value
F	46.3	52.9	22.5	35	27.3	80	6	GM (rural, 1920_1999)	21	.269
M	23.8	37.6	32.7	52	7.1	80	5	GM (rural, 1920_1999)	21	.269
F	38.9	36.5	19.2	13.6	12.5	71.4	7	SŠ (complete)	38	.49
M	29.4	28.3	10.9	17.6	12.7	43.1	9	SŠ (complete)	38	.49
F	50	54.4	18.4	17.4	37.5	80	4	GM (urban, 1940_1959)	13	.556
M	27.9	46.3	30.3	51.9	20	81.8	5	GM (urban, 1940_1959)	13	.556
F	50	54.8	28.8	28.5	28.6	85.7	3	GM (urban, 1960_1979)	25.5	.278
M	30.4	34.4	25.5	37.3	0	76	12	GM (urban, 1960_1979)	25.5	.278
F	25.9	28.9	23.4	27.8	0	60	5	PM (rural, 1940_1959)	13	.792

M	28.4	30.5	21.1	25.5	4.3	61.9	6	PM (rural, 1940_1959)	13	.792
F	45	46.7	13.2	18.9	30.8	64.7	6	PM (rural, 1960_1979)	36.5	.027 *
M	29.2	28.5	12.4	10.4	10.7	50	7	PM (rural, 1960_1979)	36.5	.027 *

A series of Mann-Whitney U tests was conducted to compare the percentage of NADJ between males and females in the selected speaker groups. The results are included in Table 4. The differences were not statistically significant in most groups, partially due to large interquartile ranges, suggesting that gender did not have a strong influence on word order preferences.

In many groups, although no significant differences were found, the median percentage of NADJ was slightly higher among female speakers than it was among male speakers. A significant difference was only observed in PM (rural, generation 1960–1979), as female speakers ($Mdn = 45.0$) used the NADJ word order more frequently compared to male speakers ($Mdn = 29.2$), $W = 36.5$, $p = .027$. This indicates that gender did, in fact, play a role in the preference for NADJ word order in this group.

It may be speculated that the decline in NADJ usage among male PM speakers born between 1960 and 1979, which made them significantly different from female speakers, was related to the opening of borders after 1990, when most of the speakers in this subsample were in their twenties. Of note, it was predominantly male speakers who engaged in transborder trade and other forms of cross-border mobility with RN Macedonia. Accordingly, the pattern may serve as a precursor to the decline in NADJ observed in the PM generation born between 1980 and 1999, as discussed above.

At the beginning of my research, I hypothesized that male speakers, being more mobile, would be more likely to adopt the contact-supported NADJ word order due to greater exposure to Albanian compared to the female speakers. However, in the case of rural male PM speakers from the 1960–1979 generation, mobility may have had the opposite effect: Their increased interaction with monolingual Macedonian speakers across the border may have reinforced the conservative Slavic ADJN word order at the expense of NADJ.

2.5 Effects of language management

I compared the speech productions that were recorded in the presence and in the absence of researchers to examine the potential effects of language management. To do so, I identified all the speakers who produced five or more target NPs and who were recorded in both conditions. Only four speakers met these criteria, as shown in Figure 5 and Table 5; therefore, they were analyzed individually.

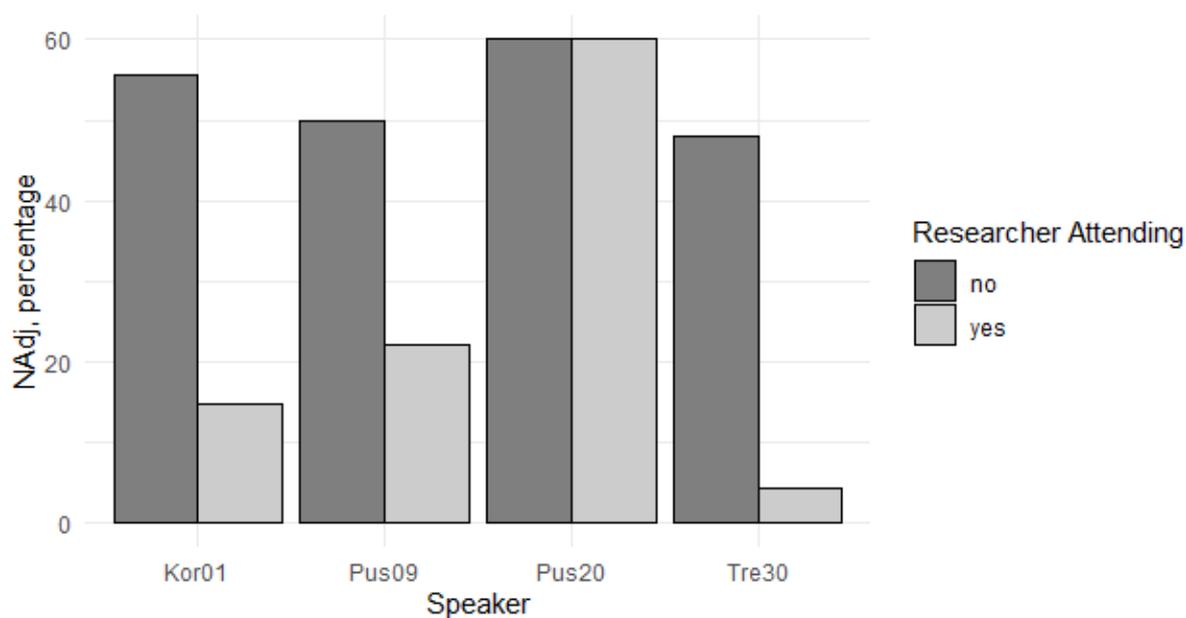


Figure 5. Percentage of NADJ speaker-wise by researcher presence

Table 5. Percentage of NADJ speaker-wise by researcher presence

speaker	NADJ, percentage	total count	researcher attending
Pus20	60	20	yes
Pus20	60	5	no
Pus09	22.2	18	yes
Pus09	50	8	no
Kor01	14.8	61	yes
Kor01	55.6	18	no
Tre30	4.3	46	yes
Tre30	47.9	169	no

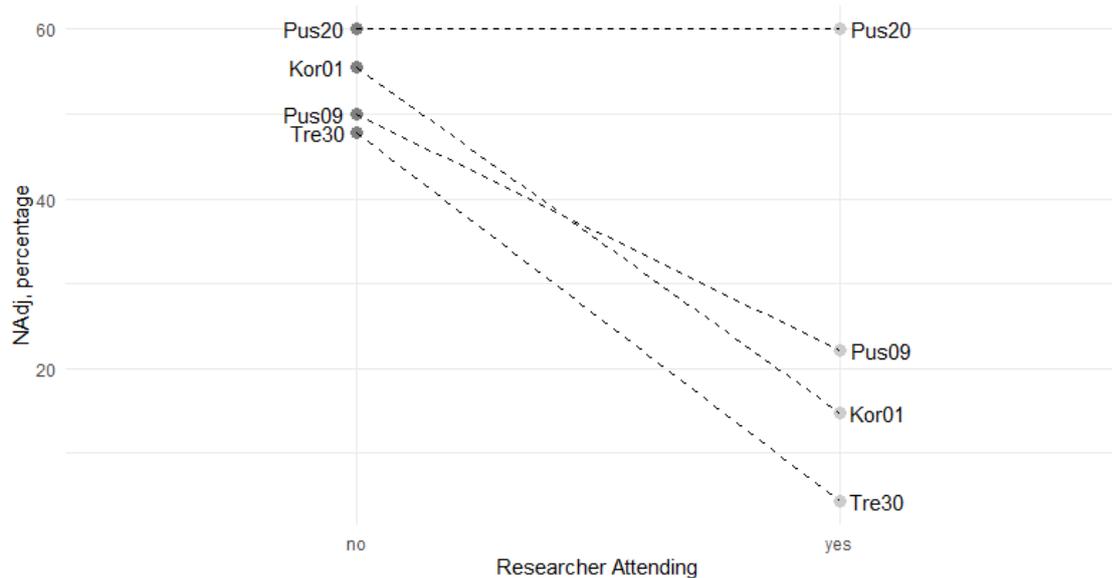


Figure 6. Speaker-wise differences in the percentage of NADJ

The very small number of speakers recorded in both contexts did not allow for a statistical test, but the presented data revealed important differences in the percentage of NADJ in the speech samples. Specifically, in the recordings that were made in the absence of the researcher, three out of four speakers produced significantly more NADJ-ordered NPs: Pus09 did so more than twice as frequently, Kor01 more than three times, and Tre30 more than 10 times.

This finding suggests that language management may, in fact, be a crucial factor in the choice of word order during speech production. This effect is not universal: Even within this very small subsample, Pus20 showed an identical distribution across both recording contexts. However, when the effect did play a role, it could reduce the percentage of NADJ usage significantly—sometimes dramatically, as in the case of Tre30.

3 Conclusion

In this study, I examined the impact of language contact on ADJN word order in the Slavic dialects that are spoken in Albania. The analysis of my corpus data showed that sociolinguistic variation could be (at least partially) linked to preferences in word order. Certain constellations of sociolinguistic parameters particularly favored an increased use of the NADJ word order due to the influence of Albanian, in which this word order is dominant. Conversely, the effects of Albanian may have been neutralized in some groups due to the influence of the standard Macedonian language.

The results of my study suggested that initial language socialization may have a greater influence on word order preferences than the increase in Albanian influence during the later decades of an individual's life, as demonstrated by the comparison of the oldest generations of speakers of different dialects.

A compact type of dwelling, which involves less everyday contact with Albanian, contributed to the stability of word order preferences, as demonstrated by SŠ in comparison to MŠ. Despite the earlier immigration of the former (in the 1880s, roughly two generations earlier than the latter in the 1920s), SŠ speakers still retained a preference for the more conservative ADJN word order, which remained stable across all the generations that could be examined.

Differences in the accessibility of regions in which the dialects are spoken resulted in varying degrees of contact with Macedonian speakers across the border. My data suggested a decline in the use of the contact-supported NADJ word order, which can be considered to be a neutralization of the Albanian influence and the retention of the conservative pre-contact word order, presumably triggered by the contact with standard Macedonian. This decline occurred in rural PM for the generation 1980–1999 and in rural GM for the generation 2000–2019. The lag in the decline can be linked to the time lag in the opening of border crossings in the respective regions.

It can be speculated that a similar decline among male rural PM speakers born between 1960 and 1979 might have been a precursor to the drop that was observed in the subsequent (1980–1999) rural PM generation; the female rural PM speakers born between 1960–1979, who were likely to have had less transborder mobility, may still have reflected the trends that developed before the border was opened.

There may also have been an effect of education in a standard Slavic language—namely, standard Macedonian in PM—as opposed to the other groups for which such education was not systematically provided, although I do not have sufficient data to explore this parameter further.

Finally, language management may have been a crucial factor in word order choice, as suggested by the comparison of speech that was produced in the presence versus in the absence of researchers. Considering that most of my data was recorded in the presence of researchers—and that language management appears to favor structures associated with the common medium, namely standard Macedonian and standard Croatian (both ADJN-dominant)—the observed data may underrepresent the use of NADJ, resulting in a more “conservative” appearance. Therefore, the effects that are discussed in this article may be even more pronounced in unmonitored, unadapted speech.

List of supplementary materials and related data

The following supplementary materials and related data are available for this article:

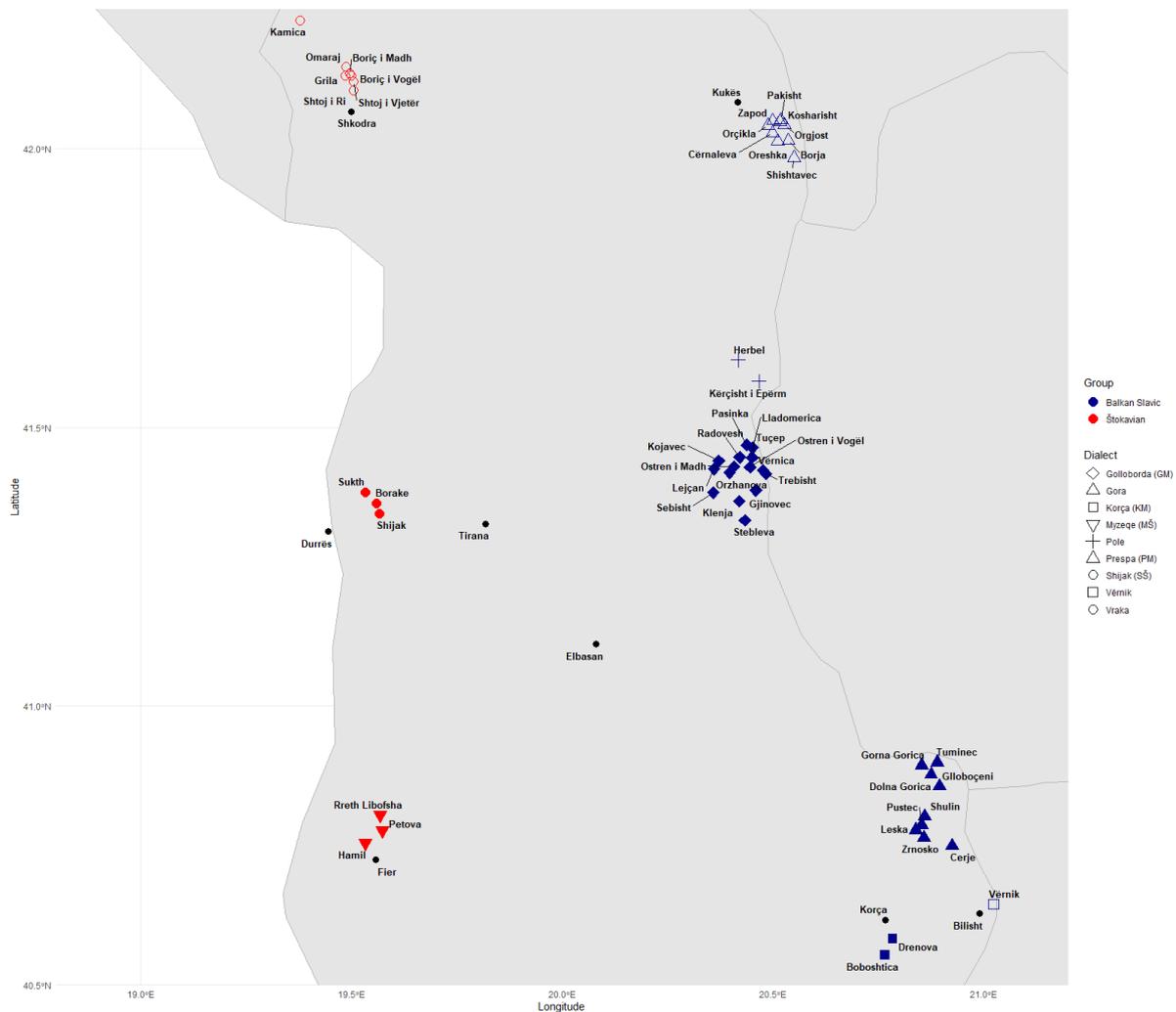
Table 6. Supplementary materials

Number	Reference	Explanation	Source
1	Corpus	Corpus of Slavic dialects in Albania	(Makartsev and Arkhangelskiy 2024)
2	Dataset	Dataset with examples analyzed in the article	DOI: 10.5281/zenodo.15332682
3	Dataset (Locations)	Dataset with names and coordinates of settlements where Slavic dialects are spoken in Albania	DOI: 10.5281/zenodo.15332682
4	Rscript	Rscript to produce tables and plots, and to run tests	DOI: 10.5281/zenodo.15332682
5	Supplement	Supplement with additional statistical tables and plots	DOI: 10.5281/zenodo.15332682

Abbreviations

1–3 — first, second, third person; ACC — accusative; ADJN — word order ADJECTIVE-NOUN; DEF — definite; DIST — distal; F — feminine gender; GM — Golloborda Macedonian dialect; IQR — interquartile range; KM — Korça Macedonian dialect; L — linker; M — masculine gender; *MDN* — median; MŠ — Myzeqe Štokavian dialect; N — neuter; NADJ — word order NOUN-ADJECTIVE; NOM — nominative; NP — nominal phrase; NTPTCP — -n/-t-participle; *p* — probability value; PM — Prespa Macedonian dialect; PRS — present; SG — singular; *r* — effect size; REFL — reflexive; *SD* — standard deviation; SŠ — Shijak Štokavian dialect; *W* — Wilcoxon rank-sum statistic.

Map of Slavic dialects in Albania²²



Bibliography

- Adamou, E., Feltgen, Q. and Padure, C. (2021). A unified approach to the study of language contact: Cross-language priming and change in adjective/noun order. *International Journal of Bilingualism*, 25 (6): 1635–1654.
- Asenova, P. (2018). Balkan syntax: Typological and diachronic aspects. In I. Krapova and B. Joseph (eds.), *Balkan Syntax and (Universal) Principles of Grammar*, 13–36. Berlin: de Gruyter.
- Bara, M., Kahl, T. and Sobolev, A. N. (2005). *Die südarumunische Mundart von Turia (Pindos) = Južnoarumynskij govor sela Tur'ja (Pind): Sintaksis, leksika, etnolingvistika, teksty*. München: Biblion.
- Bërzholi, A., Doka, D. and Asche, H. (2003). *Demographic atlas of Albania: Atlas of Albania [Atlasi gjeografik i popullsisë së Shqipërisë]*. Tiranë: Ilar.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. London: Routledge.
- Cvetanovski, G. (2010). *Govorot na Makedoncite vo Mala Prespa: zapadnoprespanski govor*. Skopje: IMJ “Krstе Misirkov”.

²² Color indicates the major dialect groups: dark blue — Balkan Slavic; red — Štokavian. The dialects included in the Corpus of Slavic Dialects in Albania (GM, KM, PM, MŠ, SŠ) are filled in with the corresponding color, while other territorial Slavic dialects (based on Steinke and Ylli 2007; 2008; 2010; 2013, with slightly updated place names) are left unfilled.

- Domi, M. (ed.) (2002). *Gramatika e gjuhës shqipe: Vëllimi II. Sintaksa*. Tirana: Akademia e shkencave e Shqipërisë, Instituti i gjuhësisë dhe i letërsisë.
- Dryer, M. S. (2013a). 87. Order of adjective and noun. In *The World Atlas of Language Structures Online (v2020.4)* [Online] Available from: <https://wals.info/chapter/87> [Accessed: 23 January 2025].
- Dryer, M. S. (2013b). Determining dominant word order. In *The World Atlas of Language Structures Online (v2020.4)* [Online] Available from: <https://wals.info/chapter/s6> [Accessed: 23 January 2025].
- Duma, G. (2007). *Mala Prespa*. Skopje: Art Print Plus.
- Friedman, V. A. and Joseph, B. D. (in press, 2025). *The Balkan languages*. Cambridge: Cambridge University Press.
- Hristova, E. (2003). *Bălgarska reč ot Albanija. Govorăt na selo Vrăbnik*. Blagoevgrad: Universitetsko izdatelstvo “Neofit Rilski”.
- Jašar-Nasteva, O. (1998). *Kontaktet gjuhësore në hapësirat ballkanike*. Shkup: Logos-A.
- Koneski, B. (1986). *Istorija na makedonskiot jazik*. Skopje: Kultura.
- Konior, D. V. (2021). Minority within a minority: Iabalcea and Caraşova in Romania. In A. N. Sobolev (ed.), *Between separation and symbiosis. South Eastern European languages and cultures in contact*, 157–188. Berlin: de Gruyter.
- Makartsev, M. (2025). Language contacts, New Year, and word order in the nominal phrase in the Corpus of Slavic Dialects in Albania. In I. Sedakova and A. Ippolitova (eds.), *Old and new / young in the linguocultural space of the Balkans (Series Balcanica 18)*, 107–115. Moscow: Institute of Slavic Studies, Russian Academy of Sciences.
- Makartsev, M. (in press, a). The inter-speaker variation in the use of t-demonstratives in a Štokavian dialect in Myzeqe (Albania) and language contact. *Balkanistica*.
- Makartsev, M. (in press, b). The elusive progressive aspect in the minority dialects in Albania: Inherited structures and the grammaticalization paths. *Zeitschrift für Balkanologie*.
- Makartsev, M. and Arkhangelskiy, T. (2024). *Corpus of Slavic Dialects of Albania*. [Online] Available from: <https://slav-dial-alb.uni-oldenburg.de/> [Accessed: 16 January 2025].
- Makartsev, M., Chivarzina, A., Chivarzin, M., Yakovleva, A. and Spyreli, D. (2016). Divided families: The borders’ perception through the human senses. Preliminary results of the field trip in August, 2014. In V. Nitsiakos, I. Manos, G. Agelopoulos, A. Angelidou, and V. Kravva (eds.), *Ethnographic research in border areas: Contributions to the study of international frontiers in Southeast Europe*, 31–37. Konitsa: The Border Crossing Network.
- Makartsev, M. and Kikilo, N. (2022). Some tendencies in the morphosyntax of the migrational Shtokavian dialects in Albania (Shijak and Myzeqe) and Slavic-Albanian language contact. *Slavic World in the Third Millennium*, 17 (1–2): 120–141.
- Marković, M. (2007). *Aromanskiot i makedonskiot govor od ohridsko-struškiot region (vo balkanski kontekst)*. Skopje: MANU.
- Marković, M. (2008). Linearizacija na imenskite sintagmi zavisni od glagolska pridavka. *Studia linguistica Polono-Meridianoslavica*, 12–13: 97–103.
- Mihajlov, P. (1954). *Gradskiot debarski govor*. Skopje: Filozofski fakultet.
- Minova Gjurkova, L. (1994). *Sintaksa na makedonskiot standarden jazik*. Skopje: Rading.
- Mladenov, M. S. (1993). *Bălgarskite govori v Rumănija*. Sofija: BANU.
- Morozova, M. (2016). Govor albancev Ukrainy: Grammatičeskij očerk. In A. Novik, Y. Bučatskaja, D. Ermolin, A. Dugushina, and M. Morozova (eds.), “*Priazovskij otrjad*”. *Jazyk i kul’tura albancev Ukrainy*, part 1 (2), 421–506. Saint Petersburg: MAË RAN.
- Popovski, A. (1970). Makedonskiot govor vo Gostivarskiot kraj. In M. Panov and J. Sinadinovski (eds.), *Gostivarskiot kraj: 2. Makedonskiot, albanskiot i turskiot govor*, 3–200. Gostivar: Opštinsko sobranie Gostivar.
- Sobolev, A. N. and Novik, A. (2013). *Golo Bordo (Gollobordë), Albanija: Iz materialov balkanskoj èkspedicij RAN i SPbGU 2008–2010 gg*. Materialien zum Südosteuropasprachatlas Bd. 6. Sankt-Peterburg: Nauka.
- Steinke, K. and Ylli, X. (2007). *Die slavischen Minderheiten in Albanien (SMA): 1. Teil. Prespa–Vërnik–Boboshtica*. Slavistische Beiträge 458. München: Otto Sagner.

- Steinke, K. and Ylli, X. (2008). *Die slavischen Minderheiten in Albanien (SMA): 2. Teil. Golloborda–Herbel–Kërçishti i Epërm*. Slavistische Beiträge 462. München: Otto Sagner.
- Steinke, K. and Ylli, X. (2010). *Die slavischen Minderheiten in Albanien (SMA): 3. Teil. Gora*. Slavistische Beiträge 474. München: Otto Sagner.
- Steinke, K. and Ylli, X. (2013). *Die slavischen Minderheiten in Albanien (SMA): 4. Teil. Vraka–Borakaj*. Slavistische Beiträge 491. München: Otto Sagner.
- Stojkov, S. (1968). Banatskijat ezikov sąjuz. *Slavjanska filologija*, 10: 97–102.
- Tisheva, J. (2007). Sintaktichni osobenosti v govora na banatskite bălgari v Sărbija. In Asenova, P. (ed.), *Bălgarskite ostrovi na Balkanite*, 155–169. Sofija: Figura.
- Tomić, M. (1984). Govor sviničana. *Srpski dijalektološki zbornik*, 30, 7–263.
- Topolinjska, Z. (2008). Linearizacija na pridavskite modifikatori (AM) vo makedonskata imenska sintagma (NP). *Studia linguistica Polono-Meridianoslavica*, 12–13: 75–95.

Supplement

Maxim Makartsev (Independent researcher)

Contact-induced changes in the order of modifying adjectives and nouns in Slavic dialects in Albania. Supplement

2.2 MŠ vs everyone else: Effects of compact vs disperse dwelling

Table 1 (suppl.) Results of the Brunner-Munzel tests

Comparison	<i>W</i>	<i>p</i> -value	adjusted <i>p</i>
MŠ (disperse) vs. PM (urban)	2.31	0.061	0.306
MŠ (disperse) vs. GM (urban)	1.16	0.276	1.000
MŠ (disperse) vs. GM (rural)	0.98	0.349	1.000
MŠ (disperse) vs. PM (rural)	0.49	0.633	1.000
MŠ (disperse) vs. SŠ (compact)	0	1.000	1.000

2.3 GM, PM, MŠ: Urban vs rural residence as a predictor

Table 2 (suppl.) Mann-Whitney U Test results for percentage of NAdj across generations in SŠ (compact)

Comparison	<i>W</i>	<i>p</i>
1920–1939 vs. 1940–1959	19	.936
1940–1959 vs. 1980–1999	11	.694
1920–1939 vs. 1980–1999	8	.905

Table 3 (suppl.) Mann-Whitney U Test results for percentage of NAdj across generations in PM (rural)

Comparison	<i>W</i>	<i>p</i>
1920–1939 vs. 1940–1959	10	.923
1940–1959 vs. 1960–1979	54	.325
1960–1979 vs. 1980–1999	44	.047 *
1920–1939 vs. 1980–1999	6	.481

2.4 GM, PM, MŠ: Gender as a predictor

In order to establish how I can best structure my data in a sensible way to still pursue some possible gender-related effects in the preference of the Albanian-supported NAdj word order, I added gender as one of the predictors and established boxplots and a table with descriptive statistics.

Table 4 (suppl.) Gender distribution per dialect, generation, and residence type

grouping	gen	gender	n	Mean	Median	SD	Min	Max
KM (rural)	1920	1939.f	1	52.9	52.9	NA	52.9	52.9
KM (rural)	1920	1939.m	2	39.05	39.05	3.75	36.4	41.7
KM (urban)	1920	1939.f	1	87.5	87.5	NA	87.5	87.5
KM (urban)	1920	1939.m	2	52.5	52.5	19.09	39	66
GM (rural)	1920	1939.m	1	64.5	64.5	NA	64.5	64.5
GM (rural)	1940	1959.f	3	52.5	40	23.85	37.5	80
GM (rural)	1960	1979.f	2	39.95	39.95	17.89	27.3	52.6
GM (rural)	1960	1979.m	2	18.15	18.15	7.99	12.5	23.8
GM (rural)	1980	1999.f	1	80	80	NA	80	80
GM (rural)	1980	1999.m	2	43.55	43.55	51.55	7.1	80
GM (rural)	2000	2019.f	1	0	0	NA	0	0
GM (rural)	2000	2019.m	1	12.5	12.5	NA	12.5	12.5
GM (urban)	1920	1939.f	1	60	60	NA	60	60
GM (urban)	1920	1939.m	2	37.85	37.85	11.81	29.5	46.2
GM (urban)	1940	1959.f	4	54.38	50	18.44	37.5	80
GM (urban)	1940	1959.m	5	46.32	27.9	30.33	20	81.8
GM (urban)	1960	1979.f	3	54.77	50	28.85	28.6	85.7
GM (urban)	1960	1979.m	12	34.37	30.45	25.53	0	76
GM (urban)	1980	1999.f	1	38.6	38.6	NA	38.6	38.6
GM (urban)	1980	1999.m	1	62.5	62.5	NA	62.5	62.5
GM (urban)	2000	2019.f	2	40	40	56.57	0	80
PM (rural)	1920	1939.f	1	28.9	28.9	NA	28.9	28.9
PM (rural)	1920	1939.m	1	19.1	19.1	NA	19.1	19.1
PM (rural)	1940	1959.f	5	28.9	25.9	23.44	0	60
PM (rural)	1940	1959.m	6	30.47	28.4	21.08	4.3	61.9
PM (rural)	1960	1979.f	6	46.7	45	13.22	30.8	64.7
PM (rural)	1960	1979.m	7	28.47	29.2	12.36	10.7	50
PM (rural)	1980	1999.f	1	28.6	28.6	NA	28.6	28.6
PM (rural)	1980	1999.m	3	19.33	15.1	8.04	14.3	28.6
PM (urban)	1940	1959.f	2	83.35	83.35	4.74	80	86.7
PM (urban)	1960	1979.f	1	53	53	NA	53	53
PM (urban)	1960	1979.m	1	24.1	24.1	NA	24.1	24.1
SŠ (compact)	1920	1939.f	3	41.37	36	27.74	16.7	71.4
SŠ (compact)	1920	1939.m	3	28.83	29.4	6.27	22.3	34.8
SŠ (compact)	1940	1959.f	1	40	40	NA	40	40
SŠ (compact)	1940	1959.m	5	26.44	18.8	14.02	12.7	43.1
SŠ (compact)	1960	1979.m	1	36.4	36.4	NA	36.4	36.4
SŠ (compact)	1980	1999.f	3	30.47	38.9	15.57	12.5	40
MŠ (disperse)	1940	1959.f	2	52.95	52.95	37.12	26.7	79.2
MŠ (disperse)	1940	1959.m	3	22.8	18.9	7.64	17.9	31.6
MŠ (disperse)	1960	1979.m	1	23.3	23.3	NA	23.3	23.3
Total:			106					

LABILE VERBS IN MODERN ALBANIAN: A PRELIMINARY CORPUS-BASED STUDY

Maria Morozova

Institute for Linguistic Studies
Russian Academy of Sciences, Saint Petersburg
morozovamaria86@gmail.com

Alexander Rusakov

Institute of Language and Literature
Academy of Sciences of Albania, Tirana
ayurusakov@gmail.com

The goal of the study is to analyze the use of labile verbs in the modern written Albanian language based on the Albanian National Corpus, which contains over 31 million words. In this paper, we present our findings from a pilot study of eight Albanian P-labile verbs belonging to different semantic groups that are typologically prone to lability. The study has shown that the selected verbs behave differently. Only phasal verbs *filloj* ‘begin’ and *mbaroj* ‘finish’ demonstrate consistent lability. The motion verb *lëviz* ‘move’ is used as a labile verb, with a predominance of active intransitive usages, while the verb *ziej* ‘boil; seethe; ferment’ shows variation between non-active and active intransitive usages. Deadjectival color verbs *zbardh* ‘whiten; brighten’ and *skuq* ‘make/color red; fry’ appear both transitively and intransitively, but only in their basic color-related meanings. The verb of distinction and change *dalloj* ‘distinguish; differentiate’ primarily employs active intransitive forms in reciprocal meanings and morphologically non-active forms in other meanings, while the related verb *ndryshoj* ‘change; vary, distinguish’ behaves as a typical labile verb, with a secondary development of non-active usages.

Keywords: P-lability, non-active voice, lexical semantics, Albanian National Corpus.

ЛАБИЛНИТЕ ГЛАГОЛИ ВО СОВРЕМЕНИОТ АЛБАНСКИ ЈАЗИК: КОРПУСНО ИСТРАЖУВАЊЕ

Марија Морозова

Институт за лингвистички истражувања
Руска Академија на науките, Санкт Петербург
morozovamaria86@gmail.com

Александер Русаков

Институт за јазик и литература
Академија на науките на Албанија, Тирана
ayurusakov@gmail.com

Целта на оваа статија е анализа на употребата на лабилните глаголи во современиот пишан албански јазик, врз основа на Албанскиот национален корпус. Во трудот ги претставуваме нашите согледувања од прелиминарното истражување на осум албански пациентивни лабилни глаголи, коишто припаѓаат на различни семантички групи. Резултатите покажуваат дека овие глаголи се однесуваат различно. Само фазните глаголи *filloj* ‘почне’ и *mbaroj* ‘заврши’ доследно покажуваат лабилност. Глаголот за движење *lëviz* ‘се движи’ се користи како лабилен глагол, но преовладуваат активните непреодни употреби, додека глаголот *ziej* ‘врие; зоврива; ферментира’ покажува варијација помеѓу неактивните и активните непреодни употреби. Деадјективните глаголи *zbardh* ‘обели; разјаснува’ и *skuq* ‘поцрвенува; пржи’ се јавуваат како преодни и како непреодни, но само во нивните основни значења, поврзани со значењето на боја. Глаголот *dalloj* ‘разликува, истакнува’ првенствено се јавува во активни непреодни форми со реципрочно значење и во морфолошки неактивни форми, во други значења, додека сличниот глагол *ndryshoj* ‘менува; варира, разликува, истакнува’ се однесува како типичен лабилен глагол со секундарни неактивни употреби.

Клучни зборови: пациентивна лабилност, неактивен залог, лексичка семантика, Албански национален корпус.

1 Introduction

In Albanian, as in most of the Balkan languages, with the exception of Romani, anticausative derivation predominates over causative (Nichols 2020). This aligns Albanian with other SAE languages (Haspelmath 2001). It has also been noted that some verbs in Balkan languages are characterized by the phenomenon of lability (ambitransitivity), i.e., the same verb can be used transitively and intransitively without changing its form (Friedman and Joseph 2025). Friedman (2010) suggests that lability represents a Balkanism of scalar type, manifested to varying degrees in different Balkan languages. Recent studies have found more labile verbs in Macedonian (especially in Western and some Southeastern dialects) and Modern Greek, and fewer in Aromanian, Bulgarian, and Albanian (Bužarovska and Mitkovska 2022; Makartsev et al. 2024).

Though Albanian has been mentioned among the Balkan languages with the least representation of lability, the number of labile verbs in this language is quite substantial. It is worth noting that Albanian can use ambitransitives for four out of the 18 pairs of contrasting transitive-intransitive verbs in the diagnostic list by Nichols et al. (2004): *die* — *kill* (1), *learn* — *teach*, *boil* — *boil*, *sleep* — *be asleep*. Dhrimo (1965) mentions around 30 Albanian verbs that can be employed both transitively and intransitively. The only existing study of lability in modern Albanian provides a list of about 50 verbs showing lability (Diveeva 2013).

- (1) a. *ata* *vdis-nin*
 they.M.NOM die-IPF.3PL
 ‘they were dying’
- b. *e=rrah-u* *sa* *e=vdiq*
 3SG.ACC=beat-AOR.3SG as.much 3SG.ACC=kill.AOR.3SG
 ‘he beat him so that he killed him’ (examples constructed by authors)

There are no corpus-based studies examining Albanian lability, and our research aims to fill this gap. The data come from the Albanian National Corpus (ANC), which comprises over 31 million words. The goal of the study is to analyze the use of different forms of labile verbs in the modern written Albanian employing both quantitative and qualitative methods.

In the present article, we share observations from a pilot study of eight Albanian labile verbs belonging to different semantic groups that are typologically prone to lability. Section 2 presents an overview of the data extracted from ANC and describes the methodology of data collection and analysis. In the subsections of Section 3, we focus on the particular verbs and analyze ratios of the active transitive, active intransitive, and non-active intransitive usages of each verb, their lexical meanings and argument structures, and the distribution of various TAM forms across the three usages (if significant differences are observed). When relevant, attention is given to the characteristics of texts (e.g., text type and period of creation) in which a certain type of usage appears. In Conclusion, we summarize the preliminary results and outline the perspectives of the study.

2 Methodology and overview of the data

In our study, we primarily address P-lability, as exemplified in (1). The most significant groups demonstrating lability of this type in Albanian, as in many other languages, are phasal verbs and motion verbs. A notable group of verbs exhibiting conversive lability (not considered in this article), such as *pëlqej* ‘like’, is also present in Albanian.

For our pilot study, we selected eight Albanian verbs that represent the main semantic groups that tend to demonstrate lability, based on the typological criteria proposed by Letuchiy (2013), which have previously been applied for Albanian in Diveeva (2013). These are: phasal verbs *filloj* ‘begin’ and *mbaroj* ‘finish’, the motion verb *lëviz* ‘move’, deadjectival color verbs *zbardh* ‘whiten; brighten’ and *skuq* ‘make/color red; fry’, verbs of distinction and change *dalloj* ‘distinguish; differentiate’ and *ndryshoj* ‘change; vary; distinguish’, and the verb *ziej* ‘boil; seethe; ferment’ denoting (in its intransitive use) a “spontaneously arising situation” (Letuchiy 2013: 111). These verbs can be classified as demonstrating the anticausative type of lability, with some also showing a degree of reciprocal lability.

In our analysis, we examined both active intransitive and active transitive usages of the selected verbs, as well as the presence of parallel morphologically non-active forms. The examples were searched in ANC by lemma. In the further analysis of the extracted samples, both the primary and secondary (metaphorical) usages were taken into account. We counted the number and share of active intransitive, active transitive, and non-active usages in the first 300 randomly ordered examples, as shown in Table 1 and Figure 1.

Table 1. Number and share of active intransitive, active transitive, and non-active usages of labile verbs (first 300 examples, numbers rounded to two decimal places)

	number of examples	intransitive		transitive		non-active	
		number	share	number	share	number	share
<i>dalloj</i>	300	48	0.16	124	0.41	128	0.43
<i>filloj</i>	300	178	0.59	118	0.39	4	0.01
<i>lëviz</i>	300	241	0.80	53	0.18	6	0.02
<i>mbaroj</i>	300	160	0.53	134	0.45	6	0.02
<i>ndryshoj</i>	300	138	0.46	137	0.46	25	0.08
<i>skuq</i>	300	10	0.03	52	0.17	238	0.79
<i>zbardh</i>	300	24	0.08	136	0.45	140	0.47
<i>ziej</i>	300	191	0.64	86	0.29	23	0.08
Total	2400	989	0.41	840	0.35	571	0.24

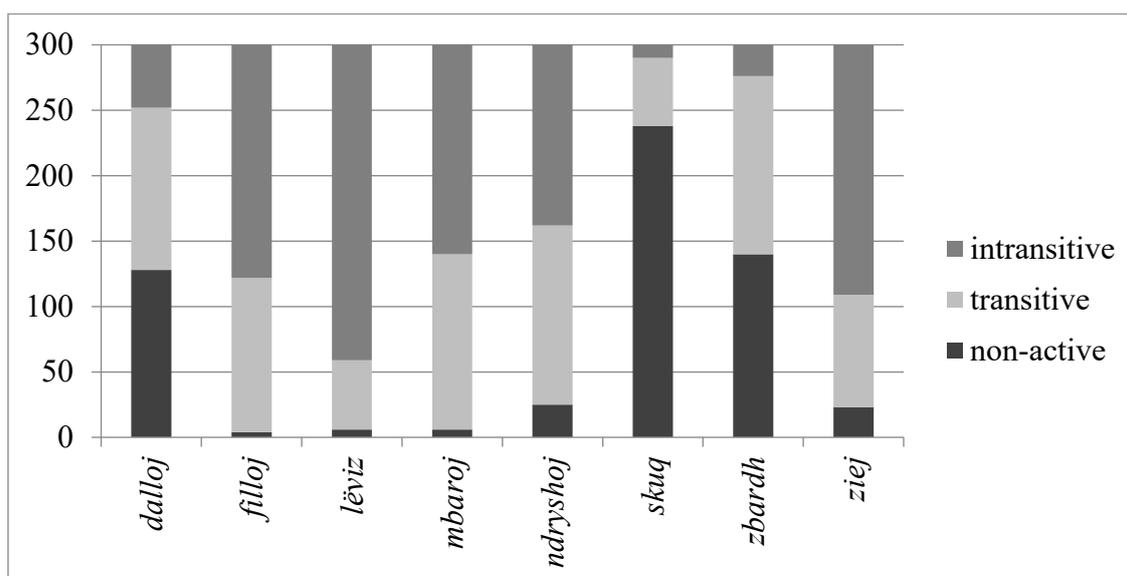


Figure 1. Number of active intransitive, active transitive, and non-active usages of labile verbs (first 300 examples)

Non-active usages are most typical for the verbs *skuq*, *dalloj*, and *zbardh*. For the verb *skuq*, they significantly outnumber active intransitive and transitive usages, accounting for 79% of all analyzed examples with this verb. As for *dalloj* and *zbardh*, active transitive usages are almost as frequent as their non-active counterparts, while active intransitive usages are significantly less frequent. The verb *lëviz* occurs mostly intransitively (80% of the sample). Phasal verbs *filloj* and *mbaroj*, as well as the verbs *ndryshoj* and *ziej*, are used mostly transitively and intransitively, with a more (*filloj*, *ziej*) or less (*mbaroj*, *ndryshoj*) pronounced prevalence of intransitive usages. Non-active forms are extremely rare for these verbs.

The distribution of active intransitive, active transitive, and non-active usages is defined by two fundamentally different factors: (a) the ratio of intransitive and transitive usages, which is determined almost exclusively by the semantics of the verb, and (b) the ratio of active and non-active intransitive usages, which can vary depending on the specific verb, its different meanings, or its semantic verb group. Below, we'll focus on the behavior of the eight selected verbs in relation to these factors. Verbs belonging to one semantic group will be also compared to each other, such as *filloj* 'begin' and *mbaroj* 'finish' in 3.1.

3 Labile verbs in the modern Albanian: Some observations on the usage and meaning

3.1 Phasal verbs *filloj* 'begin' and *mbaroj* 'finish'

It has been noted that phasal verbs "are labile even in some languages for which lability is generally uncharacteristic" (Letuchiy 2013: 171). In this article, we examine the two most common and productive phasal verbs in contemporary Albanian: *filloj* 'begin' and *mbaroj* 'finish'. Etymologically, they are derived from a noun (*fill* 'thread; source' > *filloj*, cf. *zë fill* 'begin (lit. 'get a source')') and an adjective (*i mbarë* 'proper, going well' > *mbaroj*). The two verbs notably differ in frequency: *filloj* appears 20,289 times in the ANC, while *mbaroj* appears 3,825 times (these figures include deverbal adjectives *i filluar* 'initiated' and *i mbaruar* 'finished' due to technical reasons). This aligns with the general typological tendency towards a higher frequency of inchoative verbs among phasal verbs (Letuchiy 2013: 174).

In our sample (see Table 1), active forms of both verbs are overwhelmingly dominant. For *filloj*, we observed 178 instances of active intransitive usage (2), 118 of transitive usage, and four of non-active usage (4). For *mbaroj*, the corresponding numbers are 160, 134 (3), and six. Neither verb shows a statistically significant difference in the distribution of active intransitive, active transitive, and non-active forms ($p < .05$).

- (2) *Filluan protestat nga populli dhe nga ne, nxënësit, por askush nuk na e vuri veshin.*
 'Protests began from the people and from us, the students, but no one listened to us.'
 (*Vite të vegjëlisë*, Enver Hoxha, 1976[1962])

<i>fillua-n</i>	<i>protesta-t</i>	<i>nga popull-i</i>
begin-AOR.3SG	protest(M).PL-DIR.PL.DEF	from people(M)-NOM.SG.DEF
'protests began from the people' ¹		

- (3) *Tani, tani mezi pris-te të mbaro-nte punë-n.*
 now now hardly wait-IPF.3SG SBJV finish-IPF.3SG work(F)-ACC.SG.DEF
 'Now, now he couldn't wait to finish the job.'
 (*Unaza prej floriri*, Kostaq Duka, 2009)

¹ The glossing of a long example is restricted to the verb and its immediate syntactic environment.

- (4) *Pra, të menduarit Shekspirian është një çështje që duhet të studiohet dhe të sprovohet mirë para se të fillohet performimi eventual i ndonjë vepre të këtij autorit.*
 ‘So, Shakespearean thinking is an issue that must be studied and thoroughly tested before the beginning of the eventual performance of any work by this author.’
 (Zëri, 16.03.2013)

para se të fillo-het performim-i
 before COMP SBJV begin-NACT.PRS.3SG performance(M)-NOM.SG.DEF
 ‘before the performance begins’

Some interest for us lies in the peripheral non-active forms of phasal verbs, although it is rather difficult to draw definite conclusions about them based on our dataset. It can be noted that all four instances of the non-active usage of *filloj* appear in newspapers (three of which are from Kosovo), while the six non-active examples with *mbaroj* are found both in newspapers and in fiction and religious texts. Examples with passive semantic interpretation, which are possible according to Diveeva (2013), did not occur in our sample.

There are some formally intransitive, but semantically transitive usages of the verb *filloj*, demonstrating different argument structures: with the combination of third person singular dative and accusative pronominal clitics *ia* (5) and with the preposition *me* ‘with’. Such instances seem to be extremely rare: only three cases are found in the sample of 300 examples.

- (5) *Pra, nëse ditët janë të ftohta, milingonat flenë, e kur është nxehtë, ia fillojnë punës.*
 ‘So, if the days are cold, the ants sleep, and when it’s hot, they get to work.’
 (Zëri, 01.07.2013)

ia=fillo-jnë punë-s
 DAT.3SG:ACC.3SG=begin-PRS.3PL work(F)-OBL.SG.DEF
 ‘they get to work’

An additional count was conducted for the third-person singular aorist forms of both verbs (*u fillua*, *u mbarua*) occurring in all texts of the ANC. Their distribution in the different text types mirrors the one observed in our 300 examples sample. The 17 instances of *mbaroj* (out of 3,825 usages of this verb) are found both in fiction (13, including literature from the early 20th-century) and in newspapers, while the 25 instances of *filloj* (out of 20,289) mainly come from newspapers. There are three examples in which *filloj* has a clearly passive interpretation, as in (6).

- (6) *Luft-a civil-e nuk u fillua*
 war(F)-NOM.SG.DEF civil-F NEG NACT begin.NACT.AOR.3SG
nga ballistë-t, por nga komunistë-t.
 from Ballist(M)-DIR.PL.DEF but from communist(M)-DIR.PL.DEF
 ‘The civil war was not started by the Ballists, but by the communists.’
 (Panorama, 29.07.2016)

It can be stated that *filloj* ‘begin’ and *mbaroj* ‘finish’ are typical labile verbs. Their non-active intransitive usages are negligible in number. For the verb *filloj*, such usages appear to be a recent innovation, possibly emerging in official bureaucratic registers of Albanian over the past few decades. In contrast, non-active examples of *mbaroj* seem to be more deeply rooted in the Albanian language and require a dedicated diachronically oriented study.

3.2 The verb of motion *lëviz* ‘move’

In typological terms, motion verbs are highly prone to lability, as are phasal ones (Letuchiy 2013: 183–197). The verb *lëviz* is used mainly intransitively in our sample (241 of 300 examples) and its usages are mostly interpreted as anticausative (7). The general meaning of this verb is a change of position which may entail either changing location or moving into a new place of residence, employment, etc. (conceptualized metaphorically as a movement, see [8]). Non-active forms are represented in six of 300 examples, one of which has a passive interpretation (9).

- (7) *Gjymtyrë-t* *lëviz-in* *pa* *kontroll* <...>.
 limb(F).PL-DIR.PL.DEF move-PRS.3PL without control(M).DIR.SG
 ‘Limbs move uncontrollably.’
 (Panorama, 17.11.2016)
- (8) *Garanci* *për* *persona-t* *që* *ka-në* *lëvizur*.
 guarantee(F).DIR.SG for person(M)-DIR.PL.DEF COMP have-PRS.3PL move.PTCP
 ‘Guarantee for people who have moved.’ [about immigration]
 (Agjencia Telegrafike Shqiptare, 17.01.2017)
- (9) *Shumë* *prej tyre* *ja-në* *lëvizur* *dhe* *i=ka-në*
 many from they.ABL be-PRS.3PL move.PTCP and 3PL.ACC=have-PRS.3PL
çuar *më në veri,* *madje* *edhe më larg*
 carry.PTCP CMPR in north(M).DIR.SG even and CMPR far
Tiranë-s.
 Tirana(F)-OBL.SG.DEF
 ‘Many of them were moved and they have taken them further north, even further away from Tirana.’
 (Panorama, 13.04.2015)

No significant difference was found in the distribution of active intransitive and transitive usages across the different TAM forms of this verb. The distribution of a few non-active examples across the different text types (four are found in the press and two in fiction and religious texts) does not allow for any definite conclusions. Overall, *lëviz* represents a typical labile motion verb, with a predominance of intransitive usages (presumably original) due to its semantics and the relatively low agentivity of the subject in these usages.

3.3 The verb *ziej* ‘boil; seethe; ferment’

As mentioned in Section 2, the verb *ziej* ‘boil; seethe; ferment’ is used mostly intransitively. Transitive usages are less frequent than intransitive ones, while non-active forms are rare. The distribution of the three usages in different tense forms found in our 300 examples sample is shown below in Table 2.

Table 2. Active intransitive, active transitive and non-active usages of *ziej* (first 300 examples, numbers rounded to two decimal places)

		all	active intransitive	active transitive	non-active
Present	number	83	54	20	9
	share	0.28	0.28	0.23	0.39
Imperfect	number	114	93	17	4
	share	0.38	0.49	0.20	0.17
Aorist	number	12	5	7	0
	share	0.04	0.03	0.08	0
Perfect	number	9	4	2	3
	share	0.03	0.02	0.02	0.13
Plusquamperfect	number	5	0	4	1
	share	0.02	0	0.05	0.04
Future	number	2	1	1	0
	share	0.01	0.01	0.01	0
Present Subjunctive	number	35	26	5	4
	share	0.12	0.14	0.06	0.17
Imperfect Subjunctive	number	8	4	4	0
	share	0.03	0.02	0.05	0
Perfect Subjunctive	number	2	1	0	1
	share	0.01	0.01	0	0.04
Nonfinite forms	number	11	3	7	1
	share	0.04	0.02	0.08	0.04
Imperative	number	19	0	19	0
	share	0.06	0	0.22	0
All forms		300	191	86	23

A substantial number of examples with transitive imperatives is found in recipes published in newspapers, illustrated in (10). The prevalence of active imperfective forms (present and imperfect tense) over perfective ones (aorist and perfect) can be explained by the fact that *ziej* usually describes atelic processes of boiling (11), fermenting, and distilling. From the perspective of actional meaning,² the less frequent active aorist and perfect forms can be interpreted as entry into the process (11) or state (12), while transitive usages describe a completed action directed to an object (13).

- (10) *Pris-ni panxhar-in në kubikë dhe*
 cut-IMP.2PL beetroot(M)-ACC.SG.DEF in cube(M).DIR.PL and
zie-je-ni pak minuta.
 boil-3SG.ACC-IMP.2PL few minute(F).DIR.PL
 ‘Dice the beetroot and boil it for a few minutes.’
 (Panorama, 05.12.2015)

² According to the inventory of actional meanings as shown in Tatevosov (2002).

- (11) *Pasi uj-i ka zier, hidh-ni*
 after water(M)-NOM.SG.DEF have-PRS.3PL boil.PTCP throw-IMP.2PL
hithra-t dhe lër-i-ni të zie-jnë
 nettle(F)-ACC.SG.DEF and leave-3PL.ACC-IMP.2PL SBJV boil-PRS.3PL
për një minutë.
 for one minute(F).DIR.SG
 ‘Once the water has boiled, add the nettles and let them boil for a minute.’
 (*Agjencia Telegrafike Shqiptare*, 25.02.2017)
- (12) *Pasi të ke-në zier, hap-i-ni dhe*
 after SBJV have.SBJV-PRS.3PL boil.PTCP open-3PL.ACC-IMP.2PL and
pastro-ji-ni nga lëkur-a.
 clean-3PL.ACC-IMP.2PL from husk(F)-NOM.SG.DEF
 ‘Once they are boiled, open them and peel them.’
 (*Panorama*, 29.10.2016)
- (13) *Plak-a i=dha Pandë-s*
 old.woman(F)-NOM.SG.DEF 3SG.DAT=give.AOR.3SG Panda(M)-OBL.SG.DEF
një dash, të cil-in e=pre-u
 one ram(M).DIR.SG L which-M.ACC.SG 3SG.ACC=cut-AOR.3SG
dhe e=zje-u.
 and 3SG.ACC=boil-AOR.3SG
 ‘The old woman gave Panda a ram, which he slaughtered and cooked.’
 (*Shkëlqimi dhe rënia e shokut Zylo*, Dritëro Agolli, 1972)
- In 135 of 300 examples, the verb *ziej* is used figuratively to refer to agitated emotional states. Almost all of these examples contain intransitive usages (14), while non-actives (15) and active transitive forms (16) are rare (six and three examples, respectively).
- (14) *Sekretar-i i parti-së zie-nte*
 secretary(M)-NOM.SG.DEF L.M.NOM.SG party(F)-OBL.SG.DEF boil-IPF.3SG
nga zemërim-i.
 from anger(M)-NOM.SG.DEF
 ‘The party secretary was seething with anger.’
 (*Koha e dhive*, Luan Starova, 1993)
- (15) *E=di-ja mirë se çfarë zi-hej në*
 3SG.ACC=know-IPF.1SG well comp what boil-NACT.IPF.3SG in
kokë-n e tij.
 head(F)-ACC.SG.DEF L.ACC.SG his
 ‘I knew very well what was going on in his head.’
 (*Shkëlqimi i huaj*, Beqë Cufaj, 2003)
- (16) *Vetë ia=zie-nte gjak-un*
 self 3SG.DAT:3SG.ACC=boil-IPF.3SG blood(M)-ACC.SG.DEF
Kadri-ut të varfër...
 Kadri(M)-OBL.SG.DEF L.M.OBL.SG poor.M.SG
 ‘She herself made poor Kadri’s blood boil.’
 (*Panorama*, 03.11.2015)

No significant semantic differences between the more frequent active intransitive and the less frequent non-active forms can be observed at this stage; both have an anticausative interpretation. The non-active forms occur in both newspaper texts and literature. Overall, it seems that, for *ziej*, we can speak of genuine variation between non-active and active intransitive forms.

3.4 Deadjectival color verbs *zbardh* ‘whiten; brighten’ and *skuq* ‘make/color red; fry’

It has been noted that deadjectival verbs are quite often labile (Letuchiy 2013: 200–203), and this is partially confirmed by our data on the verbs *zbardh* ‘whiten; brighten’ and *skuq* ‘make/color red; fry’.

The main feature of the verbs *zbardh* and *skuq* is that, in modern Albanian, their direct color meaning appears in only a minority of their usages. For the verb *skuq*, the most frequent meaning (more than two-thirds of the examples) is that of a psychological reaction (‘to blush from shame, excitement, etc.’), covering a spectrum of meanings from describing purely physical manifestations of such a reaction (17) to indicating a psychological state (18). These meanings are mostly expressed by non-active forms, which are predominant for this verb (238 of 300 occurrences), with only 6 transitive usages and no attested active intransitive usages. A notable feature is the large number of aorist forms, reflecting the actional meaning of ‘entering a state’.

- (17) *Faqe-t* *e* *kuq-e iu=skuq-ën*
 cheek(F)-DIR.PL.DEF L.DIR.PL red-F 3SG.DAT:3SG.ACC=reddden-AOR.3PL
edhe më fort.
 yet CMPR strongly
 ‘His red cheeks became even redder.’
 (*Vjeshta e Xheladin Beut*, Mitrush Kuteli, 1943)

- (18) *Atëherë do t’ju vijë turp për lisat që keni dashur dhe do të skuqeni për kopshtet që keni zgjedhur.*
 ‘Then you will be ashamed of the oaks you loved and blush for the gardens you chose.’
 (*Isaia, Dhjata e vjetër*, 1994)

do të skuq-eni për kopshte-t
 FUT SBJV reddden-NACT.PRS.3SG for garden(MF).PL-DIR.PL.DEF
 ‘you will blush for the gardens’

For the verb *zbardh*, the numerically dominant usage (255 occurrences) conveys the meaning ‘to bring to light, clarify, reveal’ (19). Only non-active (intransitive) and active transitive usages are attested. The grammatical forms are diverse, including many non-active present forms found in newspaper headlines.

- (19) *Zbardh-et* *skem-a* *e* *re,*
 whiten-NACT.PRS.3SG sheme(F)-NOM.SG.DEF L.F.SG.NOM new.F.SG
si do llogarit-en pensione-t.
 how FUT calculate-NACT.PRS.3SG pension(MF).PL-DIR.PL.DEF
 ‘New scheme is revealed, how pensions will be calculated.’
 (*Panorama*, 14.04.2017)

Among the non-figurative meanings of *skuq* (if the psychological reaction meaning is considered figurative), a significant number of examples pertain to the meaning ‘to fry’, etc. (referring to food), frequently occurring in the imperative form in recipes. In the case of intransitive usages of *skuq* in its literal color meaning, non-active forms (20) predominate over active ones (21) with 41 to 10 instances. Interestingly, all 10 examples of active intransitive usage belong to literary works written no later than 1985, suggesting a somewhat literary (and perhaps slightly archaic) nature of this usage.

- (20) *Piq-et* *dardh-a,* *skuq-et*
 bake-NACT.PRS.3SG pear(F)-NOM.SG.DEF redden-NACT.PRS.3SG
moll-a...
 apple(F)-NOM.SG.DEF
 ‘Pears ripen, apples redden...’
 (Pamje katundare, Asdreni, 1938)

- (21) *Mbi mjekra-t* *e* *tyre të shkurtr-a*
 on beard(F).PL-DIR.PL.DEF L.DIR.PL their L short-F.PL
këna-ja *skuq-te* *kërcënueshëm.*
 henna(F)-NOM.SG.DEF redden-IPF.3SG menacingly
 ‘On their short beards, henna was dyed menacingly red.’
 (Ura me tri harqe, Ismail Kadare, 1978)

For the verb *zbardh*, in its non-figurative meaning, only non-active and active intransitive forms are attested. Of the 25 intransitive usages, 16 describe sunrise and are semi-idiomatic expressions (22). There is also one semantically similar non-active usage. The remaining nine intransitive direct usages (23), found both in literary works and newspapers, actionally denote state and entering a state and compete with more frequent non-active forms (24), which total 19 occurrences.

- (22) *Kish-te* *një orë* *e ca*
 have.IPF-IPF.3SG one hour(F).DIR.SG and some
që kish-te *zbardhur* *dit-a.*
 COMP have.IPF-IPF.3SG whiten.PTCP day(F)-NOM.SG.DEF
 ‘It had been an hour or so since daylight.’
 (Lëkura e daullës, Ismail Kadare, 1967)

- (23) *Njëqind* *e ca* *petë-t* *e* *mbaruar-a*
 one.hundred and some noodle(F)-DIR.PL.DEF L.DIR.PL finished-F.PL
zbardh-nin *të vën-a* *një mbi një mbi*
 whiten-IPF.3SG L stacked-F.PL one on one on
ca sofrë *drur-i.*
 some table(F).DIR.PL wood(M)-OBL.SG
 ‘A hundred or so finished noodles were whitening, stacked one on top of the other on some wooden tables.’
 (Komisioni i festës, Ismail Kadare, 1978)

- (24) *Natë-n e krishtlindje-ve Lesbos-i*
 night(F)-ACC.SG.DEF L.ACC.SG Christmas(F)-OBL.PL Lesbos(M)-NOM.SG.DEF
u zbardh nga dëbor-a.
 NACT whiten.NACT.AOR.3SG from snow(F)-NOM.SG.DEF
 ‘On Christmas Eve, Lesbos was whitened with snow.’
 (Të jetosh në ishull, Ben Blushi, 2008)

Overall, it can be cautiously suggested, that in modern Albanian, deadjectival color verbs are losing their labile character. This is likely related to the tendency of these verbs to be used more frequently not in their direct color-related meaning, favoring intransitive usage, but in their other meanings, e.g., ‘fry’ for *skuq* and ‘bring to light’ for *zbardh*, for which active transitive uses and their non-active counterparts are more natural.

3.5 The verbs *dalloj* ‘distinguish; differentiate; differ’ and *ndryshoj* ‘change; vary’

The verb *dalloj* ‘distinguish; differentiate; differ’ is basically a transitive verb, as seen from the proportion of active intransitive, active transitive, and non-active forms in Table 1. Transitive forms and their non-active counterparts prevail in the sample, along with a moderate percentage of examples in which *dalloj* is used intransitively. In contrast, *ndryshoj* ‘change; vary’ functions as a verb with almost equal numbers of active transitive and intransitive usages and a low number of non-active forms.

Usages of the verb *dalloj* are of two main types. In the first type the verb means ‘distinguish, manage to discern, be distinguished’ with the direct meaning typically referring to visual perception (25–26). When used metaphorically (27–28), *dalloj* is often accompanied by an argument introduced by the preposition *për* ‘for’. Non-active forms and a few active intransitive forms with this meaning have an anticausative interpretation (26, 28).

- (25) *Tani e=dallo-va një-rë-n nga dy*
 now 3SG.ACC=discern-AOR.1SG one.of.two-F.ACC.SG from two
silueta-t te dritar-ja.
 silhouette(F)-DIR.PL.DEF at window(F)-NOM.SG.DEF
 ‘Now I recognized one of the two silhouettes at the window.’
 (Muzgu i perëndive të stepës, Ismail Kadare, 1978)

- (26) *Përmatanë përro-it tani dallo-heshin qartë*
 across stream(M)-OBL.SG.DEF now discern-NACT.IPF.3PL clearly
tre silueta.
 three silhouette(F).DIR.PL
 ‘Across the stream, three silhouettes were now clearly visible.’
 (Yjet ndritin lart, Naum Prifti, 2002)

- (27) *Ju jeni një kritik i ashpër i Kadaresë, a dalloni ju vlera të rëndësishme në veprën e tij?*
 ‘You are a harsh critic of Kadare, do you recognize important values in his work?’
 (Panorama, 22.09.2017)

a dallo-ni ju vlera të rëndësishm-e
 Q discern-PRS.2PL you value(F).DIR.PL L important-F
 ‘do you recognize important values?’

- (28) *Gjatë viteve 1970, Sali Berisha u dallua për punën kërkimore në fushën e kardiologjisë në Shqipëri dhe u emërua profesor kardiologjie në Universitetin e Tiranës.*
 ‘During the 1970s, Sali Berisha distinguished himself for his research work in the field of cardiology in Albania and was appointed professor of cardiology at the University of Tirana.’

(*Panorama*, 15.10.2017)

Sali Berisha u dallua për punë-n kërkimor-e
 Sali Berisha NACT discern.NACT. AOR.3SG for work(F)-ACC.SG.DEF research-F
 ‘Sali Berisha distinguished himself for his research work’

In the second type of usage, *dalloj* functions as a lexical reciprocal with the meaning ‘to distinguish, to differentiate’ (on lexical reciprocals, see Knjazev 2007). In its transitive usage, this reciprocal has an object-oriented character (29), whereas in its intransitive usage, it is subject-oriented (30–31). As a rule, both transitive and intransitive uses of this verb appear in the so-called discontinuous construction, with the syntactic dominance of one of the participants. The subdominant participant is introduced by the prepositions *nga* (31), *prej* ‘from’ (30) and *me* ‘with’ (29). However, there are also two examples of the so-called simple reciprocal construction, where both participants are joined by a coordinating conjunction or expressed by a noun phrase in the plural (32). The intransitive examples of reciprocal use of the verb *dalloj* have an anticausative interpretation.

- (29) *Në Shqipëri s’e=dallo-jnë flamur-in*
 in Albania(F).DIR.SG NEG:3SG.ACC=discern-PRS.3PL flag(M)-ACC.SG.DEF
e Kosovë-s me të Bosnjë-s.
 L.ACC.SG Kosovo(F)-OBL.SG.DEF with L Bosnia(F)-OBL.SG.DEF
 ‘In Albania, they don’t distinguish between the Kosovo flag and the Bosnian flag.’
 (*Koha.mk*, 21.07.2012)

- (30) *Këtu parimisht dallo-jmë ne prej jush <...>.*
 here fundamentally discern-PRS.1PL we.NOM from you.ABL
 ‘This is where we fundamentally differ from you <...>.’
 (*A.z via Koha.mk*, 06.03.2012)

- (31) *Ilirët flisnin një gjuhë që dallohej nga gjuhët e popujve të tjerë të kohës së lashtë të Ballkanit.*
 ‘The Illyrians spoke a language that was distinct from the languages of other ancient Balkan peoples.’
 (*Historia e popullit shqiptar. Ilirët*, Muzafer Korkuti et al., 2002)

një gjuhë që dallo-hej nga
 INDF language(F).DIR.SG REL discern-NA.IPF.3SG from
gjuhë-t e popuj-ve të tjerë
 language(F)-DIR.PL.DEF L.DIR.PL people(M)-OBL.PL L other.M.DIR.PL
 ‘a language that was distinct from the languages of other peoples’

- (32) *Këto pjesë të qytet-it dallojnë edhe nga infrastruktur-a.*
 this.F.DIR.PL part(F).DIR.PL L city(M)-OBL.SG.DEF discern-PRS.3PL
edhe nga infrastruktur-a.
 also from infrastructure(F)-NOM.SG.DEF
 ‘These parts of the city also distinguish in infrastructure.’
 (*Koha.mk*, 21.03.2012)

Table 3 and Figure 2 draw attention to the different distribution of the reciprocal meaning in terms of the opposition between active intransitive and non-active forms. Among the 128 passive forms, we observe 20 reciprocals, while among the 48 active intransitive forms, there are 37 reciprocals. From this distribution, at the very least, we can conclude that the active intransitive usages of the verb *dalloj* rarely express the more common non-reciprocal meaning of this verb. However, this issue requires further investigation.

Table 3. Reciprocal and non-reciprocal usages of the verb *dalloj* (first 300 examples, numbers rounded to two decimal places)

	number of examples	intransitive		transitive		non-active	
		number	share	number	share	number	share
reciprocal	81	37	0.46	24	0.30	20	0.25
non-reciprocal	219	11	0.05	100	0.46	108	0.49
Total	300	48	0.16	124	0.41	128	0.43

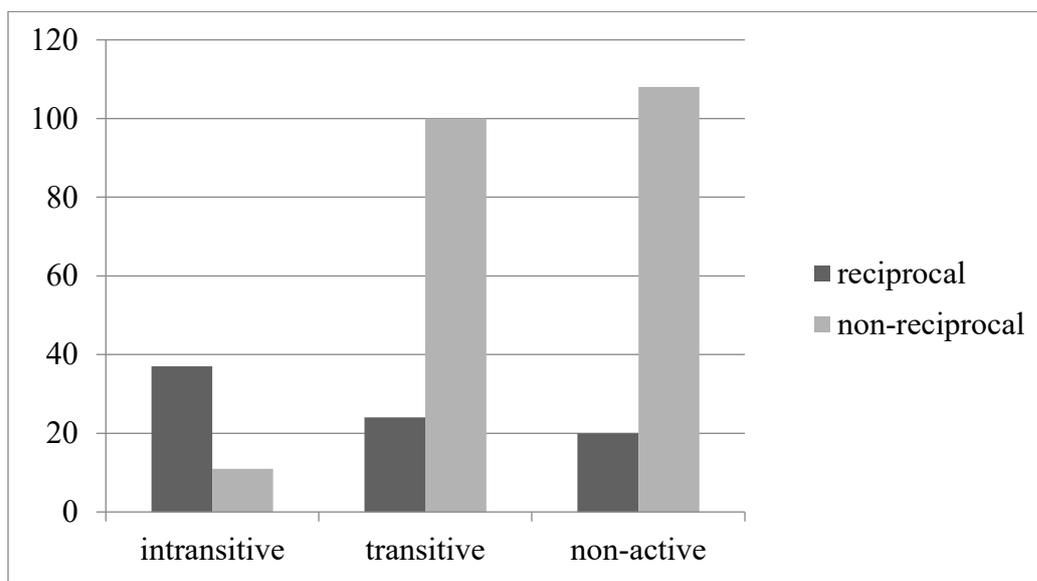


Figure 2. Reciprocal and non-reciprocal usages of the verb *dalloj* (first 300 examples)

For the verb *ndryshoj*, as mentioned above, active intransitive usages significantly outnumber the non-active ones (138 to 25; see Table 1). Almost all active intransitive usages of *ndryshoj* have an anticausative interpretation (33–34), while the majority of non-active usages have a passive interpretation (35). Nearly all non-active examples originate from newspapers and official documents. Imperfective and present tense forms of *ndryshoj* unambiguously refer to a process (33) or state (36), while perfective forms (aorist and perfect) refer to a change of state, which results in a new state (34). In our sample of 300 examples, the perfect dominated over the aorist (57 and 23 examples, respectively) and the present tense, often accompanied by the continuative particle *po*, dominated over the imperfect (77 and 8

examples), due to the high representation of newspapers in the ANC, in which “hot news perfect” is widespread (see Rusakov and Morozova 2014), and present rather than past states of affairs are usually described.

- (33) *Koncepti i sovranitetit kombëtar në kuadrin e globalizimit të botës dhe të integritimit Evropian po ndryshon.*

‘The concept of national sovereignty in the context of world globalization and European integration is changing.’

(*Gazeta Shqiptare*, 17.11.2006)

<i>koncept-i</i>	<...>	<i>po</i>	<i>ndrysho-n</i>
concept(M)-NOM.SG.DEF		PROG	change-PRS.3SG
‘the concept <...> is changing’			

- (34) *Koh-a ka ndryshuar.*
time(F)-NOM.SG.DEF have.PRS.3SG change.PTCP
‘The time has changed.’

(*Koha.mk*, 18.04.2011)

- (35) *Ndërkaq nga radhë-t e socialistë-ve do të ndrysho-het ministr-i i transport-it.*
meanwhile from row(F)-DIR.PL.DEF L.DIR.PL Socialist(M)-OBL.PL FUT
SBJV change-NACT.PRS.3SG minister(M)-NOM.SG.DEF L.M.NOM.SG
transport(M)-OBL.SG.DEF

‘Meanwhile, from among the Socialists the Minister of Transport will be changed.’

(*Rel via Zëri*, 30.07.2013)

One of the possible meanings of the verb *ndryshoj* is ‘distinguish, differ’, which is synonymous with *dalloj*. However, usages like (36) turned out to be infrequent in our sample (23 of 300 examples). All of them are active intransitive, as opposed to *dalloj*, which is used both transitively and intransitively in the same meaning. Almost half of these usages (11 of 23 examples) demonstrate lability of the reciprocal type (37), probably modeled on *dalloj*, which is described above.

- (36) *Dhe këto doreza ndrysho-jnë në gjatësi, materiale, ngjyra dhe dizajn.*
and this.F.DIR.PL glove(F).DIR.PL change-PRS.3PL in length(F).DIR.SG
material(M).DIR.PL color(F).DIR.PL and design(M).DIR.SG
‘And these gloves vary in length, material, color and design.’

(*Koha.mk*, 28.09.2012)

- (37) *Këto ëndrra ndryshonin nga njëra-tjetra, por, te secila prej tyre, heroi kryesorë ishin dhitë tona.*

‘These dreams differed from each other, but in each of them, the main heroes were our goats.’

(*Koha e dhive*, Luan Starova, 1993)

këto *ëndrra* *ndrysho-nin* *nga një-r-a*
 this.F.DIR.PL dream(F).DIR.PL change-IPF.3PL from one.of.two-F.NOM.SG
tjetr-a
 other-F.NOM.SG
 ‘these dreams differed from each other’

It can be stated that *ndryshoj* is a labile verb with a secondary development of non-active usages in bureaucratic speech style.

4 Conclusion

A pilot corpus-based study of eight verbs demonstrating P-lability has shown that Albanian verbs behave differently in this respect. Phasal verbs *filloj* ‘begin’ and *mbaroj* ‘finish’ exhibit lability quite consistently. The motion verb *lëviz* ‘move’ is also a labile verb, with a predominance of active intransitive usages (presumably original) over active transitive and non-active ones. The verb *ziej* ‘boil; seethe; ferment’ displays variation between non-active and active intransitive usages (so far, no significant semantic differences between these usages have been established).

Deadjectival color verbs *zbardh* ‘whiten; brighten’ and *skuq* ‘make/color red; fry’ demonstrate lability only in their basic color-related meanings (which are relatively infrequent), while in their figurative uses, they employ morphologically non-active forms in intransitive contexts. The verbs of distinction and change *dalloj* ‘distinguish; differentiate’ and *ndryshoj* ‘change; vary, distinguish’ behave differently. *Dalloj* primarily employs active intransitive forms in reciprocal meanings and morphologically non-active forms in other meanings. In contrast, *ndryshoj* is a labile verb with a secondary development of non-active usages.

From a semantic perspective, most of our verbs exhibit an anticausative type of lability, as well as the rarer reciprocal mobility (*dalloj* and, to some extent, *ndryshoj*).

For the next stages of the research, the following objectives are set:

- expanding the range of verbs to cover all verbs demonstrating lability in Albanian, including those with other types of lability, primarily conversive lability;
- incorporating a diachronic aspect (tracing the development of lability from the earliest Albanian written records) and a diatopic aspect (comparing regional and dialectal varieties of Albanian in terms of lability expression);
- placing Albanian lability in the broader context of how lability is realized in Balkan languages.

Abbreviations

1, 2, 3 — person; ABL — ablative; ACC — accusative; AOR — aorist; COMP — complementizer; CMPR — comparative; DAT — dative; DEF — definite; DIR — direct case; F — feminine; FUT — future; IMP — imperative; INDF — indefinite; IPF — imperfect; M — masculine; MF — ambiguous noun; NACT — non-active; NEG — negation; NOM — nominative; OBL — oblique case; PL — plural; PROG — progressive; PRS — present; PTCP — participle; REL — relativizer; SBJV — subjunctive; SG — singular.

Bibliography

- ANC: Morozova, M., Rusakov, A. and Arkhangelskiy, T. *Albanian National Corpus*. Available online at: albanian.web-corpora.net (accessed on 26.03.2025).
- Bužarovska, E. and Mitkovska, L. (2022). *On the lability of verbs in Balkan languages*. Paper presented at the 22nd Balkan and South Slavic Conference, The Ohio State University, 7–10 April 2022.
- Dhrimo, A. (1965). Një vështrim mbi foljet me më se një reksion në shqipen e sotme letrare [A look at verbs with more than one rection on the modern literary Albanian]. *Studime filologjike*, 2: 139–151.
- Diveeva, V. (2013). *Labil'nost' albanskikh glagolov v balkanskom kontekste* [Lability of Albanian verbs in the Balkan context]. Course paper. Saint Petersburg State University.
- Friedman, V. (2010). Lability as a Scalar Balkanism. In P. Assenova, A. Petrova and Ts. Ivanova (eds.). *The Verbal System of the Balkan Languages — Heritage and Neology*, 63–69. Veliko Tŕrnovo: Faber.
- Friedman, V. A. and Joseph, B. D. (2025). *The Balkan Languages*. Cambridge: Cambridge University Press.
- Haspelmath, M. (2001). 107. The European linguistic area: Standard Average European. In 2. *Halbband Language Typology and Language Universals 2. Teilband*, 1492–1510. Berlin/Boston: de Gruyter. <https://doi.org/10.1515/9783110194265-044>.
- Knjazev, J. P. (2007). Lexical reciprocals as a means of expressing reciprocal situations In V. P. Nedjalkov (ed.). *Reciprocal constructions*, 115–146. Amsterdam: John Benjamins.
- Letuchiy, A. (2013). *Tipologija labil'nykh glagolov* [A typology of labile verbs]. Moscow: Jazyki slavjanskikh kul'tur.
- Makartsev, M., Wahlström, M. and Escher, A. (2024). Lability in Balkan Slavic. *Studies in Language*. <https://doi.org/10.1075/sl.23062.mak> (published online: 19 December 2024).
- Nichols, J. (2020). The Balkan sprachbund in typological-geographical space. *Acta Linguistica Petropolitana*, 16 (2): 306–330.
- Nichols, J., Peterson, D. A. and Barnes, J. (2004). Transitivity and detransitivizing languages. *Linguistic Typology*, 8 (2): 149–211.
- Rusakov, A. and Morozova, M. (2014). Perfekti në shqipen e sotme: studimi i bazuar në tekstet e Korpusit elektronik të shqipes [Perfect in modern Albanian: A study based on the texts from an electronic Albanian corpus]. In B. Rugova (ed.). *Seminari Ndërkombëtar për Gjuhën, Letërsinë dhe Kulturën Shqiptare. Materialet e punimeve të Seminarit XXXII Ndërkombëtar për Gjuhën, Letërsinë dhe Kulturën Shqiptare: Prishtinë, 19–30.08.2013*. Vëll. 33/1, 97–104. Prishtinë: Universiteti i Prishtinës: Fakulteti i filologjisë.
- Tatevosov, S. (2002). The parameter of actionality. *Linguistic Typology*, 6: 317–401.

EVENT MODALITY IN BALKAN TURKISH: FORMAL AND SEMANTIC VARIATION IN CONTACT Part 1. Possibility and Necessity

Julian Rentzsch

Johannes Gutenberg University Mainz
rentzsch@uni-mainz.de

The present article constitutes the first part of a two-part study on event modality in selected Turkish varieties of Kosovo, North Macedonia, Bulgaria, and Eastern Thrace (Turkey). The linguistic structures will be compared with corresponding expressions in Modern Standard Turkish and pre-modern Ottoman Turkish varieties. The study identifies both common features and differences among the Balkan Turkish varieties. Variation occurs in different slots within the investigated constructions and concerns lexical, semantic and morphosyntactic features, including complementation patterns, where both infinitive and subjunctive structures can be found. The linguistic variation is partly dialect-specific and distributed differently among the eastern and western dialects of Balkan Turkish, but intra-dialectal variation is also observed. It will be argued that while some processes that have led to the present situation in Balkan Turkish may be attributed to internal developments of inherited structures and to universal tendencies, impact of language contact has also contributed to the distribution of certain structures within the Turkish dialects of the Balkans.

Keywords: dialectology, semantics, morphosyntax, complementation patterns.

**ПРЕДИКАЦИСКАТА МОДАЛНОСТ ВО БАЛКАНСКИОТ ТУРСКИ
ЈАЗИК: ФОРМАЛНА И СЕМАНТИЧКА ВАРИЈАНТНОСТ
ВО ЈАЗИЧЕН КОНТАКТ**
Дел 1. Веројатност и нужност

Јулијан Ренч

Универзитет Јоханес Гутенберг, Мајнц
rentzsch@uni-mainz.de

Оваа статија го претставува првиот дел од истражувањето на модалноста на настаните, изразена во одредени турски варијанти од Косово, Северна Македонија, Бугарија и од Источна Тракија (Турција). Јазичните структури се споредуваат со нивните функционални еквиваленти во модерниот стандарден турски јазик и со постојните варијанти во османлискиот турски јазик. Во истражувањето ги определуваме заедничките карактеристики и разлики меѓу балканските турски јазични варијанти. Варијантноста е карактеристична за различни синтаксички позиции на истражуваните конструкции и се однесува на лексички, семантички и на морфосинтаксички особини. Таа се појавува и на ниво на комплементација, како инфинитивна така и субјунктивна. Јазичната варијација е делумно присутна во дијалектите на балканскиот турски јазик, но со различна застапеност во источните и во западните дијалекти. Во статијата се тврди дека некои процеси, коишто довеле до денешната ситуација во балканскиот турски јазик, можат да се должат на внатрешниот развој на наследените структури, но контактот со соседните јазици исто така придонел за постоење на одредени структури во турските дијалекти на Балканот.

Клучни зборови: дијалектологија, семантика, морфосинтакса, стратегии на комплементација.

1 Introduction

The present study investigates expressions of event modality (root modality) in Balkan Turkish.¹ Event modality, as used in this article, encompasses possibility, necessity, and volitive modality (cf. Palmer 2001; Rentzsch 2015). Some neighbouring semantic domains such as procedural and practical knowledge, as well as intention will be included in the study. The expressions under investigation are of particular interest for the documentation and analysis of the Balkan Turkish dialects, especially in terms of their similarities and differences compared to other Western Oghuz varieties. Moreover, they are also relevant in the context of Balkan languages, as certain features of Balkan Turkish may be attributed to contact with neighbouring languages.

Special interest will be dedicated to more or less transparent morphosyntactic constructions and the problem of linguistic variation. The modal constructions under investigation generally involve a morphosyntactic mechanism of complementation (cf. Dixon 2006; Noonan 2007; Achard 2007), i.e., a strategy for linking the lexical element contributing the modal value to the state of affairs (SoA) over which it scopes. The range of formal realizations covers, on the one hand, constructions consisting of a matrix clause and a subordinate clause, and, on the other hand, auxiliary constructions. The boundary between these two poles is fluid as there are constructions that can be conceptualized in either ways, as will become evident in the course of this paper.² In order to have a terminological tool at hand that covers various degrees of syntactic integration, both predicates of matrix clauses and auxiliaries will be labelled in this paper with the umbrella term *matrix segment*. The item effectuating the connection of the matrix segment to the SoA which it scopes over will be labelled *linking segment*. In Balkan Turkish, as in Turkic languages in general, matrix segments may be nominal and verbal. SoAs minimally consist of a predicate, but may also include arguments and adjuncts. The linking segments may be of various kinds, including case-marked or unmarked verbal nouns and finite mood forms functioning as subjunctives. Complementizer particles also occur, although they play a minor role.

Given the structural complexity of the expressions of modality, there is a considerable potential for linguistic variation. Variation may concern the matrix segments, the linking segments and complementation strategies, as well as the degree of conventionalization (grammaticalization or idiomaticization).

The database for this study consists of dialect texts from Kosovo, North Macedonia, Bulgaria, and Eastern Thrace. The modal constructions will be described structurally, and questions of heritage, universal tendencies of linguistic change, and language contact will be discussed. The focus will be on synchronic data but a limited amount of historical data will be added to provide the diachronic context. Besides Old and Middle Ottoman data from the 15th to 17th century, dialect texts from Adakale³ collected by Ignác Kúnos between 1890 and 1895 (Kúnos 1907), and from Vidin (Bulgaria) collected by Gyula [Julius] Németh in 1931 (Németh 1965) represent more recent historical data. In order to convey an idea of the specific Balkan Turkish features, the corresponding Modern Standard Turkish constructions will also be shown for the sake of comparison. The present paper elaborates on phenomena mentioned in work

¹ The investigation on root modality in Balkan Turkish is divided into two parts: the first discusses possibility and necessity in the present article, whereas the second part focusing on volitive modality will be published in the next issue of the *Journal of Contemporary Philology*.

² Constructions of a matrix clause and a complement clause are biclausal by definition, while auxiliary constructions are usually considered monoclausal.

³ Adakale was an island in the Danube River that was depopulated in 1968 because of the construction of a dam and became submerged in 1971. The dialect, which belonged to Western Rumelian Turkish, is extinct.

such as Menz's study on Gagauz (1999) – an Eastern Rumelian Turkish variety with its own standard language.⁴

To narrow down the rich material, the study will be limited to a selection of matrix elements comprising *bil-* 'to know', *mümkün* 'possible', and *yok* 'absent' for the domain of possibility, *lazım* 'necessary', and *var* 'present, available' for necessity, and *iste-* 'to want', *dile-* 'to wish', and *niyet* 'intention' for the domain of volitive modality. A few other items will be touched upon to supply further relevant information.

In the course of the description and analysis, some terms will be used that are potentially ambiguous, or used in various ways across the literature. There are several approaches to the notion connected to the term *infinitive* in historical and comparative linguistics. In Turkish Studies the conventional use of this term can be particularly misleading. Among the linguistic approaches which have informed the present study is a historical account proposed by Haspelmath (1985), which essentially treats the infinitive as a category that evolved from a purposive verbal noun.⁵ In addition, a functional approach developed by Joseph (1983), with particular reference to the Balkans, has also been influential. It describes infinitives as non-finite verb forms that perform typical functions, such as expressions of purpose, systematically occupying complement slots of verbs and adjectives, and potentially fulfilling additional functions language-specifically (Joseph 1983: 30–36).

Both approaches are fruitful for the discussion in this paper, and, although departing from different perspectives, they are ultimately compatible. Note that the conventional use of the term infinitive in Turkish Studies differs significantly from these frameworks. In this tradition, the term infinitive refers to the morpheme *-mAK*, and (depending on the author) possibly to a shorter morpheme *-mA*,⁶ which may occur in some auxiliary constructions (including constructions relevant for this paper) but do not entail the purposive component, neither diachronically nor synchronically. Rather, these forms are plain non-factual verbal nouns lacking case marking and purposive semantics, though they can take nominal inflectional elements such as possessive and case markers. Using the term infinitive for these items is misleading from a general linguistic point of view and will be avoided in this paper. On the other hand, various items in certain Turkic languages, including Balkan Turkish, could be appropriately described as infinitive in terms of both Haspelmath's and of Joseph's frameworks. In the specific Balkan Turkish context, these include the dative forms of the verbal nouns *-mAK* and *-mA*,⁷ surfacing as *-mAGA* (predominantly in the west) and *-mAyA* (predominantly in the east), respectively. What makes the situation particularly confusing is the fact that the form *-mAGA* (i.e. verbal noun *-mAK* plus dative *-(y)A*) may frequently undergo a formal reduction to *-mĀ* and even *-mA*, in the extreme case resulting in a form which looks identical to the unmarked, short verbal noun in *-mA*. The synchronic distinction between these two forms is evident in their combinability: the plain verbal noun in *-mA* can take possessive and case markers, whereas the form *-mA* derived from *-mAGA* does not allow any additional suffixation. For the sake of clarity, I will use expressions such as "verbal noun in the dative"

⁴ See especially Menz 1999: 47–66), Friedman's description of subjunctive-type constructions in Western Rumelian Turkish and their parallels in Macedonian and Albanian (2003: 62–64; 2006: 38), as well as Römer's investigation (2012) of notable dative complements in Middle Ottoman texts of the 16th century.

⁵ Haspelmath (1985: 288) speaks of "purposive action nominal", without saying much about the syntactic functions of the items.

⁶ This term is widespread in the grammars, dictionaries and teaching manuals of Turkish. Among the linguistic literature consulted for this study, it is also adopted in Brendemoen (2014) and partly in Brendemoen (2013).

⁷ As the shorter form *-mA* gained a wider dissemination in Ottoman Turkish only during the 17th century, as Brendemoen (2014) has shown, the morpheme *-mA* has only a limited distribution in the western dialects of Balkan Turkish. However, it is extremely productive in Standard Turkish, where it is always preferred against *-mAK* when possessive suffixes are added, and in most case forms. The combinability of *-mAK* in ST is basically limited to the dative and the ablative, although it also frequently appears in an unmarked form.

or “the item **-mAGA*” in this paper and use the term infinitive only in quotation marks, or as a term to refer to a specific subordination strategy.

The term *subjunctive* will be used as an umbrella term for finite mood forms whose function in a given construction is to subordinate a verb to a matrix segment (whether it be a matrix clause or an auxiliary). The inventory of forms suitable for this function includes mood forms traditionally labelled *voluntative* (in Turkish, these are the first person singular and plural items *-(y)AyIm* and *-(y)Allm* and the third person singular and plural items *-sIn* and *-sInlAr*) and *optative* (in Turkish *-(y)A* plus personal markers) in the Turcology literature. These items can also serve as predicates in independent sentences, where they encode meanings such as willingness, readiness, desire, obligation, etc. (cf. Rentzsch 2015: 173). These original meanings are bleached in their subjunctive function. Importantly, then, the term subjunctive in this study does not refer to a specific morphological class but rather to finite items that function as linking elements between the matrix segment and the SoA.

Finally, the term *aorist* must be commented on, which is of marginal relevance in this study but occasionally will be used to refer to a specific morphological class of verb forms in *-Ir*, *-Ar*, and *-r* in Turkish. The term is well established in Turkish and Turkic Studies (cf., e.g., Lewis 1967: 115; Göksel and Kerslake 2005: 339). However, semantically, these morphological forms differ significantly from the aorist in Ancient Greek or Balkan Slavic, where the terms denote a completed past (preterite), to put it simply (cf. Friedman 2003: 128–131). Since the category labelled “aorist” in Turkish is not central to the analysis presented here, I see no need to introduce an alternative term or to delve into its semantic nuances.

To provide an approximate picture of the situation of Balkan Turkish within the Turkic language family, it should be noted that the Balkan Turkish dialects can be subclassified into Eastern Rumelian Turkish (ERT) and Western Rumelian Turkish (WRT). These two subbranches are not sharply delimited; Németh (1956) defines some distinctive criteria for WRT which, however, can be disputed in detail.⁸ Together with Anatolian Turkish and Azerbaijani, Balkan Turkish constitutes the western branch of the Oghuz languages. The Oghuz branch (or South West Turkic) furthermore includes Turkmen, Khorasan Turkic, and South Oghuz languages such as Qashqai. The Western Oghuz languages Turkish, Gagauz, and Azerbaijani have standard norms. Among those, Turkish and Gagauz essentially represent Eastern Rumelian varieties.⁹

This study considers both Western Rumelian and Eastern Rumelian varieties of Balkan Turkish. The primary database consists of text collections of selected varieties in Kosovo, North Macedonia, Bulgaria, and Turkey. The sources include both published books and unpublished MA and doctoral dissertations. The transcription principles of these works differ considerably. In some cases, the accuracy and adequacy of sound representation may be open to question. However, this issue is of minor importance for the present study, which primarily focuses on morphosyntactic constructions. The transcription system used in this study is a broad transcription loosely based on the orthographic conventions of Standard Turkish, supplemented with additional symbols to indicate important phonetic features. For the sake of brevity, when a given structure is attested both in WRT and in ERT, fully glossed examples will be drawn from Western Rumelian Turkish, while representative ERT equivalents will be provided in brackets without glossing.

The sources provide only a limited picture, as not all dialects and variations are covered. When I refer to “attested” items, I mean those occurring in the text corpus. The absence of a structure may be due to the corpus’s limitations or an oversight.

⁸ This is partly due to the fact that the Turkish dialects of Bulgaria and North Macedonia were still insufficiently documented when Németh wrote his study.

⁹ For details on the internal classification of Oghuz Turkic, see Doerfer (1990).

Frequent comparison with Standard Turkish (ST) in this paper does not mean that ST is considered the structural “original” from which Balkan Turkish is a “deviation”. Rather, both varieties result from distinct, though partly interacting, historical developments within their specific contact settings and may feature conservativisms and innovations in different domains.

Certain similarities may result from either shared innovations or common heritage. ST inevitably exerts a certain influence on Balkan Turkish, which has been growing due to schooling, mass media, and increased mobility and communication. On the other hand, given the massive impact of Istanbul Turkish in the formation of Standard Turkish, Balkan Turkish has also contributed to the development of both spoken and literary standard language in the late Ottoman Empire and the Republic of Turkey, as Istanbul Turkish itself is an Eastern Rumelian dialect.

Some remarks on notation principles: Capital letters in morpheme writings represent rule-based sound changes, e.g. *-mAGA* = {-*ma**γ**a*; -*me**γ**e*}; *-DA* = {-*da*; -*de*; -*ta*; -*te*}. Letters in parentheses represent sounds that occur in certain phonological environments, e.g. *-(y)A* = {-*ya*; -*ye*} after vowels and {-*a*; -*e*} after consonants. An asterisk (*) denotes archiforms or reconstructed forms in this paper; e.g., **-mAGA* represents forms such as {-*ma**γ**a*/*-me**γ**e*; -*mā*/*-mā*, -*ma*/*-me*}, etc., regardless of whether the archiform (in this case *-ma**γ**a*/*-me**γ**e*) is attested.

2 Possibility

Turkish has a fully grammaticalized marker of possibility: *-(y)Abil-* in its positive, unnegated form and *-(y)AmA-* in its negated form. In the unnegated form, the historical origin in the converb *-(y)A* and the auxiliary verb *bil-* (originally ‘to know’) remains transparent. However, the two components are rather strongly fused, with only the particle *da/de* ‘too’ able to intervene. The historical predecessor of *-(y)Abil-* has been sporadically attested in Turkic since ca. the 11th century (cf. Rentzsch 2014: 361) and became especially widespread in Oghuz Turkic, though similar forms also occur in other branches of Turkic. One hypothesis for the origin of the negated form *-(y)AmA-* is that it arose from a converb combined with the negation form of an obsolete verb *u-* ‘to be able’. This construction is very old and is firmly attested in Old Uyghur from around the 9th century.

The possibility markers *-(y)Abil-* and *-(y)AmA-* cover a broad field on the semantic map of possibility, ranging from ability to participant-external possibility and deontic possibility (permission). It is also used in epistemic expressions.¹⁰

These markers are common across all varieties of Balkan Turkish, both Eastern and Western Rumelian, and have been attested in Anatolia from the oldest written sources. In Old Anatolian Turkish, negated forms also appear with the negation suffix attached to the auxiliary, i.e. *-(y)V bilme-*. Such negation forms are nowadays frequent in Azerbaijani but they are not typical of Rumelian Turkish.

Although the exact distribution of functions between the possibility markers *-(y)Abil-* and *-(y)AmA-* on the one hand, and competing forms, on the other, may vary among varieties and even be subject to dialect-internal variation (cf. Rentzsch, Mitkovska and Nedelkoska 2020 for the Ohrid dialect), these items will not be considered further in this study. They are firmly established throughout both Anatolian and Rumelian Turkish and do not display any exciting variation in our dialect material. There are, however, other constructions denoting possibility involving the auxiliary verb *bil-* ‘to know’. Semantically, these constructions usually encode learned and/or inherent skills (the precise semantic profile has to be established language- and

¹⁰ See van der Auwera and Plungian (1998) for the terminology.

dialect-specifically), i.e., domains of participant-internal possibility. These constructions show considerable formal variation in regard to the linking segment.

Since *bil-* as a full verb is transitive and governs the accusative or the unmarked case,¹¹ the auxiliary construction involving this verb in ST generally uses the non-factual verbal noun *-mA* in the accusative as a linking segment, resulting in the construction *-mAyI bil-*, as shown in example (1). Furthermore, a possessive suffix of the third person may be added (*-mAsInI bil-*) without a noticeable change in meaning, as in example (2).

- (1) *Sen surat oku-ma-yı bil-ir mi-sin?*
 you face read-VN-ACC know-AOR Q-2.SG
 ‘Can you read faces?’
 (ST, Pamuk 1990: 195)
- (2) *Kuran bu konu-da yalnızca harf-ler-i oku-ma-sın-ı*
 Qur’an DEM topic-LOC only letter-PL-ACC read-VN-POSS.3-ACC
bil-en-ler için açık.
 know-PTCP-PL for clear
 ‘The Qur’an is clear about this only for those who know how to read the letters.’
 (ST, Pamuk 1990: 151)

These two constructions seem to be rare in the Rumelian dialects, although an instance of *-mAyI bil-* is attested in a Turkish variety spoken in the Central Rhodopes:

- (3) *[Duva] kıl-ma-yı da bil-ir-im aşā yukarı*
 prayer do-VN-ACC too know-AOR-1.SG more.or.less
kendi-m-e kadar.
 self-POSS.2.SG-DAT until
 ‘I also know how to pray more or less, according to my abilities.’
 (BG/Rhodopes/Karabulak, Mustafa-Rashidova 2024: 220)

More frequently, however, we find *bil-* combined with the unmarked verbal noun in *-mAK*, a construction that resembles the expression of wish in *-mAK iste-* (see part 2). This construction *-mAK bil-* is attested both in Eastern and Western Rumelian dialects, although not frequently.

- (4) *Ben oku-mak bil-mëm, yaz-mak bil-mëm*
 I read-VN know-NEG.AOR.1.SG write-VN know-NEG.AOR.1.SG
nasıl müneccim ol-ayëm.
 how astrologer be/come-VOL.1.SG
 ‘I cannot read, I cannot write, how could I become an astrologer?’
 (MK/Ohrid, Kakuk 1972: 261)
 [cf. ERT: *İlle var bi tane turun ne duy-mak bil-ir* ‘Anyhow, there is one grandchild that knows to listen’ (TR/Edirne/Uzunköprü, Kalay 1998: 248)]

In the Balkans, the same verbal noun in the dative is more widespread as a linking segment in this type of construction. The linking segment may, depending on dialectal and idiolectal parameters, either appear in its full form *-mAGA* or in contracted forms such as *-mĀ* or *-mA*. The construction **-mAGA bil-* is attested both in the East and the West of the Balkan Peninsula,

¹¹ Turkish has differential object marking.

as evidenced by the examples from Eastern Thrace and Kosovo. However, no instances have been identified in North Macedonia.

- (5) *Cid-alım çağır-alım cüzel hanım-i cür-sün*
 go-VOL.1.PL call-VOL.1.PL beautiful lady-ACC see-VOL.3
nasıl bun-lar bül-me-y=miş-le ekmek ye-ma.
 how DEM-PL know-NEG-PROG=EVID-PL bread eat-VN.DAT
 ‘Let us go and summon the beautiful lady so that she sees that [“how”] they are unable to eat bread.’
 (RKS/Mamusha, Hafız 1985: 235)
 [cf. ERT: *E motor hayda-mā bil-en var mı?* ‘Is there anybody who can drive a motorbike?’ (TR/Edirne/Enez, Kalay 1998: 202)]

The construction itself is not particularly recent as evidenced by the examples such as *Biz daha baba de-meg-e bil-meyiz* ‘We cannot yet say father [to him]’ from Adakale (Kúnos 1907: 261). Furthermore, it is also attested in the *Colloquia Familiaria Turcico-Latina* by Jakab Nagy de Harsányi (ed.: Hazai 1973), a so-called transcription text in Latin script from 1672, which probably represents a variety of Istanbul Turkish of that time: *Arpaşu bu vilajetlerde jap-mag-a bil-mezler* ‘They don’t know how to make beer here’ (Hazai 1973: 68).

Since *bil-* as a full verb governs direct objects, the shift toward the dative in this construction is noteworthy. It reflects a universal tendency in the development of infinitives, which originate in purposive verbal nouns. In fact, the form *-mAyA* (corresponding to *-mAGA* in Middle Ottoman) can also be used to form purpose clauses in Modern Standard Turkish.

In addition to this non-finite (“infinitival”) complementation strategy, Balkan Turkish also employs another strategy, in which a finite mood form (typically the voluntative in the first and third persons and the optative in the second persons) serves as a subordinator and linking segment. This strategy, which can be termed the subjunctive strategy, is attested with the auxiliary verb *bil-* in WRT:

- (6) *Ama çöyce bil-mez=miş lafet-sın,*
 but boy know-NEG.AOR=EVID speak-VOL.3
çok kirli imiş saç-lar-i, uzun tırnak-lar-i var imiş.
 very dirty EVID hair-PL.POSS.3 long nail-PL.POSS.3 present EVID
 ‘But the boy does not know to speak properly, his hair is very dirty and he has long fingernails.’
 (MK/Struga, Ahmed 2004: 319)

It is likely that the precursor to this construction can be traced to non-factual complement clauses governed by the matrix verb *bil-*, which contain a question word and express how, where, when, to whom or what shall be done, as in examples (7)–(9).

- (7) *Valla bil-me-y-m nasıl de-e-m, ādet.*
 by.god know-NEG-PROG-1.SG how say-OPT-1.SG custom
 ‘I don’t know how to say it [= how I shall say it], it is a custom.’
 (MK/Skopje, Erdem et al. 2024: 211)

- (8) *Bu kız hu yalnız ev-de kal-i, bil-me-y*
 DEM girl all.alone house-LOC remain-PROG.3 know-NEG-PROG.3
ne yap-sın, başla-y ağla-sın.
 what do-VOL.3 begin-PROG.3 weep-VOL.3
 ‘The girl is left home all alone, she does not know what to do [= what she shall do], and starts crying.’
 (RKS/Mitrovica, Hafız 1985: 204)
- (9) *Bil-mez=dın kim-e selam, kim-i sor-a-sın,*
 know-NEG.AOR=PST.2.SG who-DAT greet who-ACC ask-OPT-2.SG
kim-a ne ver-e-sın. boyle idi.
 who-DAT what give-OPT-2.SG such PST.3
 ‘You didn’t know whom to greet, whom to ask, whom to give what. It was like this.’
 (MK/Gostivar, Erdem et al. 2024: 250)

This type does not exist in ST and it is not attested in the ERT data investigated for this study. In ST, the same notions are expressed by a construction involving a question word, a prospective verbal noun *-(y)AcAK* with a possessive suffix and the accusative: *ne yap-acağın-ı bil-mi-yor* ‘s/he does not know what to do’ (cf. Rentzsch, Mitkovska and Nedelkoska 2020: 89–90).

While expressions with *bil-* ‘to know’ of the type just described clearly represent participant-internal renewals of the less specific possibility markers *-(y)Abil-* and *-(y)AmA-*, expressions based on the Arabic adjective *mumkin* ‘possible’ have been broadly attested across the Islamicized Turkic world since the early Middle Turkic era. These expressions usually cover participant-external domains of possibility and, in many Turkic languages, also have epistemic uses. In Modern Standard Turkish, two constructions dominate, one using the verbal noun *-mAK* in impersonal expressions (example [10]), and another using the verbal noun *-mA* with a possessive marker in expressions in which the projected performer is present (example [11]). The performer is encoded by a possessive suffix; if expressed overtly, it occurs in the genitive case to agree with the possessive marker.

- (10) *Üzerinde ‘polis’ bile yaz-ıyor, yanlış-mak mümkün değil.*
 on.it police even write-PROG err-VN possible NEG
 ‘It even says ‘police’ on it, it is not possible to go wrong.’
 (ST, Pamuk 1990: 147)
- (11) *Çok ünlü bir kadın yıldız-ın [...] bir hanımefendi olarak*
 very famous one woman star-GEN one lady as
film hayat-ın-a devam et-me-si de mümkün=dü.
 film life-POSS.3-DAT continue-VN-POSS.3 too possible=PST
 ‘It was even possible for a very famous female star to continue her film career as a lady.’
 (ST, Pamuk 2008: 366)

The range of morphosyntactic variants is considerably broader both diachronically and synchronically. Example (12) from North Macedonia represents the subjunctive complementation strategy, with the third person voluntative functioning as a linking segment between the auxiliary and the main verb. Notably, the negation is marked by *yok* ‘absent’, which usually operates on nouns rather than adjectives. Additionally, *mümkün* carries a possessive suffix. Both facts suggest that *mümkün* in this example is structurally interpreted as

a noun.¹² In Adakale, the complementizer *ki* occasionally occurs between *mümkün* and the subordinate clause, a construction not found in the recent data.¹³

- (12) *O gelin gel-sın ev-in-e mümkün-i yok.*
 DEM bride come-VOL.3 house-POSS.3-DAT possible-POSS.3 absent
 ‘It [was] impossible for the bride to come his house.’
 (MK/Tetovo, Erdem et al. 2024: 230)

A different kind of linking strategy is observed in the next example, where the verbal noun in *-mAK* is followed by the postposition *için*. This strategy resembles the infinitive strategy, as *-mAK için*, similarly to *-mAYA*, can be used to form purpose clauses in ST and other varieties. It also reflects the universal affinity of purposive and infinitive.

- (13) *Yürü-r yürü-r, bi dere-ye cel-ır.*
 march-AOR march-AOR one river-DAT come-AOR
On-i apuş-mak için mümçün yok.
 DEM-ACC transgress-VN for possible absent
 ‘He walks and walks and comes to a river. It is impossible to cross it.’
 (RKS/Prizren, Hafiz 1985: 189)

More non-finite linking strategies are found in the Adakale data, including the plain verbal noun in *-mAK* and the dative-marked form in *-mAGA*. The latter, often realized in the contracted form *-mĀ*, is illustrated in the following ERT example from Bulgaria.

- (14) *Eh tä bereket vā-sın çocūm=län, kız-lar-ım=lan*
 PTCL PTCL luckily son.POSS.1.SG=with daughter-PL-POSS.1.SG=with
geçin-eme-mā mümkün yok. Torun-lar-ım=la
 get.along-NEG.POT-VN.DAT possible absent grandchild-PL-POSS.1.SG=with
geçin-ēm çok şükür allah-ım-a, bu gün-ä.
 get.along-PROG.1.SG many thank god-POSS.1.SG-DAT DEM day-DAT
 ‘Well, look, luckily, it is impossible not to [be able to] get along with my sons and daughters. Thank God I get along with my grandchildren until now.’
 (BG/Silistra, Karaşinik 2011: 181)

This example is particularly intricate as it combines two negated expressions of possibility, the one in *-(y)AmA-* and *mümkün yok*. The exact pragmatic force of this complex construction is not entirely clear but the context suggests that semantically at least the combination – negated possibility within the scope of another negated possibility – conveys a meaning akin to ‘it is absolutely possible’ or ‘it is not impossible’. Given that this is a singular attested instance, its relevance to the overall language system is unclear. This recorded instance may represent an idiolectal feature or even a slip of the tongue.

In contrast, at least some of the aforementioned constructions can be traced back to pre-modern Turkish varieties. The Old Ottoman *Ferec ba’d eş-şidde*, a text from the 15th century or potentially earlier, contains two types discussed above: the **-mAK mumkin (değil)* construction (example [15]) and the subjunctive-type construction (example [16]). The Modern Standard Turkish type *-mAsI mümkün* can be identified in the Middle Ottoman *Tārīh-i Peçevī*

¹² In ST the adjective *mümkün* is usually negated with *mümkün değil* as in example (10). There is an alternative nominal construction *imkan-ı yok*, literally ‘its possibility is absent’.

¹³ *Mümkün dīl mi ki kendi kelligini de geçirtir-e-sın güzel bir delikanlı ol-a-sın?* ‘Isn’t it possible that you have your baldness removed and become a handsome young man?’ (Kúnos 1907: 133).

(example [17]), a chronicle by İbrāhīm Peçevî (1574–ca. 1649), an Ottoman from Pécs in Hungary, completed between 1642 and 1649 (Schaendlinger 1972: 186). The *-mAsI mümkün* type itself is unlikely to be much older, because the “short” verbal noun in *-mA* did not emerge significantly earlier than 17th century, as shown by Brendemoen (2014).

- (15) *Yalan=u gerceg-ünüz hod mühre-sin-de ma'lüm=dur*
 lie=and truth-POSS.2.PL self marble-POSS.3-LOC known=COP
kimse bun-dan söz yaşur-mak mümkin degül.
 somebody DEM-ABL word conceal-VN possible NEG
 ‘Whether you speak a lie or the truth is understood in her marble; no-one can conceal the truth from it.’

(Old Ottoman/FBŞ 183a13–14, Hazai and Tietze 2006: 506)

- (16) *Cihān-ı gez-er=se-η mümkin degül=dür*
 world-ACC tour-AOR=COND-2.SG possible NEG=COP
nazır-in-i bul-a-sın.
 match-POSS.3-ACC find-OPT-2.SG

‘If you travel around the world it is not possible that you will find its match.’

(Old Ottoman/FBŞ 136b20–21, Hazai and Tietze 2006: 406)

- (17) *Bir vakt-i mu'ayyen-de gel-me-si mümkün ol-an-lar*
 one particular.time-LOC come-VN-POSS.3 possible be/come-PTCP-PL

‘Those for whom it is possible to come at a particular time’

(Middle Ottoman/TP 91b5, Özbal 2005: 41)

An interesting type of impossibility markers is based on the copula element *yok* ‘absent’. In most of the instances, this type appears to encode negated participant-external possibility, including negative deontic possibility. In the latter domain, there is an overlap with deontic necessity, resulting from the translatability of negated permission (i.e., negative deontic possibility, $\neg \diamond p$) into a prohibition (i.e., the obligation not to do something, $\square \neg p$).¹⁴ This affinity is reflected also in the interpretation of such constructions, which may oscillate between ‘cannot’ and ‘must not’, depending on the context.

A variant also available in ST combines the verbal noun in *-mAK* with *yok*. As expected, this variant occurs frequently in ERT, but it is also found in WRT, as illustrated in example (18) from Kumanovo.

- (18) *Pope, darıl-mag yok=tur.*
 pope be.offended-VN absent=COP

‘Pope, you should not be offended.’

(MK/Kumanovo, Eckmann 1962: 128)

[cf. ERT: *Türkçe bilir, yannız, kızıl çin tarafınna konuş-mak yok* ‘They know Turkish, but it is forbidden to talk to the Red Chinese’ (TR/Tekirdağ/Naip Köyü, Tosun 2003: 335)]

The verbal noun in *-mAK* seems to be interchangeable with *-mA* both in Eastern and Western varieties. However, since the available data do not contain instances with *-mAGA* and

¹⁴ For more details, see van der Auwera and Plungian (1997: 99–100).

-mā combined with *yok*, it is not totally clear whether *-mA* in this construction represents the **-mAGA* type or the unmarked short verbal noun in *-mA*.¹⁵

- (19) *Salde yok=tor darıl-ma.*
 only absent=COP be.offended-MA
 ‘You only should not be offended.’
 (RKS/Prizren, Hafız 1985: 231)
 [*Büle sāya sola gez-me yok* ‘It is not possible to simply walk around here and there’
 (TR/Edirne/Lalapaşa, Kalay 1998: 219)]

In WRT, *yok* as an auxiliary can also combine with finite mood forms (the subjunctive complementation strategy). This construction is especially common in North Macedonia.

- (20) *Mare adam, açan gid-eceys anam-da, ev-de,*
 PTCL man when go-FUT.1.PL mother-LOC house-LOC
yok yēsin sen boyle bitevi el=le.
 absent eat.OPT.2.SG you so continuously hand=with
 ‘Hey, man, when we go to my mother’s house, you must not/cannot eat with your fingers like this all the time.’
 (MK/Resen, Ahmed 2001: 141)

As with the case of *bil-*, it is reasonable to assume that formally similar constructions containing a question word form the basis from which the construction exemplified in (20) has developed:

- (21) *Dövlet bil-mes, şaşır-ır o, yok*
 ruler know-NEG.AOR.3 be.confused-AOR.3 DEM absent
ne yap-sın.
 what do-VOL.3
 ‘The ruler does not know what is going on, there is nothing he can do.’
 (MK/Resen, Ahmed 2001: 134)

- (22) *Em işte, orda yok iç kimse, yok kim gör-sün*
 and PTCL there absent at.all somebody absent who see-VOL.3
orda biz-i.
 there we-ACC
 ‘And there is nobody, there is no-one who could see us there.’
 (MK/Resen, Ahmed 2001: 170)

- (23) *Açan ol-ur akşam, yok nerde kal-sın.*
 when be/come-AOR.3 evening absent where stay-VOL.3
 ‘When it becomes evening, there is no place for her to stay.’
 (MK/Resen, Ahmed 2001: 157)

¹⁵ In terms of syntax, in the construction *darılmak yok*, *darılmak* is the subject and *yok* is the predicate, so one might ask how the use of **-mAGA* in this slot can be justified. However, the same can be said of the construction *-mAK mümkün* (cf. ex. [15]), where we have alternative constructions with **-mAGA* nonetheless, see ex. (14). It seems possible that at least in the east, where the verbal noun *-mA* is highly frequent, *-mA* might partly be the unmarked form of the short verbal noun in *-mA*. In contrast, in the west, where this verbal noun is scarce, it seems more likely that this segment belongs to the **-mAGA* type and is inserted into this auxiliary construction by analogy with other auxiliary constructions containing **-mAGA*.

- (24) *O sokak-lar... yog idi ner-den geç-e-sın araba=yle.*
 DEM street-PL absent PST where-ABL pass-OPT-2.SG car=with
Dar sokak-lar idi-ler, çamur-lar, aman aman...
 narrow street-PL PST-PL mud-PL for.goodness.sake
 ‘Those roads... There was nowhere you could pass with a car. The roads were narrow,
 and there was mud, my goodness!’
 (MK/Resen, Ahmed 2001: 190)

All these constructions have parallels in Macedonian: (20') *Nema da jadeš so raka* ‘You shall not eat with the hand’,¹⁶ (21') *Nema što da pravi* ‘There is nothing she can do’, (22') *Nema koj da nè vidi* ‘There is no-one who can see us’, (23') *Taa nema kade da prestojuva* ‘She has no place to stay’, (24') *Nema kade da pomineš so kolata* ‘There is nowhere to pass with the car’. Thus, it seems reasonable to assume that the structures underlying examples (20)–(24) result from code-copying from Macedonian.¹⁷ Regarding (20), the existing autochthonous structure *-mAK yok* may have played a role in the formation of the item through selective copying of combinational features from Macedonian onto a Turkish structure.

As a final point in the documentation and discussion of expressions of possibility in Balkan Turkish, constructions using a form of the verb *ol-* ‘to become, to be’ as an auxiliary deserve mention. The form commonly called aorist in Turkish studies of this verb – *olur* – is a widespread independent expression of consent in Turkish which can be analysed as a lexicalized interjection denoting ‘alright, okay’. The construction relevant for the present discussion predominantly occurs in interrogative sentences. Semantically, this type covers participant-external possibility, including deontic possibility (permission). In Eastern varieties, the linking segment is *-mAK* (example [25]), while in Western varieties the non-finite complementation strategy involving **-mAGA* as in example (26) coexists with a finite complementation strategy (examples [27] and [28]). Example (28) differs from (27) in two respects: it includes the complementizer particle *ki* between the matrix verb and the complement clause, and the predicate of the complement clause is negated.

- (25) *Hep otur-mak ol-ur mu beyāv?*
 always sit-VN be/come-AOR Q PTCL
 ‘Is it possible to always sit around, eh?’
 (TR/Edirne/Merkez, Kalay 1998: 200)
- (26) *Deli mi-sın, akilli mi-sın? Devlet-in kız-ın-i*
 crazy Q-2.SG clever Q-2.SG rich-GEN daughter-POSS.3-ACC
ol-or mi ara-ma?
 be/come-AOR Q ask-VN.DAT
 ‘Are you in your right mind? Can one ask for the hand of the rich man’s daughter?’
 (RKS/Prizren, Hafız 1985: 213)

¹⁶ Also negative future: ‘You will not eat with the hand’ (i.e., with fingers), cf. Koneski (1967: 487).

¹⁷ The constructions with question word (i.e. constructions similar to [21]–[24], but not [20]) are also well attested in Gagauz, where possibility is mainly expressed by *var/yok* with one of the question words *nasıl* and *nicā* ‘how’ and either *-mAA* (< *-mAGA*) or a subjunctive mood form (Menz 1999: 59). Other question words can be used in the same construction types for procedural and practical knowledge (Menz 1999: 63–66). Hence, we may assume that such constructions also exist in other Eastern Rumelian Turkish dialects, although they are not attested in the material considered for this study.

- (27) *Ol-or mi bu cece bu ineg-i siz-de koyu?*
 be/come-AOR Q DEM night DEM cow-ACC you.PL-LOC put.VOL.1.SG
 ‘Can I leave this cow with you tonight?’
 (RKS/Prizren, Hafız 1985: 187)
- (28) *Hiç ol-ur mi ki dünyā-de bun-ın yara-lar-ın-a*
 ever be/come-AOR Q COMP world-LOC DEM-GEN wound-PL-POSS.3-DAT
çāre bul-un-ma-sın?
 cure find-PASS-NEG-VOL.3
 ‘Is it possible that no cure for his wounds is found in this world?’
 (Adakale, Kúnos 1907: 74)

3 Necessity

The most frequent expressions of necessity in Balkan Turkish are constructions based on the Arabic adjectives *lāzim* ‘necessary’ and *mecbur* ‘forced’. The former occurs significantly more often than the latter, which is not considered in detail here. Notably, the nominal auxiliary *gerek* ‘necessity’, which is frequently used in ST alongside *lazım* and is well-attested in Old and Middle Oghuz, rarely occurs in the texts under investigation, even in Eastern Rumelian varieties. The same applies to the necessitative in *-mAlI*, a mood form directly attached to verb stems and highly frequent in ST. Several authors remarked on the absence (or almost absence) of this item in the dialects they investigated, including İgci (2010: 68) for Vushtrria (Kosovo), Karasinik (2011: 136–137) for Silistra (Bulgaria), and Mustafa-Rashidova (2024: 156) for the Central Rhodopes (Bulgaria).

In ST, *lazım* (alternative writing: *lāzim*) is constructed similarly to *mümkün*, i.e., with the verbal noun *-mAK* expressing impersonal necessity (example [29], cf. [10], and with the verbal noun *-mA* followed by a possessive suffix when the projected performer of the action is indicated (example [30], cf. [11]).

- (29) *Bacak-lar-ı uygun ol-ma-yan-a mini eteğ-i*
 leg-PL-POSS suitable be/come-NEG-PTCP-DAT miniskirt-ACC
yasakla-mak lazım.
 forbid-VN necessary
 ‘One should ban miniskirts for those with unfit legs.’
 (ST, Pamuk 2008: 101)
- (30) *Konuş-ma-nız lazım.*
 talk-VN-POSS.2.PL necessary
 ‘You must talk.’
 (ST, Pamuk 2008: 204)

Not surprisingly, non-finite linking strategies with verbal nouns in the *casus rectus* are frequent in Eastern Rumelian dialects, but they do also occur in the west, as the following examples from North Macedonia illustrate.

- (31) *O şey-ler-e lazım austos-ta gel-mek.*
 DEM thing-PL-DAT necessary august-LOC come-VN
 ‘One has to come in August for these things.’
 (MK/Debar, Erdem et al. 2024: 359)
 [cf. ERT: *Allah ne kadā verisā o kadā yaşa-mak lāzım* ‘One must live as long as God wants you to’ (BG/Silistra, Karaşinik 2011: 177)]
- (32) *On-un için aç-lar-ı kes-me-n lazım.*
 DEM-GEN for tree-PL-ACC cut-VN-POSS.2.SG necessary
 ‘Therefore you have to cut down the trees.’
 (MK/Budakovo, Alievska 2003: 130)
 [cf. ERT: *Cenāzā günündā az da olsa yemek yapıl-ma-sı lāzım* ‘On the day of the funeral at least a bit food must be prepared’ (BG/Silistra, Karaşinik 2011: 187)]

In WRT, however, *lazım* much more frequently combines with finite mood forms. This structure is similar to the one exemplified by (12) with *mümkün*. An example from Prizren (Kosovo) is given in (33). The pattern is common throughout Kosovo, North Macedonia, Western Bulgaria (including Vidin), and historically also documented in the Adakale texts (where it is represented roughly on par with patterns known from Modern Standard Turkish). It is also not uncommon in ERT and very widespread in Gagauz (Menz 1999: 54–58). In addition, the **-mAGA* type is found in WRT, particularly in Kosovo (example [34]). This non-finite item is especially suitable for impersonal constructions, as *-mA* (< *-mAGA*) does not accept further suffixation.¹⁸ Note the **-mAGA* type is also found in Gagauz (Menz 1999: 55); therefore, its presence in ERT dialects must be considered, even though it did not appear in the texts consulted for this study.

- (33) *Ne iste-yeceg-ımız-i lazım düşün-alım.*
 what want-PRO.VN-POSS.1.PL-ACC necessary think-VOL.1.PL
 ‘We have to think what we shall ask for.’
 (RKS/Prizren, Hafız 1985: 215)
 [*Gelin güvā, güvā lāzım o çāşırı giy-sin, gelin de bindallıyı giy-sin lāzım* ‘Bride and groom, the groom has to wear that *çāşır*, and the bride has to wear the *bindallı*’ (BG/Silistra, Karaşinik 2011: 171)]
- (34) *Dert dane lazım=dır cütür-ma o yer-e*
 four piece necessary=COP bring-VN.DAT DEM place-DAT
da o kardaş-i çik-sın ora-dan.
 and DEM sibling-POSS.3 come.out-VOL.3 there-ABL
 ‘It is necessary to bring there four [rams] so that his brother comes forth.’
 (RKS/Prizren, Hafız 1985: 225)

In 19th century WRT as documented for Adakale, the complementizer particle *ki* could intervene between the matrix segment and the subjunctive form (as with *mümkün*).¹⁹

Another matrix segment frequently used in expressions of necessity is *meçbur* ‘forced’, an adjective of Arabic origin. It primarily governs the dative and occurs in **-mAGA* type constructions, but also appears with the subjunctive and in various other constructions. In

¹⁸ It cannot be emphasized enough that this statement applies to the item *-mA* ~ *-mĀ* ~ *-mAGA*. As already mentioned, there is another plain verbal noun in *-mA*, experiencing an upsurge in the 17th century and highly frequent in ST. It readily accepts inflectional suffixes such as possessive markers and case markers.

¹⁹ *Lāzım ki şimdi sen onnarı dāvet ed-e-sin* ‘It is necessary that you invite them’ (Kúnos 1907: 43–44).

addition, there is a significant amount of miscellaneous expressions of necessity, some of which resemble items from other Turkish varieties such as Standard Turkish or Anatolian dialects, including constructions such as *-mAGA hacet yok* (Adakale, Kúnos 1907: 148), *-mAGA mutaç* (MK/Struga, Ahmed 2004: 304), *-mAsA olmaz* (Adakale, Kúnos 1907: 202), *-mAK mecburiyetinde kal-* (RKS/Vushtrria, Hafiz 1985: 242), *-mAGIn luzumu yok* (Adakale, Kúnos 1907: 207), etc. These examples will not be discussed in detail here; however, two more types (with their variants) deserve mention in this study due to their potential significance for Balkan linguistics. Both types contain the auxiliary segment *var* ‘present, existent’.

The first type consists of *var* with the subjunctive as in the following example from Kosovo:

- (35) *Biz meclis-te karar al-misık. Çoban*
 we assembly-LOC decide-EVID.PST.1.PL shepherd
var as-ıl-sın!
 present hang-PASS-VOL.3
 ‘We have decided in the assembly. The shepherd must be hung!’
 (RKS/Mamusha, Hafiz 1985: 251)

This example seems to be paradigmatically related by opposite polarity to the construction <yok+SBJV> in example (20) from Resen. As previously mentioned, there is a logical relation between possibility and necessity in combination with negation, which also manifests linguistically. While <yok+SBJV> encodes impossibility and prohibition, <var+SBJV> in example (35) seems to convey necessity. Moreover, Macedonian has a similar pattern *ima da* (‘be present, have’ + subjunctive), which expresses, among others, obligation (Mitkovska and Bužarovska 2012). A similar construction *ka për të* (active) and *ka për t’u* (non-active) with subjunctive exists in Albanian (Buchholz and Fiedler 1987: 85–86). Given these structural similarities, it seems highly likely that the WRT construction has developed under contact influence.

Another construction involves *var* ‘present, existent’ and a prospective verbal noun. This type, which is very old²⁰ and attested in many different Turkic languages, has been described by Rentzsch (2015: 154–156) as a volitive marker. Some Balkan Turkic examples of this construction can, in fact, be interpreted in terms of a desire, but others exhibit clear readings of participant-internal necessity (urge, need). In this construction, *var* is interchangeable with the verb *gel-* ‘to come’,²¹ often in a preterite form (simple past in *-DI* or evidential past in *-mİş*), without any noticeable semantic difference. The negative variant of the type <PRO.VN+var> is <PRO.VN+yok>. Consider the following two examples from ST (cf. Rentzsch 2015: 155–156), before turning to the Balkan Turkish examples.

- (36) *Gör-d-ün mü derviş efendi,*
 see-PRET-2.SG Q dervish master
döv-üş-esi-m yok=tu ama döv-üş-t-üm.
 beat-COOP-PRO.VN-1.SG absent=PST but beat-COOP-PRET-1.SG
 ‘Did you see that, Mister dervish, I didn’t want to fight but I fought.’
 (ST, Şafak 2009: 52)

²⁰ The oldest known occurrence is from the Ongi Inscription (ca. 732–734 AD) in present-day Mongolia. The Old Turkic specimens of this type are still semantically diffuse between necessity, volition and possibility (Rentzsch 2015: 51–52).

²¹ The oldest attested occurrence of the variant with *kel-* ‘to come’ is even older; it is found in the Tonyukuk Inscription (726 AD).

- (37) *A ne gör-eceğ-im gel-di!*
 PTCL what see-PRO.VN-1.SG come-PRET
 ‘Oh, how much I would like to see them!’
 (ST, Tanpınar 1973: 148)

In the above examples, we observe two different prospective (“future”) verbal nouns as linking segments, the archaic *-(y)AsI* and the widespread, fully productive *-(y)AcAK*. Both are combined with possessive markers to indicate the “wisher”; and both occur with two different auxiliary segments, *var/yok* ‘present/absent’ and *gel-* ‘to come’. Both parameters – linking segment and auxiliary segment – are interchangeable without altering the semantic interpretation, which produces a clear volitive reading.

Turning to the data from Balkan Turkish, it is notable that only the verbal noun **(y)AcAK* occurs as a linking segment in this type of construction; the verbal noun in *-(y)AsI* is not attested.²² It remains unclear whether this absence is a coincidence or it reflects a broader pattern in the distribution of the morpheme *-(y)AsI* across the Turkish dialects. This question necessitates further investigation. The auxiliary segment *var* ‘present’ is found in examples (38) and (39). Example (38) evokes a volitive reading, while the reading of example (39) is necessitative.

- (38) *Adam benim şarko inek et-i*
 man my spotted cow meat-POSS.3
yeci-m var, yarın kasap-lar-ı
 eat.PRO.VN-POSS.1.SG present tomorrow butcher-PL-ACC
çar-tır bizim bu şarko ini-y kes-tir.
 call-CAUS.IMP.SG our DEM spotted cow-ACC cut-CAUS.IMP.SG
 ‘Man, I want to eat the meat of a spotted cow, summon the butchers tomorrow and have them slaughter this spotted cow of us.’
 (MK/Kanatlarci, Alievska 2003: 151)

- (39) *Otur-a otur-a uyku-sı gel-ir,*
 sit-CVB sit-CVB sleep-POSS.3 come-AOR
çiş ed-eceg-i de var=miş.
 pee-PRO.VN-POSS.3 too present=EVID
 ‘While he is sitting and sitting, he becomes sleepy, and he has to pee.’
 (Adakale, Kúnos 1907: 181)

The remaining examples contain the auxiliary *gel-* ‘to come’. There is some morphological variation: in example (41) from Resen (North Macedonia) the third person possessive marker appears in its postvocalic variant *-sI*, indicating that the suffix **(y)AcAK* has undergone phonetic reduction to *-(y)AcA*.²³ In contrast, the third person variant in *-eciy* in the Budakovo dialect (examples [40] and [42]) builds on underlying **-ecegi* and represents a more conservative formation.

While examples (40)–(42), drawn from folk tales, use the evidential past in *-miş* as default TAM form of this discourse type, examples (43) and (44) represent direct speech. In spite of past tense marking, they express synchronic modalities, which aligns with usage in ST. The

²² Generally, it can be stated that the verbal noun in *-(y)AsI* is not particularly frequent and productive in Turkish, except in curses and benedictions, and specifically in the construction *-(y)AsI var/yok/gel-*.

²³ In the future, contraction forms such as *içecem* (1.SG), *içecen* (2.SG), *içecek* (1.PL) in the finite paradigm and *içecem* (1.SG), *içecen* (2.SG), *içeceyi* (3.SG) in the non-finite paradigm can be found even in casual spoken Standard Turkish, but the form *içecesi* for 3.SG non-finite seems rather unusual.

semantic development underlying this idiomatic expression can be conceptualized as follows: *su iç-eceğ-im gel-di* (lit.) ‘my prospective water-drinking has come’ ≈ *su iç-eceğ-im var* (lit.) ‘my prospective water-drinking is present’ → ‘I want to drink water’.

Examples (42) and (43) demonstrate that the order of the segments is variable in this type of construction as well. Examples (42) and (44) mention the subject overtly; the possessive suffix in the predicate requires the subject to appear in the genitive (as in ST). In terms of interpretation, examples (40)–(42) are understood as expressing a wish, whereas examples (43) and (44) convey necessity.

- (40) *Kız baba-sın-ı çok özle-miş,*
 girl father-POSS.3 much miss-EVID.PST
gör-eciy gel-miş.
 see-PRO.VN.POSS.3 come-EVID.PST
 ‘The girl missed her dad a lot and wanted to see him.’
 (MK/Budakovo, Alievska 2003: 150)

- (41) *Su iç-ece-si gel-miş.*
 water drink-PRO.VN.POSS.3 come-EVID.PST
 ‘He was thirsting for water.’ (= ‘He wanted to drink water.’)
 (MK/Resen, Ahmed 2001: 132)

- (42) *Asan-in gel-miş su iç-eciy.*
 NP-GEN come-EVID.PST water drink-PRO.VN.POSS.3
 ‘Hasan wanted to drink water.’
 (MK/Budakovo, Alievska 2003: 147)

- (43) *Abdes boz-acay-ım g’âl-di.*
 abolutions anul-PRO.VN-POSS.1.SG come-PRET
 ‘I have to go to the toilet.’
 (BG/Vidin, Németh 1965: 156)

- (44) *Kız-ım, benim ül-eceğ-im gel-di.*
 daughter-POSS.1.SG my die-PRO.VN-POSS.1.SG come-PRET
 ‘My daughter, my time has come to die.’ (= ‘I must die.’)
 (Adakale, Kúnos 1907: 181)

Some general observations about the examples (38)–(44) can be made: In the texts consulted for this study, only positive (i.e., unnegated) examples are attested. As noted previously, only **(y)AcAK* occurs as the linking segment; *-(y)AsI* does not appear. Semantically and pragmatically, the examples involving events such as eating, drinking, and meeting people evoke a volitive reading, i.e., a reading based on desire. In contrast, examples involving actions as urinating, defecating, and dying produce a necessitative reading, i.e., a reading based on need. A plausible hypothesis (which will require further investigation) is that controllable actions trigger volitive readings while uncontrollable or less controllable actions trigger a necessitative reading. Thus, the interpretation seems to be related to the degree of control encoded in the state of affairs.

It seems that volitive modality and necessity are not semantically coded by the construction itself, but rather emerge as pragmatic interpretations of a broader, more diffuse semantic base. These context-dependent interpretations are influenced particularly by the degree of control associated with the state of affairs. Historical data show that in Old Turkic, this type of

construction was semantically more “fuzzy” than in most modern Turkic languages, where the volitive meaning has become dominant. This raises the question whether the necessitative usages of this item in Balkan Turkish reflect a retention of an earlier situation compared to ST, or represent a semantic shift influenced by the semantics of the item <var+SBJV> (see example [35]), which itself most probably developed under the influence of contact with Balkan languages. This issue remains an area for future investigation.

4 Conclusion

The expressions of possibility and necessity investigated in this part of the study constitute a subset of a broader system of event modality markers. This system also includes additional, less frequent expressions of possibility and necessity, as well as expressions of volitive modality, which will be explored in Part 2 of the study.

The items focussed on in this study share the morphosyntactic property of including a component of complementation, in which a matrix segment takes an SoA into its scope. The complementation in Balkan Turkish is achieved by a linking segment, which may be finite or non-finite, giving rise to various complementation strategies.

The matrix segments considered in this study are either verbal or nominal in nature. The combinational features of the underlying lexical items vary, ranging from either casus indefinitus (nominative) or accusative as seen with *bil-* ‘to know’, to nominative as in the case of *mümkün* ‘possible’, *lazım* ‘necessary’, *yok* ‘absent’ and *olur mu* ‘is it possible’, as they originally occur in subject-predicate constructions. In addition, some items may also require other cases such as the dative with *mecbur* ‘forced’ and *muhtaç* ‘dependent’.

In the Balkan Turkish dialects we observe a tendency for complementation patterns to shift into two directions: constructions involving a non-factual verbal noun in the dative, yielding an infinitive-type pattern; and constructions using a finite mood form resulting in a subjunctive-type pattern. Both strategies are attested with *bil-*, *mümkün*, *lazım*, *yok*, *olur mu*, and also *mecbur*.

Regarding the distribution of these strategies, both constructions are attested in eastern and western varieties of Balkan Turkish, though there is a preference for the subjunctive strategy in WRT. In contrast, the dialects in North Macedonia employ non-finite complementation strategies far less frequently than other Balkan Turkish varieties, a phenomenon consistent with the so-called infinitive loss, which is particularly pronounced in Macedonian, the dominant contact language.

The construction *var/yok* ‘present/absent’ plus subjunctive (examples [20] and [35]) can be argued to have developed under the contact influence of both Macedonian and Albanian.

In contrast, a different construction involving *var* ‘present’ or *gel-* ‘to come’ and a future verbal noun, appears to have older roots in the Turkic languages. Semantically, this construction occupies a space between participant-internal necessity and volitive modality.

Acknowledgement

The initial research for this study was conducted during a post-doctoral fellowship at FON University Skopje between April 2016 and March 2017, which was supported by the Deutsche Forschungsgemeinschaft (DFG, RE 2727/4-1).

Abbreviations

AOR	aorist
BG	Bulgaria
ERT	Eastern Rumelian Turkish
EVID	evidential
FBŞ	Ferec ba'd eş-şidde
MK	Republic of North Macedonia
OPT	optative
POT	potential
PRET	preterite
PRO	prospective
PTCL	particle
RKS	Kosovo
SoA	state of affairs
ST	Standard Turkish
TP	Tārīḥ-i Pečevī
TR	Turkey
VN	verbal noun
VOL	voluntative
WRT	Western Rumelian Turkish

Bibliography

- Achard, M. (2007). Complementation. In D. Geeraerts and H. Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 782–802. Oxford: Oxford University Press.
- Ahmed, O. (2001). *Fonetički i leksički karakteristiki na turskiot govor vo Resen i Resensko*. MA thesis, UKIM, Skopje.
- Ahmed, O. (2004). *Morfosintaksa na turskite govori od ohridsko-prespanskiot region*. PhD dissertation, UKIM, Skopje.
- Alievska, M. (2003). *Ženata kako lik vo turskite narodni prikazni od prilepsko i bitolsko*. MA thesis, UKIM, Skopje.
- Auwera, van der, J. and Plungian, V. A. (1998). Modality's semantic map. *Linguistic Typology*, 2: 79–124.
- Brendemoen, B. (2013). The infinitive in -me/-ma in the Balkan dialects. *Turkic languages*, 17: 31–37.
- Brendemoen, B. (2014). Some remarks on the infinitive in -mA in 17th century Ottoman Turkish. In N. Demir, B. Karakoç and A. Menz (eds.), *Turcology and Linguistics. Éva Ágnes Csató Festschrift*, 103–113. Ankara: Hacettepe Üniversitesi Yayınları.
- Buchholz, O. and Fiedler, W. (1987). *Albanische Grammatik*. Leipzig: VEB Verlag Enzyklopädie.
- Dixon, R.M.W. (2006). Complement clauses and complementation strategies in typological perspective. In R.M.W. Dixon, and A. Y. Aikhenvald (eds.), *Complementation. A cross-linguistic typology*, 1–48. Oxford: Oxford University Press.
- Doerfer, G. (1990). Die Stellung des Osmanischen im Kreise des Oghusischen und seine Vorgeschichte. In G. Hazai (ed.), *Handbuch der türkischen Sprachwissenschaft*. Teil 1, 13–34. Budapest: Akadémiai Kiadó.
- Eckmann, J. (1962). Kumanova (Makedonya) Türk Ağzı. In J. Eckmann, A. S. Levend, and M. Mansuroğlu (eds.), *Németh Armağanı*. (Türk Dil Kurumu Yayınları 191), 111–144. Ankara: Türk Tarih Kurumu Basımevi.
- Erdem, M. D., İsmaili, B., Beyoğlu, B. and Bölük, R. (2024). *Kuzey Makedonya Türk Ağzıları. Giriş-İnceleme-Metinler*. Elazığ: Asos Yayınları.
- Friedman, V. A. (2003). *Turkish in Macedonia and beyond. Studies in contact, typology and other phenomena in the Balkans and the Caucasus*. (Turcologica 52.) Wiesbaden: Harrassowitz.

- Friedman, V. A. (2006). West Rumelian Turkish in Macedonia and adjacent areas. In: H. Boeschoten and L. Johanson (eds.). *Turkic languages in contact*. (Turcologica 61), 27–45. Wiesbaden: Harrassowitz.
- Göksel, A. and Kerslake, C. (2005). *Turkish. A comprehensive grammar*. London /New York: Routledge.
- Hafiz, N. (1985). *Kosova Türk Halk Edebiyatı Metinleri*. Priştine: Kosova Üniversitesi Priştine Felsefe Fakültesi.
- Haspelmath, M. (1989). From purposive to infinitive – a universal path of grammaticalization. *Folia Linguistica Historica*, 10/1–2: 287–310.
- Hazai, G. (1973). *Das Osmanisch-Türkische im XVII. Jahrhundert. Untersuchungen an den Transkriptionstexten von Jakab Nagy de Harsány*. Budapest: Akadémiai Kiadó.
- Hazai, G. and Tietze, A. (eds.) (2006). *Ferec ba 'd eş-şidde. "Freud nach Leid". Ein frühosmanisches Geschichtenbuch*. (Studien zur Sprache, Geschichte und Kultur der Türkvölker 5.1–5.2.) Berlin: Klaus Schwarz Verlag.
- İgci, A. (2010). *Viçitrin-Kosova Türk Ağzı*. MA thesis. Ege Üniversitesi, İzmir.
- Joseph, B. D. (1983). *The synchrony and diachrony of the Balkan infinitive. A study in areal, general, and historical linguistics*. Cambridge: Cambridge University Press.
- Kakuk, S. (1972). Le dialecte Turc d'Ohrid en Macédonie. *Acta Orientalia Academiae Scientiarum Hungaricae*, 26: 227–282.
- Kalay, E. (1998). *Edirne İli Ağzları. İnceleme-Metin*. (TDK Yayınları 694). Ankara: Türk Dil Kurumu.
- Karaşinik, B. (2011). *Silistre (Silistra) İli ve Yöresi Ağzları*. MA thesis. Trakya Üniversitesi, Edirne.
- Koneski, B. (1967). *Grammatika na makedonskiot literaturni jazik*. Skopje: Kultura.
- Kúnos, I. (1907). *Materialien zur Kenntnis des rumelischen Türkisch. I: Türkische Volksmärchen aus Adakale, gesammelt, in Transkription herausgegeben und mit Einleitung versehen*. Leipzig /New York: Rudolf Haupt.
- Lewis, G. L. (1967). *Turkish grammar*. Oxford /New York: Oxford University Press.
- Menz, A. (1999). *Gagausische Syntax. Eine Studie zum kontaktinduzierten Sprachwandel*. (Turcologica 41). Wiesbaden: Harrassowitz.
- Mitkovska, L., Bužarovska, E. (2012). Tipovi modalnost vo nemenlivite ima da-konstrukcii. *Makedonski jazik*, LXIII: 83–103.
- Mustafa-Rashidova, S. (2024). *Bulgaristan'daki Orta Rodop Türk Ağzlarının Fonetik, Morfolojik ve Sentaktik Özellikleri*. PhD dissertation. İstanbul Üniversitesi, İstanbul.
- Németh, J. (1956). *Zur Einteilung der türkischen Mundarten Bulgariens*. Sofia: Bulgarische Akademie der Wissenschaften.
- Németh, J. (1965). *Die Türken von Vidin. Sprache, Folklore, Religion*. Budapest: Akadémiai Kiadó.
- Noonan, M. (2007). Complementation. In T. Shopen (ed.). *Language typology and syntactic description*. Vol. 2: *Complex constructions*, 52–150. Cambridge: Cambridge University Press.
- Özbal, M. (2005). *Peçevî Tarihi. 80b–114a metin, edisyon kritik, dizin, özel adlar sözlüğü*. MA thesis. Marmara Üniversitesi, İstanbul.
- Palmer, F. R. (2001). *Mood and modality*. (Second edition.) Cambridge: Cambridge University Press.
- Pamuk, O. (1990 [2004]). *Kara Kitap*. İstanbul: İletişim.
- Pamuk O. (2008 [2018]). *Masumiyet Müzesi*. İstanbul: Yapı Kredi Yayınları.
- Rentzsch, J. (2014). Zur Rolle der Analogie in der Grammatikalisierung türkischer Modalauxiliare. In N. Demir, B. Karakoç and A. Menz (eds.). *Turcology and Linguistics. Éva Ágnes Csató Festschrift*, 357–374. Ankara: Hacettepe Üniversitesi Yayınları.
- Rentzsch, J. (2015). *Modality in the Turkic languages. Form and meaning from a historical and comparative perspective*. [Studien zur Sprache, Geschichte und Kultur der Türkvölker 18]. Berlin: Klaus Schwarz Verlag.
- Rentzsch, J. (2018). SoA-type complement clauses in the Tārīḥ-i Peçevî. In É. Á. Csató, J. Parslow, E. Türker and E. Wigen (eds.). *Building bridges to Turkish. Essays in honour of Berndt Brendemoen*. (Turcologica 116), 265–284. Wiesbaden: Harrassowitz.
- Rentzsch, J., Mitkovska, L. and Nedelkoska, G. (2020). Complementation strategies with the verb 'to know' in the Turkish dialect of Ohrid compared to Standard Turkish and Macedonian. *Turkic Languages*, 24 (1): 78–109.

- Römer, C. (2012). On some peculiar datives in colloquial Middle Ottoman texts. In T. Kahl, M. Metzeltin and H. Schaller (eds.). *Balkanismen heute – Balkanisms today – Balkanizmy segodnja*. (Balkanologie. Beiträge zur Sprach- und Kulturwissenschaft 3), 185–194. Münster: LIT.
- Şafak, E. (2009). *Aşk*. İstanbul: Doğan Kitap.
- Schaendlinger, A. C. (1972). Die Probleme der Redaktionen der Chronik des Ibrāhīm Pečevī (Peçuyly). *Wiener Zeitschrift für die Kunde des Morgenlandes*, 63/64: 176–186.
- Tanpınar, A. H. (1973 [1999]). *Sahnenin Dışındakiler*. İstanbul: Dergâh Yayınları.
- Tosun, İ. (2003). *Tekirdağ Merkez İlçe ve Köyleri Ağızları*. MA thesis. Trakya Üniversitesi, Edirne.

АКЦЕНТОТ НА ЗБОРОТ КАКО БАЛКАНСКА КАРАКТЕРИСТИКА

Ирена Савицка

Универзитет „Никола Коперник“, Торунь
irsawicka@gmail.com

Во статијата се разгледуваат две фонетски појави кај балканските дијалекти во Македонија: ограничувањето на местото на акцентот на последните три слога и двојното акцентирање. Тие се карактеристични за македонско-албанско-грчкото подрачје во центарот на Балканскиот јазичен сојуз. Авторот не се сложува со прифатената теза во балканистичката литература дека западните и централните македонски говори имаат заеднички акцентен тип со грчките, ароманските и албанските дијалекти во овој ареал. Во статијата се даваат аргументи, со коишто се побиваат претходните ставови за централнобалканскиот карактер на ограничувањето на местото на акцентот на последните три слога. Таа појава не заслужува посебно проучување бидејќи претставува универзална тенденција, којашто произлегува од природните надворешнојазични преференции. Во врска со втората појава – двојното акцентирање – авторката издвојува еден тип кој може да има статус на балканска црта. Таквото акцентирање е карактеристично за една балканска микрообласт, каде што словенските и грчките дијалекти се во контакт.

Клучни зборови: Балкански јазичен сојуз, фонетика, меѓујазичен контакт, дијалекти.

ACCENT AS A BALKAN FEATURE

Irena Sawicka

Nicolaus Copernicus University, Toruń

irsawicka@gmail.com

The article examines two phonetic phenomena in the Balkan dialects of Macedonia: the restriction of accent placement to the last three syllables and double accentuation. These phenomena are characteristic of the Macedonian-Albanian-Greek area at the center of the Balkan Sprachbund. The author disagrees with the widely accepted thesis in Balkan literature that the western and central Macedonian dialects share a common accentual system with the Greek, Aromanian, and Albanian dialects in this region. The article presents arguments that challenge previous views on the central Balkan nature of the restriction of accent placement to the last three syllables. The author argues that this phenomenon does not merit special attention, as it represents a universal tendency driven by natural linguistic preferences. Regarding double accentuation, the author identifies a specific type of this phenomenon that may be considered a distinct Balkan trait. This type of accentuation is characteristic of a Balkan micro-region where Slavic and Greek dialects are in contact.

Key words: Balkan Sprachbund, phonetics, language contact, dialects.

1 Вовед

Постоењето на Балкански јазичен сојуз е факт што не може да се оспори. Неговото несомнено единство се заснова на морфосинтаксички карактеристики. На ова треба да се додаде и турскиот слој на лексика, кој сè уште постои, и покрај активностите на јазичните „прочистувачи“ во некои балкански земји. Во поглед на балканската фонетика, се наметнува впечатокот дека за неа не е доволно дискутирано во литературата. Иако во основните описи на Балканскиот јазичен сојуз (на пр. Schaller 1975; Decsú 1972 и др.) се споменуваат две фонетски карактеристики, тие лесно можат да бидат доведени во прашање. Првата се однесува на присуството на темен глас во ранг на фонема – всушност, тоа е карактеристика на многу поширок простор во кој Балканскиот сојуз е периферија (сп. Sawicka 2023; Топоров 1965). Втората карактеристика се однесува на појавата на хомооргански консонантски групи – „назален сонант плус оклузив“ на почетокот на зборот. Во предметната литература, оваа карактеристика е претставена на многу поедноставен начин. Двојниот извор на оваа карактеристика (латински и грчки) значи дека таа функционира различно во различни области на Балканот. Свкупно, оваа особеност најдобро се карактеризира како тенденција за функционална еквиваленција на овие групи со соодветните звучни оклузивни. Оваа тенденција на различни места се јавува со различен интензитет, а во некои од најјужните области на Балканот води до целосна фонологизација.

Треба да се истакне дека формирањето на националните држави значително ги ограничи процесите на конвергенција на Балканот. Сепак, сè уште постои повеќејазична зона во која се среќаваат грчките, албанските, ароманските и македонските дијалекти, и каде што меѓусебната пенетрација на јазичните структури е сè уште многу жива. И овде има целосен збир на класични балкански јазични карактеристики, збогатен со дополнителни заеднички карактеристики, меѓу кои се и некои фонетски особини. Се создаваат и многу интересни микрообласти, во кои одредени појави се заеднички.

2 Акцент

Вообичаените прозодиски појави, забележани во споменатиот ареал, го опфаќаат и типот на акцентирање на зборовите. Се верува дека западните и централните дијалекти во Република Северна Македонија, потоа соседните грчки, аромански и албански дијалекти, претставуваат сличен акцентен тип, иако механизмите за одредување на местото на акцентот не се идентични. Се вели дека сличноста е во тоа што зборовниот акцент паѓа на еден од последните три слога. Иако јас порано многупати сум го поддржувала овој став, сега сметам дека и оваа карактеристика не може да се смета како општобалканска. Да се потсетиме дека во европските јазици се најчести зборовите што имаат два и три слога. Затоа, акцентот во повеќето зборови мора да падне на првиот, вториот или третиот слог, без оглед на тоа како броите, од крајот или од почетокот, и без оглед на тоа кој механизам управува со одредувањето на местото на акцентот. И ова е случај, *mutatis mutandis*, и со другите европски јазици. Напротив, балканскиот акцент е многу разновиден, а сличностите се тривијални. Што се однесува на јазичното ниво на кое се донесуваат одлуките за местото на акцентот, во поголем дел од македонската дијалектна област тоа е фонетското ниво, што значи дека се одредува специфичен слог на зборот на кој паѓа акцентот. Во другите јазици, односно дијалекти од дадената област, напротив, се означува со специфична морфема. Тоа значи дека регулирањето на местото на акцентот доаѓа од подлабоко ниво.

Во источномакедонските дијалекти со „морфолошки“ акцент, постои силна тенденција за стабилизирање на акцентот на истата морфема во парадигмите на номинална флексија, на пр. во формите на лексемата *човек* [ʃov¹ek], det. [ʃov¹eko], pl. [ʃov¹etsci], pl. det. [ʃov¹etsite].¹ Од друга страна, во западниот дел на Егејска Македонија и мал дел од југоцентралниот регион на Република Северна Македонија, акцентот е исто така слободен, но постои силна тенденција да се стабилизира местото на акцентот на претпоследниот слог, како на пр. во единската и во множинската форма на лексемите *гулаб* и *човек*: [g¹oʎup], [guʎ¹ombi], [ʃ¹ovek], [ʃov¹etsi].

Во албанскиот јазик, акцентот е фиксиран, и во номиналната и во глаголската флексија, на пр. *v¹ajzë* ‘девојка’, det. *v¹ajza*, pl. det. *v¹ajzat*, gen., dat. pl. *v¹ajzave* итн.

Во грчките дијалекти, акцентската подвижност е можна во парадигмата на флексијата. Акцентот генерално се поврзува со одредена морфема, но има и голем број дистрибутивни ограничувања, коишто произлегуваат од фактот дека акцентот треба да падне на еден од трите слога кои се бројат од крајот на зборот. Значи, во грчкиот јазик ограничувањата доаѓаат од фонетско ниво. Овој факт (заедно со другите фактори) ја фаворизира мобилноста на акцентот во парадигмата, и покрај јасните тенденции за стабилизација (на пример, при конјугација на глаголот, акцентот има тенденција да се стабилизира на третиот слог, сметајќи од крајот на зборот, а не на одредена морфема).

Ниту еден од овие типови, како што може да се види, не е чист. И во грчкиот и во албанскиот јазик, акцентот на зборот паѓа на еден од последните три слога, иако правилата за избор на нагласениот слог се различни. Исто така, невозможно е јасно да се одреди нивото на кое се определува местото на акцентот на кој било од овие два јазика; покрај тоа, не е сосема јасно дали акцентната единица е прозодиски збор или лексема.

Различни дијалекти од ист јазик имаат различни типови на акцент. На исток од македонското говорно подрачје со фонетскиот акцент, местото на акцентот е поврзано со морфолошката структура на зборот, но има и случаи во кои влијае фонетското рамниште. Ова особено се однесува на ограничувањата на нагласувањето на одредени слогови, на пр., во еден мал регион од дијалекти со морфолошки акцент не може да се нагласува последниот слог.

Во албанскиот јазик, акцентот најчесто паѓа на последниот слог од зборообразувачката основа, но има многу исклучоци, кои најчесто се поврзани со тоа што многу наставки го привлекуваат акцентот, на пр. *punët¹or* ‘работник’, *sport¹ist* ‘спортист’ и др. Интерпретациите, засновани на современите теории, особено оптималната и метричката теорија, формулираат правила од фонетска природа. Така, според Каналис (Canalis 2007), нагласена е претпоследната мора на зборот.

Во конкретни случаи, фонетиката може да влијае на промената на местото на акцентот и таму каде што местото на акцентот е суштински определено од морфологијата, на пример, акцентот во југоисточните македонски говори не останува на истата морфема, иако е парадигматски, кај именки со подвижен вокал, од типот *дош¹ел*, *дош¹а*, *дош¹о*. Од друга страна, и функционалното рамниште може да влијае на промената на местото на акцентот во фонетски регулираниот тип. Тоа влијае на промената на единицата што го носи акцентот и резултира добивање на две различни функционални јазични единици (на пример, *име на n¹азар* – посесивност односно генитивен однос, но *одам n¹а назар* – прилошка синтагма, односно акузативен однос).

Во чистиот „фонетски тип“, нагласената единица е фонетски збор, под што се подразбира низа од слогови со заеднички акцент, познат и како прозодиски збор, стапка, такт, акцентска група итн. Ова значи дека приклучувањето на клитиките мора да доведе до поместување на местото на акцентот во единицата. Така е во западномакедонските

¹ Примерите потекнуваат од југоистокот на Република Северна Македонија.

дијалекти и (со неколку исклучоци) во стандардната норма. Треба да истакнеме дека поместувањата на акцентот на проклитиката низ македонскиот простор зависат од типот на клитиката, а не само од тоа дали имаме работа со неакцентоген збор. Меѓу другото, акцентот одамна не се префрла на предлозите во голем дел од македонскиот дијалектен ареал. Значи, поместувањето на местото на акцентот не е толку важно за сегментацијата на текстот,² бидејќи со текот на времето движењето на акцентот слабее. Така, основната акцентска единица станува лексема, а не прозодиски збор. И тоа донекаде ги ослабува разликите меѓу видовите акценти на Балканот.

Виктор Фридман (2011) и Ронел Александер (Alexander 1995) сметаат дека ограничувањето на местото на акцентот на последните три слога довело до трансформација на енклитиките во проклитики. Мислам дека вреди да се погледне подетално романскиот и другите медитерански јазици (на пример италијанскиот), во кои клитиката се јавува и во проклиза.

Во ова мноштво различни акцентски системи во дијалектите од предметната област, тешко е да се најде заеднички именител. Сепак, фактот дека веројатно сите европски јазици припаѓаат на т.н. *stress-timed* и фактот дека повеќето метрички стапки насекаде се стапки со два или три слога,³ дава основа да се тврди дека акцентот припаѓа на збирот на циклични феномени, коишто спроведуваат когнитивен ритам. Сепак, во литературата има контроверзи на темата поврзана со квалификацијата на концептот ритам. Некои лингвисти веруваат дека ритамот е резултат на фонетски и перцептивни појави и затоа е чисто површински феномен. Постои и верување дека видовите ритам се дериват на други јазични појави, како што се редукиите, акцентот на зборовите итн., а лингвистичкиот ритам е резултат, а не причина за други фонетски појави (Dauer 1983; Auer 1993; Ramus, Nespor and Mehle 1999; Grabe and Low 2002; Dellwo 2010, и др.). Исто така, постојат ставови што целосно го негираат постоењето на изохронизмот, третирајќи го ритамот како случаен феномен (Bolinger 1965; Lehiste 1977; Dauer 1983; Jassem et al 1984). Застапниците на овие гледишта се обидуваат да покажат дека времетраењето на метричките стапки, слогови или мори, е многу различно. По мое мислење, оваа позиција не е адекватна.

Фактот дека перцепцијата на ритамот не се потпира на јасни фонетски корелации не мора да значи дека се потпира на некој фонолошки механизам. Според тоа, лингвистичкиот ритам е независна мрежа, која на говорниот текст му наметнува специфични барања, па затоа е феномен што функционира и на фонолошко ниво.

Стратегиите за сегментација, специфични за јазикот, се изведени од ритмичкиот тип (Nazzi et al 2006). Пред сè, ритамот ја регулира распределбата на нагласените слоговни центри. Нивото на кое го дефинираме јазичниот ритам е централното ниво на прозодиската организација на исказот. Овој став лежи, меѓу другото, во основата на метричката теорија на акцентот, наспроти ставот дека акцентот е сегментална карактеристика (види Liberman 1975; Halle and Vergnaud 1987; Hayes 1981, 1995).

Природниот лингвистички ритам⁴ не само што комбинира две пократки единици во една акцентска единица, туку и ги пресекува подолгите единици на помали. Како резултат на тоа, се појавува т.н секундарен или дополнителен акцент, обично на почетниот слог на подолгите единици, или пак некои морфеми добиваат независен акцент.

Двојното акцентирање на Балканот се јавува во две форми. Овој феномен треба да се гледа како нешто невообичаено и релативно оригинално. Станува збор за двојно

² Сепак, препознавањето на текстуалните единици главно се потпира на семантиката, со оглед на фактот што фонетските гранични сигнали се многу помалку ефективни.

³ Овој факт се смета за европска универзалија во сферата на прозодијата (Златоустова 1983).

⁴ Веројатно условен од когнитивниот ритам, а со тоа и (квази)универзален, барем во европските јазици.

нагласување на зборовите, кое се користи за изедначување на времетраењето на акцентните стапки и за елиминирање на нарушувањата на ритмот. Се појавува во повеќесложни зборови и обично е привремен феномен. Меѓутоа, овде овој дополнителен акцент станува толку силен што луѓето, коишто ги запишувале дијалектите во предметната област, сметале дека е неопходно да одразат два исто толку силни акценти во фонолошкото претставување на зборот. Ваква појава е забележана во многу грчки дијалекти во Егејска Македонија, особено во нејзиниот западен дел. Во некои дијалекти, со текот на времето, еден од овие акценти исчезнал, оригинален или нов, но во некои двата опстојуваат и денес. Двојниот акцент бил карактеристичен за грчките дијалекти низ цела Егејска Македонија. Има почетен акцент што е развиен како реакција на премногу долги стапки, т.е. на стапки подолги од три слога, кои во основата имале парокситонски акцент, на пр. [l¹exasl¹ami] ‘изгубивме’. Целосната парадигма на свршената форма *xaso* (од глаголот *xano* ‘губам’) е: 1 sg. [l¹exasa], 2 sg. [l¹exasis], 3 sg. [l¹exasi(n)], 1 pl. [l¹exasl¹ami], 2 pl. [l¹exasl¹eti], 3 pl. [l¹exasan]. Понатамошниот развој на двојно нагласените зборови бил тринасочен: (1) биле зачувани и двата акценти, (2) бил зачуван оригиналниот (подвижниот) акцент, (3) бил зачуван почетниот акцент. Колонообразен акцент (т.е. оригиналниот акцент плус почетниот, или само почетниот акцент), во основа, ги карактеризира грчките дијалекти од цела Егејска Македонија.⁵

Сличен ефект се постигнува во грчкиот јазик со додавање на енклитиката. Ако зборот има пропарокситонски акцент, тогаш, по присојединувањето на енклитиката кон личната замена, тој добива дополнителен акцент на последниот слог, на пр. *σκοτωσε τον* ‘убиј го’ и *το αλογο μου* ‘мој коњ’ обично се изговараат [sk¹otosl¹eton] и [t¹aloyl¹omu].

Двојно акцентирање на овие простори е забележано и во словенските дијалекти – во македонските дијалекти од источниот дел на Егејска Македонија (во Серскиот и Драмскиот Регион) и во соседните бугарските дијалекти. Истражувачите на тие дијалекти сметале дека е неопходно да се запишат два „главни“ акценти во еден збор (на пр. на неколку места кај Видоески, 2000). Во овој регион, двоен акцент се јавува кај зборовите со четири и повеќе слогови, на пр. во *гасеница* [g¹ɛsinl¹itsa], *градовето* [gr¹adovl¹eto], *ластувица* [l¹astuvl¹itsa], *биволица* [b¹ivoɫl¹itsa] и др. Дополнителниот акцент се јавува и во другите македонски дијалекти, но тој е редок и е често ограничен на одредени морфолошки категории. Во Пиринска Македонија се јавува доста често дури и во тросложни зборови. Така, двојниот акцент се јавува во бугарските дијалекти во широката област на Југозападна Бугарија, во Западна Рупа, во некои родопски дијалекти и низ цела Пиринска Македонија, на пр. [kl¹adenl¹ec] *кладенец* или [k¹azvaml¹e] *казваме*. Истражувачите овде забележуваат два силни акценти (сп. Alexander 1995, 2004; Kolev 2004; Стојков 2002: 224).

Згора на тоа, во Бугарија постои друг вид двојно акцентирање на прозодиските зборови, што подразбира деклитизација. Се дава прозодиска независност на некои кратки зборови, што функционираат како клитики во другите словенски простори. Овој тип двојно акцентирање, поточно поделба на прозодиските зборови на две единици под влијание на семантиката, се јавува дури и во Североисточна Бугарија. Важно е дека не постои ниту структурна ниту географска референца за типот на двојно акцентирање што беше дискутирано погоре.

Во врска со ова прашање, Ронел Александер (Alexander 2004) разликува два главни системи на двоен акцент: (1) два главни акценти во еден збор, чие место е регулирано од ритмот (Југозападна Бугарија), при што обично секој втор слог е под акцент; (2) додадениот акцент е поврзан со одредени морфолошки категории (одбрани клитики што се дел од прозодискиот збор). Овој тип акцентирање се јавува во различни варијанти во

⁵ Повеќе за тоа види во Sawicka (2023: 87), исто така во Papanastasiou и Papadamou (2013).

различни делови на Бугарија, на пример во многу дијалекти, па и во стандардниот бугарски јазик, проклитиката добива независен акцент ако ѝ претходи негацијата, како на пр. во [ne g^lo rozn^lavam] ‘не го познавам’. Во Еркеч, во Североисточна Бугарија, на членските морфемии им се дава дополнителен акцент, на пр. [d^loktorit^le] ‘докторите’.

3 Заклучок

Во статијата се разгледуваат две фонетски појави, карактеристични за балканските јазици: ограничувањето на местото на акцентот на последните три слога и двојното акцентирање. Прво ги наведувам аргументите против тезата дека западните и централните македонски говори имаат заеднички акцентен тип со грчките, ароманските и албанските дијалекти. Сметам дека ограничувањето на местото на акцентот на последните три слога треба да се третира како општа, ако не и како универзална тенденција што произлегува од природните надворешнојазични преференции. Затоа, оваа појава има тривијален карактер и не заслужува посебно проучување. Што се однесува на прашањето за двојното акцентирање, сметам дека само првиот тип може да се смета за заедничка карактеристика на балканската микрообласт, во која грчките и словенските дијалекти се во контакт.

Библиографија

- Видоески, Б. (2000). *Фонолошки бази на говорите на македонскиот јазик*. Скопје: МАНУ. [Vidoeski, B. (2000). *Fonološki bazi na govorite na makedonskiot jazik*. Skopje: MANU. [available at <http://ical.manu.edu.mk/index.php/publications>]. Accessed 29 август 2021.
- Златоустова, Л. В. (1983). Универсалии в просодической организации текста (на материале славянских, германских и романских языков). *Вестник Московского Университета, серия 9, Филология* 4: 69–79. [Zlatoustova, L. V. (1983). Universalii v prosodicheskoj organizacii teksta. *Vestnik Moskovskogo Universiteta, serija 9, Filologija* 4: 69–79]
- Стойков, С. (2002). *Българска диалектология*. София: Академично издателство Марин Дринов. [Stojkov, S. (2002) *Bălgarska dialektologija*. Sofija: akademicheskoto izdatelstvo Marin Drinov]
- Топоров, В. Н. (1965). Несколько замечаний о фонологической характеристике центрально-азиатского языкового союза. [Топоров, В. Н. (1965). Neskolkol'ko zamechaniy o fonologicheskoj harakteristike central'no-aziatskogo jazykovogo sojuza] In S. Drewiak (ed.). *Symbolae linguisticae in honorem Georgii Kuryłowicz*. 322–328. Wrocław: Ossolineum Institute.
- Фридман, В. (2011). *Македонистички студии*. Скопје: МАНУ. [Friedman, V. (2011). *Makedonistički studii*. Skopje: MANU]
- Alexander, R. (1995). The Balkanization of Wackernagel's law. *Indiana Slavic Studies*, 7: 1–8.
- Alexander, R. (2004). The scope of double accent in Bulgarian dialects. In R. Alexander and V. Zhobov (eds.), *Revitalizing Bulgarian dialectology*, 193–210. Published in association with University of California Press. [Available from handle: <http://books.library.upenn.edu/webbin/book/lookupid?key=olbp46045>]. Accessed on: October 10th, 2024.

- Auer, P. (1993). Is a rhythm-based typology possible? A study of the role of prosody in phonological typology. *KontRI Working Paper*, 21. Hamburg: Universität Hamburg.
- Bolinger, D. (1965). *Pitch accent and sentence rhythm. Forms of English: Accent, morpheme, order*. Cambridge, MA: Harvard University Press.
- Canalis, S. (2007). To what extent is Albanian word stress predictable?. In A. Bisetto and F. Barbieri (eds.). *Proceedings of the XXXIII Incontro di grammatica generativa*, 1–14. Bologna: University of Bologna.
- Dauer, M. (1983). Stress-timing and syllable-timing reanalysed. *Journal of Phonetics* 11: 51–62.
- Decsy, G. (1973). *Die linguistische Structure Europas*. Wiesbaden: Harrasovitz.
- Dellwo, V. (2010). *Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence*. PhD Dissertation. Bonn, Universität Bonn.
- Grabe, E., and Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven and N. Warner (eds.), *Papers in Laboratory Phonology 7*: 515–546. Berlin: Mouton de Gruyter.
- Halle, M., and Vergnaud, J. R. (1987). *An Essay on Stress*. Cambridge, Mass.: MIT Press.
- Hayes, B. (1981). *A metrical theory of stress*. Bloomington: Indiana University Linguistic Club.
- Hayes, B. (1995). *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press.
- Jassem, V., Hill, D. R., and Witten, I. (1984). Isochrony in English Speech: its Statistical Validity and Linguistic Relevance. In D. Gibbon and H. Richter (eds.). *Intonation, accent, and rhythm: studies in discourse phonology*, 203–225. Berlin: De Gruyter.
- Kolev, V. (2004). Dialectal accent shifts and double accent in the Bulgarian linguistic region. In R. Alexander and V. Zhobov (eds.). *Revitalizing Bulgarian dialectology*, 215–226. Berkeley: University of California Press. [Available from handle: <http://books.library.upenn.edu/webbin/book/lookupid?key=olbp46045>]. Accessed: October 10th, 2024.
- Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5: 253–263.
- Nazzi, I., Jakiova, G., Betoncini, J., Fedonie, S., and Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*: 54 (3), 283–299.
- Papanastasiou, G., and Papadamou, E. (2013). The position of the northern Greek dialects of Kastoria among the Modern Greek dialects. In M. Janse, B. Joseph, A. Ralli, and M. Bağriaçik (eds.). *Studies in Modern Greek dialects and linguistic theory*, 389–401. *Proceedings of the 5th International Conference of Modern Greek Dialects and Linguistic Theory. Ghent 20–22 September 2012*. Patras: University of Patras.
- Ramus, F., Nespors, M., and Mehle, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73: 265–292.
- Sawicka, I. (2023). *Studies in Balkan Phonetics (Język na Pograniczach)*. Warszawa: Instytut Slawistyki Polskiej Akademii Nauk and Fundacja Slawistyczna 2023, seria: Język na Pograniczach 48. [Available from handle: <https://hdl.handle.net/20.500.12528/1966>]. Accessed: May 5th, 2024.
- Schaller, H. (1975). *Die Balkansprachen: Eine Einführung in die Balkanologie*. Heidelberg: Carl Winter Universitätsverlag.

THE MACEDONIAN AND BALKAN SLAVIC THREE-PART ARTICLE AND THE TYPOLOGY OF DEICTIC ARTICLES

Max Wahlström

University of Helsinki
max.wahlstrom@helsinki.fi

Don Killian

University of Helsinki
donald.killian@helsinki.fi

Macedonian's tripartite article system (proximal, neutral, and distal) aligns with similar systems found in other Balkan Slavic varieties, but also exhibits parallels with non-Slavic languages, such as Wolof, which feature deictic definite articles. This paper investigates the morphosyntactic and semantic features of these articles in Balkan Slavic, aiming to clarify their status within the broader typology of definiteness and deixis. We show through the study of the Macedonian spoken corpus that there are significant differences between the articles regarding their relative frequency. Cross-linguistic comparison further informs the typological status of these articles. In both Balkan Slavic and Atlantic languages like Wolof, definite articles are enclitics, with one form generally occurring significantly more frequently than the others. While interesting secondary functions such as nominal tense have been proposed for these articles, these features are not diagnostic of their "articlehood" and may be expressed by demonstrative pronouns in other languages.

Keywords: deixis, definiteness, proximal, distal, anaphoric reference.

МАКЕДОНСКИОТ И БАЛКАНСКО СЛОВЕНСКИОТ ТРОЕН ЧЛЕН И ТИПОЛОГИЈАТА НА ДЕИКТИЧКИТЕ ЧЛЕНОВИ

Макс Валстром

Хелсиншки универзитет
max.wahlstrom@helsinki.fi

Дон Килиан

Хелсиншки универзитет
donald.killian@helsinki.fi

Системот на тројниот член во македонскиот јазик (проксимален, неутрален и дистален) има заеднички црти со таквите системи во другите балкански словенски дијалекти, но има допирни точки и со несловенските јазици што имаат деиктички определени членови. Во овој труд ги истражуваме морфосинтаксичките и семантичките карактеристики на тројните членови во балканските словенски јазици, со цел да го разјасниме нивниот статус во пошироката типологија на определеноста и на деиксисот. Спроведената анализа на податоците од македонскиот говорен корпус покажа дека постојат значителни разлики помеѓу тие членови, во однос на нивната релативна застапеност. Споредбата со другите јазици дополнително го открива типолошкиот статус на овие членови. Во балканските словенски јазици и во атлантските јазици, како што е Волоф, определените членови се енклитики, од кои еден член значително доминира над другите во поглед на употребата. Постои мислење дека секундарните својства на овие членови не ја определуваат нивната категоријална припадност и често се изразуваат преку демонстративни заменки во другите јазици.

Клучни зборови: деиксис, определеност, проксимален, дистален, анафорска референција.

1 Introduction

The study of the Macedonian tripartite article system over the past decades has both questioned (Topolinjska 2006) and defended (Boronnikova 2014) the status of the proximal and distal markers as definite articles. Some researchers offered a more nuanced interpretation of their functions with no expressed stance on their article status (Sonnenhauser 2009). However, none of these attempts have sought to compare the Macedonian system cross-linguistically beyond other Balkan Slavic varieties, nor employed a definition of the definite article that allows to distinguish it from demonstrative pronouns. This paper seeks to contribute to the discussion by suggesting a theoretical basis for crosslinguistic comparison, briefly introducing findings from other languages with so-called deictic articles, and performing a short corpus study examining the distribution of the articles in spoken Macedonian.

Deictic definite articles are found in several unrelated languages, and are characterized by having more than one element thought to be a definite article, some of which carrying a deictic “flavor” that contrasts with a deictically neutral article. Yet many influential definitions of the definite article are based on the absence of deictic functions – this is also the premise of Topolińska’s (Topolinjska 2006) analysis, which treats the proximal and distal markers as clitic demonstrative pronouns. Moreover, the grammaticalization of demonstrative pronouns into articles is typically described as involving a loss of deixis (see, e.g., Lyons 1999: 331–332).¹

These seemingly contradictory perspectives raise two primary questions: How do the three articles, especially the proximal and distal, differ from deictic demonstrative pronouns, and how does their distribution compare cross-linguistically? Additionally, what added functions do the deictically marked articles have? In this paper, we propose adhering to a typologically tested definition of definiteness that allows us to organize the observations and impressions regarding the deictic article systems. Second, we discuss non-Slavic languages, including Wolof (Atlantic), Ambel (Austronesian), and Classical Armenian (Indo-European), that have been said to display deictic definite articles. Third, we perform a small corpus study on the Macedonian spoken corpus (Escher and Winistörfer [eds.] 2021) to extract basic findings regarding the relative frequency of the three articles, and how they combine with other determiners such as possessive and demonstrative pronouns. The Macedonian findings are then discussed together with what has been reported regarding other Balkan Slavic varieties. In the conclusion, we seek to contextualize Macedonian among other languages claimed to have deictic definite articles, and propose further avenues for research.

2 Defining definite articles

This paper relies on a recent typological study by Laura Becker (2021) on definite articles that summarizes much of the debate of the past decades and establishes concise definitions, informed by an exhaustive cross-linguistic study of article systems in the languages of the world. According to Becker (2021: 86): “A definite article is an article that systematically marks anaphoric, recognitional, establishing, situationally unique, contextually unique, and bridging referents. It may also occur in other types of definite or generic contexts. It does not mark a referent as specific or non-specific”.

Let us now look at the six defining contexts closer.

- *Anaphoric* reference recalls something mentioned earlier in the conversation: *Leon found a book on the table. The book was surprisingly old, given its condition.*

¹ Yet Lyons specifically mentions Macedonian when stating that some articles have not lost deictic distinctions.

- *Recognitional* use contrasts with anaphora: the article signals that both the speaker and the hearer recognize the referent through shared experience or knowledge, but there is no previous mention:
We should go to the park this weekend. (The speaker assumes the listener knows which park is being referred to, because it is already familiar.)
- *Establishing* referents are not identifiable to the hearer at the time of uttering; the speaker signals to the hearer that they can or should construct the referent as (soon to be) identifiable:
Did you hear the news? Some otters escaped from the zoo.
- *Situationally unique* referents are identifiable by being the only available ones in the discourse context:
Please write your answer on the blackboard.
- *Contextually unique* referents, on the other hand, are identifiable in a larger context in which they are unique:
The prime minister (of the country we live in) held a preference conference today about the security leak.
- *Bridging referents* are identifiable through a link that exists between something that has been said before in the discourse; Becker (2021: 81) distinguishes between two different types of bridging referents:
 - *Unique bridging referents* are identifiable by both the speaker and the hearer through an unambiguous link such as a part–whole relationship:
I was driving my bike when the rear wheel suddenly went flat.
 - *Relational bridging referents* in contrast display a looser, less direct link, but the speaker signals that the referent is identifiable in relation to a previously mentioned referent:
We tried out this new restaurant yesterday, and the food was excellent.

While some of these defining contexts are shared with demonstratives, Becker (2021: 103) goes further and identifies contexts in which only definite articles may occur. According to her, only definite articles encode situationally and contextually unique referents, and unique bridging referents. The distinction between unique and relational bridging referents may seem subtle, but modern spoken Finnish offers evidence that this is a real cut-off point between demonstratives and articles. Finnish has often been discussed in the context of developing grammatical marking of definiteness (e.g., Laury 1997). Crucially, the candidate incipient article, the deictically neutral demonstrative pronoun *se*, cannot be used to mark unique bridging referents (example 1), whereas it can be used with relational bridging referents (example 2).²

- (1) *mä ajoin pyörällä ja ?se takapyörä hajos*
 I drove bike and DEM rear_wheel broke
 ‘I was driving a bike and the rear wheel_broke.’
 (Finnish, own knowledge)

² Glosses follow Leipzig standards except for the following: AOR Aorist, ART Article, CLSB Class B, EST Established, FOC Focus, I Inclusive, INSTR Instrumental, INT Interrogative, MID Mid-distance, NHUM Non-Human, NVIS Non-visible, ORI Orientative, PRO Pronoun, REC Recipient, REF Referential, REM Remote Deixis, REMPST Remote Past, TR Transitive, VIS Visible.

- (2) *käytiin yhes ravintolas ja se ruoka oli hyvää*
 we_went one restaurant and DEM food was good
 ‘we went to a restaurant and the food was good.’
 (Finnish, own knowledge)

What this definition entails, is that in order to consider the Macedonian proximal and distal elements as definite articles, they must occur in contexts in which demonstratives cannot, as suggested also by Topolińska (Topolinjska 2006). In what follows, we observe a few parallel systems featuring so-called deictic articles.

3 Towards the typology of deictic definite articles

Deictic definite articles follow the same definition outlined previously; that is, they occur in “anaphoric, recognitional, establishing, situationally unique, contextually unique, and bridging contexts.” However, they also include a deictic component in their meaning, the precise function of which varies from language to language. Lyons (1999: 55) states that “The kind of deictic distinctions which typically occur in demonstratives (distinctions of distance from the speaker, association with different persons, for example) are occasionally found in simple definites.” He further observes that proximity features appear independently of demonstratives, and that, in some languages, the same deictic features are present on both demonstratives and the definite article. Yet, as the data below show, the situation is more complex than Lyons describes.

The referent of a deictic definite article does not have to refer to an object present in the immediate discourse situation; Becker states that “deictic referents can but do not need to be marked by definite articles” (2021: 93). Instead, the languages claimed to have deictic definite articles divide these articles into different categories based on spatial deixis. Examples of such languages include deictic articles in the Atlantic languages Saafi-Saafi, Wolof and Noon, the Austronesian languages Nemi and Ambel, South Slavic Rhodopian, Torlak varieties, and Macedonian.

Wolof, for example, exhibits a two-way deictic distinction in its definite article system, with proximal *Ci* and distal *Ca*, *C* indicating the noun class of the referent the article modifies. Becher (2001:65) contrasts Wolof definite articles using *buur* (king) as an example:

- *buur bi*: the king here; the king who currently reigns
- *buur ba*: the king there; the king of another country or of times gone by

Although the examples Becher gives are of archaic Wolof, Modern Dakar Wolof functions similarly with respect to *Ci* and *Ca*. However, despite clear morphological connections to the demonstratives, distributional analyses suggest that the definite articles do not operate in the same way as demonstratives. The proximal is far more common than the distal, and is also used when deixis not considered relevant for the speaker. This analysis has also been adopted by researchers specializing in Atlantic.³

In a small corpus of 5 texts of Wolof (Robert 2015), out of 170 total definite articles, 147 were marked with proximal and 23 with distal. The distal examples are typically used to indicate clear distance in time or space, and referent tracking does not seem to be relevant to the distinction.

³ Denis Creissels (p.c.) mentions that *Ci* is clearly not proximal, but rather default, contrary to the demonstrative *Cii* from which it originates.

- (3) *waaw* *ñoom* *ñoo* *sanc* *dëkk* ***b-a***
 yes PRO.3PL FOC.SBJ.3PL implant village CLSB.DET-DIST
 ‘Yes, they're the ones who founded the village.’
 (Robert 2015: WOL_SR_CONV_01_SADDA)
- (3) *te* *mu* *ne-aan* *sama* *naar* ***b-a***
 and AOR.3SG say-REMPST POSS.1SG moor CLSB.DET-DIST
des *c-a* *Aalëyétu*
 remain LOC-DIST Aalëyétu
 ‘and about whom he said, "my Moor who remained at Aalayetu.”’
 (Robert 2015: WOL_SR_NARR_04_FALLU-STICK)
- (4) *moo-y* ***y-a*** *maam* ***y-a*** *yor-oon*
 FOC.SBJ.3SG-IPFV CLSY.PL-DIST grandparent CLSY.PL-DIST hold-PST
te *ñu* *sàggan-e=ko*
 and AOR.3PL be_careless-APPL.NHUM=OBJ.3SG
 ‘is that, what the ancestors held and which has been neglected.’
 (Robert 2015: WOL_SR_NARR_03_FALLU.EAF)

All of the languages listed above except for Ambel have similar article systems in that spatial deixis is obligatory across the entire system. While we do not have sufficient data for most of these languages to confirm this definitely, we predict that each language of this type also has some kind of “default” form for when deixis is not relevant.

Ambel has a considerably more complex system, with two types of non-spatial definite articles based on accessibility to the hearer, 32 different spatially-oriented deictic definite articles, and three indefinite articles that differentiate specificity.

- (5) *mokoné: 'nya-kabút* *kalamlí* ***lu-pa*** *be* *tuta-la*
 say.3SG 2SG-*hold* scoop **sea-MID** and 1DU.I-ORI
líl *tuta-mát* *mi* *láp* ***li-ma'***
 land 1DU.I-die INSTR fire **land-DIST**
 ‘He said: “Grab the (seawards) scoop and let's go landwards to extinguish the (landwards) fire.”’
 (Laura Arnold, p.c.)

In addition to definite deictic articles, languages such as Musqueam Halkomelem (Salishan) may incorporate spatial deixis as a component in their article system, but with a broader range of reference types including indefinite (specific) and indefinite (nonspecific) reference. Becker (2021) classifies Musqueam Halkomelem as having “referential articles,” split into three deictic categories: proximal and visible, proximal and non-visible, and remote.

- (6) *niʔ* *skʷtəxʷ* *ʔə* ***tə*** *léləm* ***kʷθə*** *sqʷəməy*
 be.there inside OBL ART:REF.VIS house ART:REF.NVIS dog
 ‘The dog is in the house.’
 (Suttles 2004: 342)
- (7) *ʔi* *ʔəmí* *ʔəʔqəl* ***tə*** *sqʷəməy*
 be.here come exit ART:REF.VIS dog
 ‘The dog came out.’
 (Suttles 2004: 342)

These articles also occur with indefinite semantics (hence the label of referential rather than definite). Nonspecific indefinite referents are generally marked by the remote form of the article, while specific indefinite referents are marked by either the non-visible or visible articles.

- (8) *s-ni²-ct* *wə-k^wác-nəx^w* *tə* *čiči²q² ən*
 NMLZ-AUX-our EST-see-TR ART:REF.VIS mink
 ‘Then we saw a mink.’
 (Suttles 2004: 347)
- (9) *ʔéx^w-əs-t-Samx* *čx^w* *ʔə* *k^wə* *léctən*
 give-REC-TR-me you OBL ART:REF.REM knife
 ‘Give me a knife.’
 (Suttles 2004: 51)

Classical Armenian is also said to have deictic definite articles. Classical Armenian features a three-part system of proximal (-s-), medial (-d-), and distal (-n-) deixis (Müth 2011: 12–13). These elements occur, in addition to, for instance, demonstrative pronouns, as enclitics affixed to nouns, and are considered in the literature definite articles denoting personal affinity: -s refers to the speaker, -d to the addressee, and -n as a neutral definite article. According to Müth (2011: 14–15), Classical Armenian, unlike Greek, generally avoids definite articles in expressing “semantic definiteness,” such as with proper nouns, unique reference nouns, generic reference nouns, and nouns determined by superlative, comparative, or ordinal attributes. Importantly, Müth offers examples of the Classical Armenian article use in “pragmatic” contexts that include the marking of anaphoric, establishing, and both bridging and unique referents. However, in all these examples, only the distal article -n is employed; the study thus does not provide evidence of the proximal and medial elements being used as definite articles.⁴

Finally, another language with proposed deictic definite articles, often discussed in connection with Armenian due to a long-standing areal connection, is Common Kartvelian. According to Harris (1985: 75–77), in what the author refers to as Oldest Georgian, there existed a system of three definite articles: “proximate,” expressing closeness to the speaker, “contingent,” close to the addressee, and “remote,” distanced from both discussants. Hodgson (2022: 128) suggests that prehistoric Kartvelian may have played a role in the development of the Classical Armenian system.

4 Macedonian and Balkan Slavic

In this section, we summarize findings regarding Balkan Slavic. We discuss first what has been argued in the previous literature. We then observe the distribution of articles in a Macedonian speech corpus, and briefly discuss the key features of the deictic articles in other South Slavic varieties such as Torlak and Rhodopian.

Topolińska (Topolinjska 2006: 9–10), who does not classify the Macedonian proximal and distal elements as definite articles, contrasts the neutral definite article with the full

⁴ The status of the definite article that no longer displays deictic distinctions is complicated in Modern Armenian as well. Zolyan (2024) speaks of incomplete grammaticalization of the Modern Eastern Armenian definite article, which retains some of its original demonstrative and possessive meanings. The definite article is used in the nominative and accusative cases to mark subjects and direct objects, and its use is not strictly tied to the semantic characteristics of definiteness or indefiniteness but is more related to syntactic positions. Some argue it is used to mark specific referents (see Hodgson 2022: 146).

demonstrative pronoun containing the same element *-t-*. According to her, using the demonstrative pronoun with situationally and contextually unique referents is either not possible, or results in a change of meaning. Moreover, she argues that the neutral definite article is the only unmarked device that conveys anaphoric reference.

Another context in which the proximal and distal elements do not occur is the expression of generic referents (Usikova 2000: 79).⁵ In Becker's (2021: 86) typology, however, definite articles optionally mark generic reference in some languages, but it is not a defining characteristic.

Topolińska (Topolinjska 2006: 11–13) proceeds to analyze the particular contexts for the proximal and distal elements. These include the expression of spatial deixis functioning similarly to demonstrative pronouns. The contexts in which one could use the neutral *-t-* may further convey a sense of familiarity or closeness to the speaker with the proximal, or temporal and emotional distance with the distal. These functions have been further elaborated by Sonnenhauser (2009), who emphasizes the subtle nuances they introduce, particularly in constructions involving coordination of anaphoric reference. Boronnikova (2014), who defends the status of all three clitic elements as definite articles, highlights their expressive value, describing the proximal as carrying a positive tone, and the distal a negative one.

Both Boronnikova and Topolińska agree that the status of the proximal and distal clitic elements hinges upon their relationship with the demonstrative pronouns. Topolińska, on the one hand, demonstrates how the neutral clitic element differs from the corresponding neutral demonstrative pronoun, carrying certain characteristics that are exclusive to definite articles. However, she argues that the proximal and distal forms reflect the same deictic distinctions as the full demonstrative pronouns and thus do not qualify as definite articles. Boronnikova, in contrast, is less interested in defining the grammatical criteria of definiteness and instead challenges Topolińska's claim about the functional equivalence of the clitics and pronouns. Boronnikova (2014: 63–65) highlights the fact that in spoken Macedonian the proximal and distal clitics may occur in the same NP, which, according to her, proves that they do not serve the same function as demonstratives.⁶

4.1 Corpus study of spoken Macedonian

This small corpus study utilizes two subcorpora of the Macedonian Spoken Corpus (Escher and Winistörfer 2021): The first subcorpus consists of field data from Western Macedonian dialects (in this paper we call the subcorpus *Dialects*), collected from informants in Resen, Janche, Krani, and Arvati, with 90% of the speakers born between 1948 and 1956. In this paper, we only use data from speakers with Macedonian as their first language, resulting in a dataset of 15 informants and 18,542 tokens. The second subcorpus, called *Bombi*, consists of 141,990 tokens and represents the modern urban variety of Skopje, including traits from various

⁵ See also Karapejovski (2022) for recent study on genericity and definiteness in Macedonian.

⁶ Additionally, Boronnikova (2014: 64) gives an example of “generic” use of the distal element: *I ko ke beše blinkerot, trebaše četiri ribi da se fatet, tie fašćea po deset, petnaeset, koj kako, bez kontrola. Ribar-on da fatit deset kila, tri kila ke odnesit na ribarnica i od kaj znajt ovoj, nemat financova kontrola.* ‘And when they were allowed to catch four fish with a lure, they caught ten to fifteen kilograms, as much as they could, without control. The/a/that fisherman catches ten kilos but only takes three to the fishmonger, and how can they know; there is no financial control.’ It is true that the referent is non-specific, but even non-specific referents can be sometimes picked up in discourse; these are what Karttunen (1976) calls short-term discourse referents. In this case, what the highlighted noun represents is rather a relational bridging referent (see Example 2). Although this function can also be conveyed by a demonstrative pronoun, this is particularly interesting because, as Boronnikova points out, the distal element seems to convey a disdainful attitude towards the described practice.

regional dialects due to migration to the capital. There are more than a hundred speakers, with men being overrepresented by 5 to 1.⁷

Research resources for modern spoken Macedonian remain very limited, and this corpus represents therefore an incredibly valuable trailblazing endeavor. However, due to the preliminary nature of the corpus, the accuracy of the automated POS tagging, lemmatization, and morphological parsing is not sufficiently high to replace the search of word forms.⁸ We therefore restrict ourselves to establishing only some of the more basic characteristics of the articles in the data along with a few additional observations about their distribution. This topic undoubtedly merits a more thorough analysis, but due to time constraints, that will have to wait for a later opportunity.

We first sought to study the relative frequencies of the three articles. The three articles of Macedonian belong to entirely different frequency categories, as shown in Figure 1. The distal one is extremely rare, the proximal is more than 50 times more common, yet still only one-sixth as frequent as the neutral article.

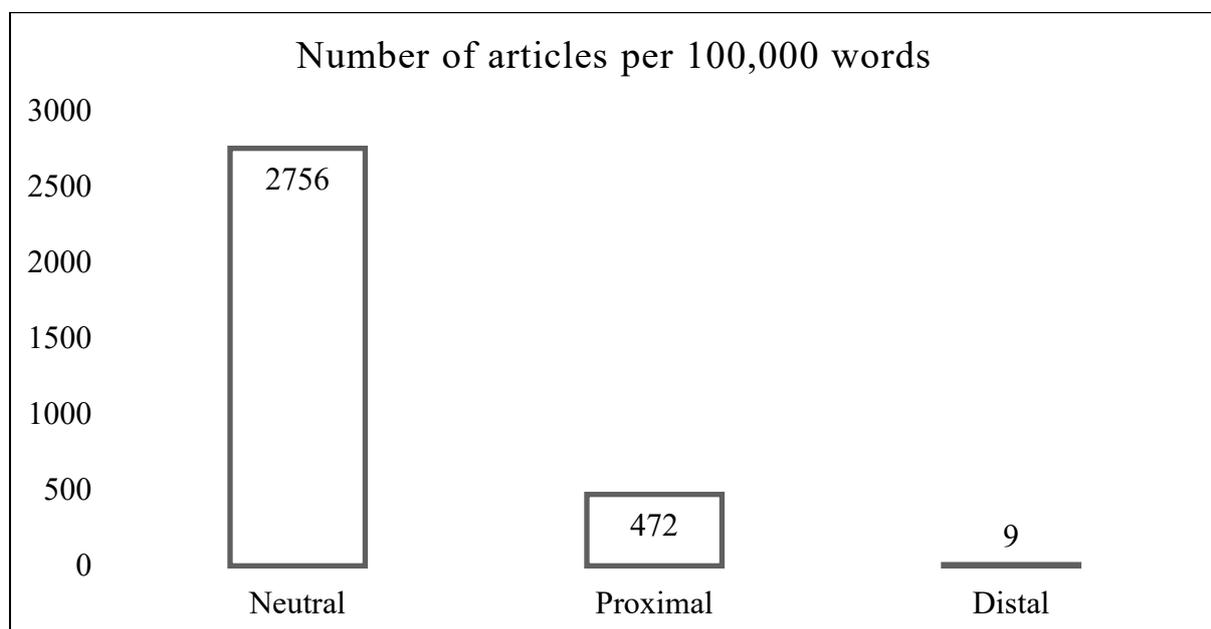


Figure 1. Frequency of the articles in the *Dialects* and *Bombi* subcorpora of the Macedonian Spoken Corpus (Escher and Winistörfer 2021), n=160,532

The corpus provides some opportunities to assess the homogeneity of the data through its metadata. Figure 2 illustrates the distribution of articles across the two subcorpora. The frequency of the proximal article is consistent across the subcorpora; however, the occurrences of the distal article are too few to allow for a reliable determination. Overall, the frequency scale for all articles remains relatively stable.

⁷ The subcorpus contains transcripts of wiretapped conversations of Macedonian political elites from 2008 to 2015 in modern colloquial Macedonian used by educated elites. These transcripts were published by the opposition party SDSM in 2015, revealing the government's high level of corruption. (Friedman 2017.)

⁸ Also, at the time of writing this mid-March 2025, the home page of the corpus is offline.

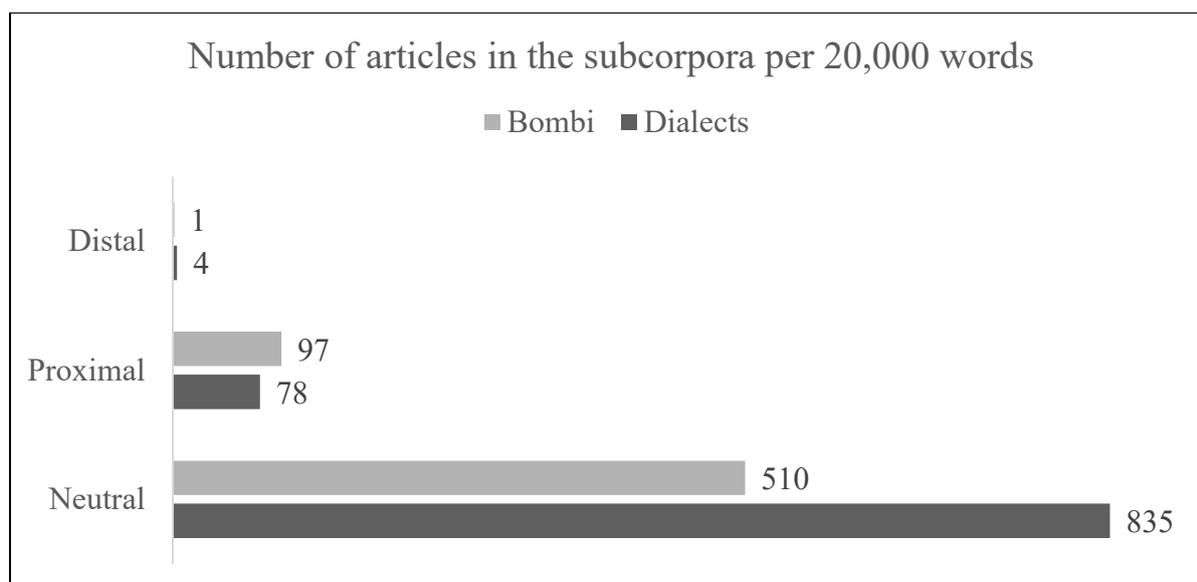


Figure 2. Frequency of the articles per two subcorpora of the Macedonian Spoken Corpus (Escher and Winistörfer 2021), n=160,532

The *Dialects* subcorpus further enables the evaluation of variation related to specific interviews or speakers, although there are enough occurrences only for the neutral article. The frequency of the neutral article in the *Bombi* subcorpus is 2.6%, while the frequencies among speakers in the *Dialects* subcorpus range from 3.1% to 6.9%. This high variance is likely not indicative of permanent linguistic idiosyncrasies of the speakers, but rather is more likely due to chance factors, such as the topics of conversation. Notably, the shortest inputs from a single speaker consist of fewer than 500 words in total.

The variance is a crucial indicator of the sufficiency of corpus size in assessing the definite article. With a corpus of fewer than 200,000 tokens, determining the relative frequencies of the three articles is barely feasible. However, obtaining a representative sample of contexts for especially the distal article, given its low frequency, would require a corpus at least ten times larger.

For the reasons outlined above, the lack of diagnostic contexts for definite articles regarding the proximal and distal clitic elements in this corpus does not serve as evidence against their articlehood. The corpus is simply too small to definitively assess this. However, the low frequency of the proximal as well as the distal elements in particular raises questions about one general condition in Becker's (2021: 86) definition: a definite article must *systematically* mark referents in the six defining contexts. With certainty, we can only confirm that the neutral article fulfils this condition.

4.2 Torlak and Rhodopian

We now turn briefly at the closely related South Slavic varieties of Torlak, spoken in Southeastern Serbia, as well as the Rhodopian dialects of Bulgaria, including Xanthi Pomak, which is part of the same macrodialect.

Vuković (2023) examines clitic demonstrative elements, traditionally classified as definite articles in Torlak, also known as Prizren-Timok, a transitional linguistic variety or a set of varieties between Macedonian and Bulgarian on the one hand and Serbian on the other. The variety Vuković focuses on in particular is Timok Torlak, also known as Prizren-Timok. The *-v-*, *-t-*, and *-n-* elements in Timok Torlak are cognates with the Macedonian ones discussed above. According to Vuković (2023: 265–266), the Timok Torlak demonstrative clitics are not

full definite articles, but, at least, in the case of the neutral *-t-* clitic, they are reminiscent of an anaphoric article. The proximal and distal elements, on the other hand, are more commonly used deictically.

The relative frequencies Vuković (2023: 254) presents are interesting when compared to spoken Macedonian: there are 147 *t*-stem (neutral), 26 *v*-stem (proximal), and 4 *n*-stem (distal) clitics per 10,000 nouns. Since the compared items are nouns, rather than words as in Figures 1 and 2, these results are not directly comparable regarding absolute frequencies. However, assuming that nouns make up no more than half of all the words in the corpus, the neutral clitic is at least three times rarer than its Macedonian counterpart. This provides giving frequency-based support for Vuković's conclusion that Timok Torlak does not have a definite article.

The Rhodopian dialect of Bulgarian, along with closely related Pomak spoken by Muslim Slavs in Northern Greece, both display a three-part system of postposed clitic elements. The key sources regarding the Rhodopian three-part system are Kanevska-Nikolova's (2006) monograph, along with Fanciullo's (2019) in depth analysis. The two works together offer a wealth of discussions and examples, much beyond the scope of this paper. The morphological elements of the three-part system is similar to those of Macedonian and Torlak, but the proximal element derives from a different Slavic demonstrative stem, *-s-*, not *-v-*. However, in other respects Rhodopian differs significantly from the other Balkan Slavic three-part systems: both authors present examples involving other elements of contexts that suggest a much more complicated situation.

According to Fanciullo (2019), the choice of elements depends on spatial interpretations, evidential values, temporal relations, the degree of familiarity, the speaker's attitude toward the information, the word class of the referent, and whether the referent is an inalienable property of the speaker.

Kinship nouns, for instance, can associate the deictic elements with spatial meanings, but these elements can also indicate the relationship of the members (close relationships associated with the *-s-* element, and distant kinship relations with *-n-*). In example (10), *-s-* indicates physical proximity (come *here* often), whereas *-s-* in (11) indicates both a close degree of kinship and possession.

- (10) Rhodope dialect
'Dɔftera-sa ti du'hoda li si 'fostif?
 daughter-DEF.S POSS.2SG come.PRS.3SG INT REFL often
 'Does your daughter come home often?'
 (Kanevska-Nikolova 2006: 68, cited in Fanciullo 2019: 58)

- (11) Rhodope dialect
ja 'ima-m si gra'dinka, ra'zdeli-me
 1SG have.PRS-1SG REFL garden divide.PRS-1PL
i na sna'ho-sa
 and to daughter-in-law-DEF.S
 'I have a vegetable garden, we divide [everything] to (my) daughter-in-law...'
 (Fanciullo 2019: 109)

Definite deictic articles can also express nominal tense. In the following examples, the speaker explains the choice of a particular deictic element, that *-s-* refers to the present tense.

- (12) Rhodope dialect
zaštò kàzva glàdni-se i ne kàzva
 why say.PRS.3SG hungry-DEF.S and NEG say.PRS.3SG
glàdni-te
 hungry-DEF.T
 ‘Why does he say the[s] hungry ones and not the[t] hungry ones?’
 (Fanciullo 2019: 153–154)
- (13) Rhodope dialect
glàdni-se zàštòtu f mum'èntə sə glàdni
 hungry-DEF.S because in moment be.PRS.3PL hungry
 ‘The[s] hungry ones, because it’s right now that they are hungry!’
 (Fanciullo 2019: 153–154)

This expression of temporal values is also possible with unique referents, as in example (14).

- (14) 'mnogo pe'tʃe 'slontse-so
 very shine.PRS.3SG sun-DEF.S
 ‘The sun is shining a lot.’
 (Fanciullo 2019: 131)

According to Fanciullo's (2019) corpus study of Rhodopian, the total amount of deictic *-t-* in the analyzed corpora was 899 (relative frequency 77.5 %), the deictic *-n-* occurred 170 times (relative frequency 14.6 %), and the total number of the deictic *-s-* was 90 (relative frequency 7.7 %). However, assessing relative or absolute frequencies of the elements is complicated by a number of confounding factors, such as the influence of standard Bulgarian, which has only one clitic originating in a demonstrative pronoun, the definite article *-ăt*. Moreover, the use of the elements varies depending on the age and location of the speakers; the use of the element *-t-* is relatively constant, but *-s-* is mainly used by older speakers, and its frequency in the corpus is proportional to the age of the speaker (Fanciullo 2019). The element *-n-* is used by all age groups, but is also more frequent among elderly speakers. Geographically, the tripartite system of deictics is best represented near Smolyan; *-s-* elements in particular are more attested in villages near the center. The frequency of use of *-s-* decreases proportionally with distance from Smolyan (Fanciullo 2019).

5 Conclusions

This article has aimed to contribute to the understanding of the Macedonian so-called tripartite article in several ways: by comparing it with other similar systems in the languages of the world known as deictic articles, by observing frequency-related characteristics of the articles in spoken Macedonian, and by summarizing the previous discussion on the Balkan Slavic systems of deictic articles. Crucially, through introducing a clear-cut definition of the definite article, we have sought to provide answers to the open questions regarding their status.

Regardless of whether all three elements of the Macedonian tripartite article should merit the status of definite article, Macedonian clearly belongs to a very small global group of languages that have what are known as deictic articles. In fact, based on our limited typological sample, with one exception, deictic articles seem to have a default form for when deixis is not

relevant – and that this is likely to be reflected in the relative frequencies of the different articles, with one article dominating over others.

What this entails for the less common articles remains unclear. It seems that the nomenclature defining all the contributing elements to be definite articles is based primarily on morphosyntactic observations: the language in question clearly has grammaticalized marking of definiteness, and all complementarily distributed morphosyntactically and semantically similar elements are called articles, often without much consideration for contexts that differentiate definite articles from other linguistic elements. Therefore, the debate surrounding articlehood of the Macedonian proximal and deictic clitics may mirror the debate faced by other languages with deictic articles, which are, by default, less studied and more poorly resourced than Macedonian.

The debate surrounding the Macedonian language has already brought to light most of the crucial aspects regarding the status of the tripartite article. It appears that there is little disagreement about the empirical facts themselves, but rather about how these elements should be categorized and named. However, without a solid and widely accepted definition of the definite article, it is impossible to resolve the key question. Topolińska (Topolinjska 2006) rightly suggests that only the neutral clitic consistently appears in contexts typically associated with definite articles. Our observations, along with those concerning the Torlak, regarding the relative frequencies of the elements support the argument for recognizing a single definite article.

Boronnikova (2014), on the other hand, makes a valuable point by arguing that clitics and full demonstrative pronouns should not be treated as synonymous. Additionally, observations from Macedonian (Sonnenhauser 2009), Rhodopian and Xanthi Pomak (Fanciullo 2019; Adamou 2011), and Wolof (Becher 2001) suggest that a functional divergence from demonstrative pronouns is indeed characteristic of deictic articles.

Phenomena related to the grammatical or grammaticalized marking of definiteness require extensive data, as we have demonstrated. It is therefore of utmost importance that larger speech corpora of modern spoken Macedonian become available, especially given the likelihood of significant areal variation.⁹ Other methods, such as grammaticality judgments, can be used to determine whether the proximal and distal elements occur in diagnostic contexts of situationally and contextually unique referents and unique bridging referents. However, to fully understand the Macedonian deictic article, it must be ultimately examined within a large, annotated corpus.

Bibliography

- Adamou, E. (2011). Temporal uses of definite articles and demonstratives in Pomak (Slavic, Greece). *Lingua*, 121(5): 871–889. DOI: 10.1016/j.lingua.2010.12.001.
- Becher, J. (1991). *Untersuchungen zu den Verbalweiterungen des Wolof*. MA thesis. Universität Hamburg.
- Becker, L. (2021). *Articles in the World's Languages*. Berlin/ Boston: de Gruyter. DOI: 10.1515/9783110724424.
- Boronnikova, N. V. (2014). Status trojnogo chlena v makedonskom jazyke. *Filologicheskie nauki. Voprosy teorii i praktiki*, 10(3): 60–65.
- Cissé, M. (2007). *Dictionnaire Français-Wolof*. Paris: L'Asiathèque, Dictionnaire des Langues O'.
- Diessel, H. (2012). Deixis and demonstratives. *An International Handbook of Natural Language Meaning*, 3: 2407–2431. Berlin: de Gruyter. DOI: 10.1515/9783110724424.

⁹ We would like to thank an anonymous reviewer for reminding us about the dialectal variation, specifically regarding the Ohrid dialect, an example of which is cited in this paper in the footnote on Boronnikova 2014.

- Escher, A. and Winistörfer, O. (eds.). (2021). *Macedonian Spoken Corpus*. Zürich: UZH Institute of Slavic Studies. escher.pythonanywhere.com, 11 April 2024.
- Fanciullo, D. (2019). *Temporal expression in nominals: tripartite deictics in the Bulgarian Rhodope dialects*. Berlin: Peter Lang. DOI: 10.3726/b15767.
- Friedman, V. (2017). Self and other (svoj i chuzhoj) in the Macedonian ‘bombs’ of 2015. In M. Makartsev and I. Sedakova (eds.), *Balkanskij tezaurus: Vzgljad na Balkany izvne i iznutri. Balkanske Chtenija 14, Moscow, 19–21 April 2017*, 216–229. Moscow: Institut slavjanovedenija, RAN.
- Hanks, W. F. (1992). The indexical ground of deictic reference. In A. Duranti and C. Goodwin (eds.), *Rethinking context*. Cambridge: Cambridge University Press.
- Harris, A. C. (1985). *Diachronic syntax: the Kartvelian case*. Orlando: Academic Press.
- Hodgson, K. (2022). Grammaticalization of the definite article in Armenian. *Armeniaca. International Journal of Armenian Studies*, 1: 125–150.
- Kanevska-Nikolova, E. (2006). *Trojnoto chlenuvane v rodopskite govori*. Plovdiv: Universitetsko izdatelstvo “Paisij Hilendarski.”
- Karapejovski, B. (2022). Razvojot na postpozitivniot morfološki člen vo makedonskiot jazik kako eksponent na generičnata referencija. *Slovene*, 11(2): 244–263. DOI: 10.31168/2305-6754.2022.11.2.11
- Karttunen, L. (1976). Discourse referents. In J. McCawley (ed.), *Syntax and Semantics, 2: Notes from the linguistic underground*, 363–385. New York: Academic Press.
- Laury, R. (1997). *Demonstratives in interaction: The emergence of a definite article in Finnish*. Amsterdam: John Benjamins.
- Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.
- Müth, A. (2011). Categories of definiteness in Classical Armenian. *Oslo Studies in Language*, 3(3): 11–25.
- Robert, S. (2015). *Recorded and annotated oral corpus of Wolof*. CorpOrAn: Corpus Oraux Annotés, https://corporan.huma-num.fr/Archives/corpus.php?codeLangue=code_WOL, 12 August 2024.
- Sonnenhauser, B. (2009). The Macedonian tripartite article: a discourse-oriented account. *Makedonski jazik*, 60: 123–136.
- Usikova, R. (2000). *Makedonskij jazyk. Grammaticeskij ocherk, teksty dlja chtenija s kommentarijami i slovarem*. Skopje: Filološki fakultet “Blaže Koneski.”
- Vuković, T. (2023). A corpus-based analysis of the grammatical status of short demonstratives in the Timok dialect. *Journal of Slavic Linguistics*, 31(1–2): 245–270.
- Topolinjska, Z. (2006). Trojniot člen: Da ili ne? *Južnoslovenski filolog*, 62: 7–15.
- Zolyan, S. (2024). On the so-called “definite article” in Eastern Armenian: Grammatical constraints and pragma-semantic functions. *Indoevropskoe jazykoznanie i klassičeskaja filologija*, 28: 641–672.

LOST IN THE BALKANS: DIFFERENTIAL PLACE MARKING IN THE AROMANIAN VARIETIES

Olivier Winistörfer

University of Zurich
olivier-andreas.winistoerfer@uzh.ch

Anastasia Escher

Federal Institute of Technology Zurich
escher@nexus.ethz.ch

Daria Konior

Independent researcher
toamna13@gmail.com

The phenomenon of Differential Place Marking (Haspelmath 2019), also called zero-marking of spatial relations (Stolz et al 2014), has often been mentioned in the languages of the Balkans. Examples of such differential marking have been documented in the Aromanian varieties (Kramer 1981; Caragiu-Marioțeanu 1975), Modern Greek (Holton et al 1997), Macedonian (Koneski 1965), Ancient Greek (Luraghi 2017), and Latin (Haspelmath 2019; Kramer 1981). However, while the presence of Differential Place Marking has been widely acknowledged, detailed descriptions of such patterns in different varieties are still lacking. Our aim is to present and discuss linguistic data from Aromanian and other Balkan Romance varieties (Istroromanian and Meglen Vlach) to better understand the inter- and intra-dialectal variation of Differential Place Marking. We study and compare their occurrences in the linguistic transcripts from different synchronic Aromanian varieties: from Kruševo (Gołąb 1984), Ohrid and Struga (Marković 2007), and Turia/Kranéa (Bara et al 2005). The results of the comparative analysis suggest that the dialectal and diachronic picture is not uniform. Various semantic factors, such as the type of noun indicating location (proper vs. common) and whether the location is perceived as proximal or distant seem to play a key role.

Key words: Aromanian dialects, Differential Place Marking, language contact, Balkanisms.

ИЗГУБЕНИ НА БАЛКАНОТ: ЛОКАТИВНОТО ДИФЕРЕНЦИЈАЛНО МАРКИРАЊЕ ВО АРОМАНСКИТЕ ДИЈАЛЕКТИ

Оливер Винисторфер

Универзитет во Цирих
olivier-andreas.winistoerfer@uzh.ch

Анастасија Ешер

Федерален технолошки институт во Цирих
escher@nexus.ethz.ch

Дарија Кониор

Независен истражувач
toamna13@gmail.com

Феноменот на диференцијално локативно обележување (Haspelmath 2019), познат и како нулто обележување на просторните релации (Stolz et al. 2014), е карактеристичен за јазиците на Балканот. Примери за вакво диференцијално обележување се документирани во ароманските говори (Kramer 1981; Saragiu-Marioțeanu 1975) и во балканските јазици: грчкиот (Holton et al. 1997), македонскиот (Koneski 1965); во класичните јазици: старогрчкиот (Luraghi 2017) и латинскиот (Haspelmath 2019; Kramer 1981). Но, иако диференцијалното локативно обележување се смета за балканска карактеристика, сè уште нема детални описи за нејзината дистрибуција во различни говори. Целта на нашето истражување е да ги анализираме јазичните податоци од ароманските и од другите балкано-романски варијанти (исторороманскиот и мегленовлашкиот) за подобро да ја разбереме дијалекталната варијабилност на диференцијалното локативно обележување. Ги анализираме примерите без предлог во јазичните транскрипти од неколку современи аромански варијанти: од Крушево (Goiaș 1984), Охрид и Струга (Марковиќ 2007), како и од Турија/Кранеа (Bara et al. 2005). Нашата компаративна анализа покажува дека дијалектолошката и дијахрониската слика на оваа појава не е униформна. Различни семантички фактори, како што е типот на именката во топонимот (општа наспроти лична) и растојанието до определеното место играат главна улога во употребата на диференцијалното локативно маркирање.

Клучни зборови: Аромански дијалекти, диференцијално локативно обележување, јазичен контакт, балканизми.

1 Introduction

Phenomena of Differential Place Marking (Haspelmath 2019), also known as zero-marking of spatial relations (Stolz et al. 2014), have received increasing attention in the linguistic research over the past decades. In such zero-marking constructions, speakers omit any overt marking that indicate spatial roles of Location, Direction or Source/Ablative.¹ Such patterns are also observable in many spoken varieties of Europe, e.g., in spoken British English (1) and Swiss German (2).²

- (1) You're "going Marbella" for the hols. (English, <https://www.thetimes.com>)
- (2) *Gö-mmer* *HB?*
 go.1SG.PL-we.NOM main.station(M)
 'Are we going to HB (main station in Zurich)?' (Swiss German, Zurich; own example)

In English and Swiss German, one would expect an overt marking with a preposition, in (1) *to* and in Swiss German 2 *zu* + *DEF*. In both cases, zero-marking of spatial relations is highly marked and appears to signal the use of a specific spoken register.³

Such phenomena are not unknown to Balkan linguistics. Sandfeld (1930) noted instances of zero-marking, i.e., without any preposition, in marking patterns of Direction in Albanian (3), Balkan Romance (4), Balkan Slavic (5), and Greek (6).

- (3) *Laskoviq-Ø* *vajta*
 Laskovik-NOM/ACC.SG.INDF go.1SG.AOR
 'I went to Laskovik.' (Albanian, Sandfeld 1930: 111)
- (4) *mi duc Sărun-ă*
 I.ACC go.1SG.PRS Thessaloniki(F)-NOM/ACC.SG.INDF
 'I go to Thessaloniki.' (Aromanian, Sandfeld 1930: 111)
- (5) *site sel'an-i-Ø 'od-el-e cărkov-Ø*
 all.PL villager(M)-PL-INDF go.IPFV-PTCP-PL church(F)-SG.INDF
 'All the villagers went to church.' (Balkan Slavic, Sandfeld 1930: 111)
- (6) *πά-ω σχολεί-ο*
 go.PFV-1SG.PRS school(N)-NOM/ACC.SG
 'I go/will go to school.' (Greek, Sandfeld 1930: 110)

Examples (3) to (6) show that the zero-marking is not only attested in all four Balkan varieties, but that it can also occur with different noun types, both proper and common nouns. Vidoeski (1999: 25) considered the "loss of the preposition" one of the more recent Balkanisms. Stolz et al. (2014: 76) argue that the zero-marking patterns are "suggestive of an areal feature which, however, does not count as a fully-blown Balkanism."

¹ See Stolz et al. (2014) or Haspelmath (2019) for the definitions of these comparative concepts (Haspelmath 2010).

² We would like to thank Prof. Dr. Adam Ledgeway for pointing out the existence of this structure in spoken British English.

³ Future studies might be able to shed new light on the use of the pattern in these varieties.

These zero-marking patterns have also attracted attention from a typological perspective. Stolz et al. (2014) and Haspelmath (2019) observed that such patterns tend to be more frequent with proper nouns than with common nouns, and are more common in patterns of Location and Direction than those indicating Source or Ablative. In their cross-linguistic study of the phenomenon, Stolz et al. (2014) examined these patterns in various Balkan varieties. In addition to South Macedonian and non-standard Greek varieties, they included the Aromanian variety spoken in Kruševo (Republic of North Macedonia), as a representative case. Their analyses were based on the linguistic data presented in Gołąb (1984). According to the findings of their large-scale study, zero-marking in Aromanian is limited to toponyms denoting cities and only appears in contexts of Location and Direction, but not Source (Stolz et al 2014: 73–75). This observation is largely in line with Sandfeld's (1930: 111) claim that these patterns in Aromanian mostly occur with toponyms.

Similar phenomena of zero-marking of spatial relations were also mentioned in the description of the Pindean variety of Aromanian in Turia/Kranéa in Greece (Bara et al 2005: 54–55). Among the examples provided, however, there are also cases of zero-marking with common nouns like *filāk'ia* 'prison'.

Hence, the question arises as to whether there is inter- and intra-diatopic variation in the zero-marking patterns of the Aromanian varieties. To address this, we aim to shed new light on these patterns analysing data from different diatopic varieties.

1.1 Research Questions

Given the divergence among existing descriptions of zero-marking patterns across the individual varieties, we decided to conduct a pilot study to gain a preliminary, comparative impression of the diatopic variation of zero-marking of spatial relations in the Aromanian varieties. This study is guided by the following research questions:

1. Are phenomena of Differential Place Marking observable in different branches of Aromanian, i.e., Fărshărot variety in Ohrid and Struga, Gramostean variety in Kruševo (both in North Macedonia), and in the Pindean variety of Turia/Kranéa (Greece)?
2. Are these Differential Place Marking or zero-marking patterns consistent across varieties, or there is evidence of both inter- and intra-diatopic variation?
3. Which semantic factors play a major role in explaining the cases of zero-marking in spatial relations? Do the patterns align with the proposed typological hierarchies, i.e., are they restricted or more wide-spread with proper nouns than with common nouns (Haspelmath 2019; Stolz et al 2014)? Is there also an asymmetry in the marking of Source compared to Location and Direction?

To address these research questions, we analyse the transcripts from linguistic fieldwork conducted on different Aromanian varieties. The data collection methodology and analysis are presented in the following section.

2 Methods

In our study, we consider the following three Aromanian varieties in the Republic of North Macedonia and Greece:



Figure 1. The Aromanian varieties included in the study.

We chose these three varieties for a number of reasons. On the one hand, they represent three different branches of Aromanian according to Saramandu (2014). Nonetheless, there is no clear consensus among scholars on the dialectological classification of Aromanian. Caragiu-Marioțeanu (1975: 264–265) and Kahl (2007: 131) argue for a bi-partition (Fărshărot vs. non-Fărshărot varieties) of the Aromanian varieties based on phonological and morphological features. However, this disagreement does not affect our study, as the three selected varieties still cover both proposed branches. On the other hand, there is sufficient linguistic data to allow for a deeper and broader understanding of the situation in the individual varieties thanks to the descriptions provided by Gołab (1984), Bara et al (2005), and Marković (2007). Moreover, these transcripts have a linguistic, dialectological focus and therefore do not suffer from interference stemming from literary tradition or political views.⁴ Since zero-marking phenomena are already considered dialectal (and probably less prestigious) even in spoken Greek and Macedonian (Stolz et al 2014: 304–305), one can reasonably assume that similar tendencies are observable in Aromanian literary texts and/or culturally-oriented collections (e.g., collections of local tales, etc.).

In the next step, we examined the transcripts for possible marking patterns of Location, Direction, and Ablative/Source. Since there is no systematic corpus of the Aromanian varieties (yet), we relied on OCR to process the texts. We searched for toponyms, different common nouns that are typically prone to zero-marking in different Balkan languages (Sandfeld 1930;

⁴ We have also considered literary texts, local newspapers and other sources in Aromanian. However, many of them seemed to show interferences from other literary languages, such as Daco-Romanian, Greek, and/or Albanian due to their respective written traditions.

Holton et al 1997: 335), such as the local forms for *school*, *church*, *prison*,⁵ as well as verbs of motion, e.g., *to go*, *to arrive*, *to leave*. We then compiled all identified occurrences (more than 330) into the XML file in the supplementary materials to ensure the data adhere to FAIR principles⁶ (Wilkinson et al. 2016). We also included marking patterns with the preposition *pãñã*⁷ ‘till, until, towards’ in the compiled file, but decided to exclude these from the analysis for two reasons: first, there are only a few examples in the linguistic data; second, its semantics can be considered more marked than those of the other prepositions of Location and Direction *la/tu/ən*. In the compiled XML file, we used different variables to describe the constructions, e.g., type of verb/noun, different characteristics of the noun/prepositional phrase (Location vs. Direction vs. Ablative/Source semantics, proper vs. common noun, definite vs. indefinite), whether there is any preposition and the expected preposition for the given variety. This approach enabled a deeper understanding of the different factors that might influence marking patterns in Aromanian.

3 Results

As the results of this comparative analysis show, we can observe phenomena of zero-marking patterns of spatial relations in all three Aromanian varieties. We classify the general patterns as follows: zero-marking patterns of Location/Direction with toponyms (section 3.1), with common nouns (3.2), and Differential Place Marking of Source (3.3).

3.1 Zero-marking of spatial relations (Location and Direction) with toponyms

In all three Aromanian varieties, Differential Place Marking patterns are attested with toponyms. However, these patterns are only observable with local toponyms, i.e., names of villages and towns/cities, as in examples (7) to (10), and not regions or countries (11).

- (7) *Am ti nãžeri Skopj-a*
 have.1SG.PRS of go.INF Skopje(F)-NOM/ACC.SG.DEF
 ‘I have to go to Skopje.’ (Fărshãrot Aromanian, Ohrid & Struga; Markovič 2007: 158)

- (8) [...] *ci s-ãre fapt-ə*
 [...] that 3SG/PL.REFL=have.3SG.PRS.AUX make.PTCP-F.SG
Óhãrd-a tu mähãl-ă.
 Ohrid(M)-NOM/ACC.SG.DEF in/to district(F)-NOM/ACC.SG.INDF
 ‘[...] that was born in Ohrid, in a district.’ (Gramostean Aromanian, Kruševo; Gołab 1984: 145)

- (9) [...] *si-d'uți gr'ebin-e.*
 [...] 3SG.REFL.DAT=go.3SG.PRS Grevena-NOM/ACC.SG.INDF
 ‘[...] he goes to Grevena.’ (Pindean Aromanian, Turia/Kranéa; Bara, et al 2005: 115)

⁵ Unlike Sandfeld (1930) and Holton (1997), we had to exclude the vast majority of occurrences for house/home as they were mainly marked through the adverb *acasã*, like in many Romance varieties.

⁶ FAIR stands for findable, accessible, interoperable, and reusable data.

⁷ There is no standardised transcript for Aromanian. In the examples, we keep the transcription provided by the authors. In our own transcripts we follow Cunia’s (2010).

- (10) [...] *m'ini mi te'a gr'ebin-e, tã*
 [...] I.ACC I.ACC bring.3SG.IMP Grevena-NOM/ACC.SG.INDF in/to
 nosokom'i-u.
 hospital(N)-NOM/ACC.SG.INDF [...]
 '[...] he brought me to the Grebine, to the hospital [...].' (Pindean Aromanian,
 Turia/Kranéa; Bara et al 2005: 241)
- (11) [...] *š vrém ta-s-fúg tu*
 [...] and want.1SG.IMP FUT-SBJV-flee.1SG.PRS in/to
 Elád-ə.
 Greece(F)-NOM/ACC.SG.INDF
 '[...] and I wanted to flee to Greece.' (Gramostean Aromanian, Kruševo; Goľab
 1984: 144)

However, zero-marking does not occur with all toponyms to the same extent. In the Färshärot varieties of Ohrid and Struga, it seems to be systematic with larger cities such as Ohrid, Bitola, and Skopje, but not with local village names like Gorna Belica (12) or Gorica (13).

- (12) *Tu Beala z-'dormi bună.*
 In/to Gorna.Belica.NOM/ACC.SG.DEF 3SG/PL.REFL=sleep.3SG.PRS well
 'One sleeps well in Gorna Belica.' (Färshärot Aromanian, Ohrid & Struga; Marković
 2007: 130)
- (13) *Sil'-a 'esti dus-Ø tu*
 Sila(M)-NOM/ACC.SG.DEF be.1SG.PRS.AUX go.PTCP-M.SG in/to
 Goric-a
 Gorica(F)-NOM/ACC.SG.DEF
 'Sila went to Gorica.' (Färshärot Aromanian, Ohrid & Struga; Marković 2007: 146)

In the other two varieties, the size of the settlement does not seem to play a significant role. Zero-marking is overwhelmingly predominant with local toponyms such as Bitola, Kruševo, Ohrid, and Skopje. There seem to be occurrences of prepositional marking with more distant toponyms like *Biligrádu* 'Belgrade' and *Póle* 'Istanbul', but not with *Bešli* 'Vienna' and *Parísl'i* 'Paris':⁸

- (14) [...] *mi dúk^u ən Pól-e.*
 [...] I.ACC.REFL go=1SG.PRS in/to İstanbul-NOM/ACC.SG.INDF
 '[...] I go to İstanbul.' (Gramostean Aromanian, Kruševo; Goľab 1984: 244)
- (15) *mi dúku Beš-li.*
 I.ACC go=1SG.PRS Vienna(M).NOM/ACC.SG.DEF
 '[...] I go to Vienna.' (Gramostean Aromanian, Kruševo; Goľab 1984: 208)

The only apparent explanation for this difference in marking patterns could be the presence of a Definiteness marker with the toponyms Paris and Vienna. However, other indefinite

⁸ Note here also the use of the preposition *ən* and not *tu* as in the Färshärot variety of the region.

toponyms, such as *Bitule*, also exhibit zero-marking. Therefore, additional data are needed to clarify the underlying factors.

In the Pindean variety of Turia/Kranéa, variation is also attested. In this case, there seems to be a general variation between nouns with and without preposition (16).

- (16) *nu-ave'a* *k'al-i* *au'a* *t ā*
 NEG=have.3SG.IMPF STREET(F)-NOM/ACC.SG.DEF here in/to
gr'ebine
 Grevena(F).NOM/ACC.SG.INDF
 'There was no street in Grevena.' (Pindean Aromanian, Turia/Kranéa; Bara et al 2005: 253)

Nonetheless, zero-marking is clearly predominant with toponyms in the Aromanian variety of Turia/Kranéa. Moreover, there is no significant difference in the tendency towards either zero- or overt-marking of spatial relations between Location and Direction in any of the varieties.

3.2 Zero-marking of spatial relations

In this subsection, we look at Location and Direction with common nouns. In the linguistic transcripts from the Aromanian varieties spoken in Kruševo and Turia/Kranéa, we also observe instances of zero-marking for Location and Direction with common nouns. Interestingly, the two varieties do not show these patterns with the same common nouns. We found cases of zero marking with the noun *skulīe* 'school' (17)⁹ in Kruševo, while in Turia/Kranéa it occurs with *filāk'ii* 'prison' and *ho'arā* 'village', as in (19) and (20).

- (17) *Mīne* *ši* *frác-l'i* *amei* *ənvicám*
 I.NOM/ACC and brother(M)-NOM/ACC.PL-DEF 1SG.POSS.M.PL learn.1SG.IMP
skulīe [...] [*skulī-e* [...]]
 school(F)-NOM/ACC.SG.INDF [...] [*skulī-e* [...]]
 'My brothers and I learned at school [...].' (Gramostean Aromanian, Kruševo; Gołąb 1984: 144)

- (18) *əmvéc^u* *la* *skulīe*.
 teach.1SG.PRS at/to school(F)-NOM/ACC.SG.INDF
 'I teach at school.' (Gramostean Aromanian, Kruševo; Gołąb 1984: 143)

- (19) *ālu* *bāg'arā* *filāk'ii*
 he.ACC.M.SG put.1SG.PRS prison(F)-NOM/ACC.SG.INDF
L'ārs-a.
 Larissa(F)-NOM.ACC.SG.DEF
 'They put him into prison in Larissa.' (Pindean Aromanian, Turia/Kranéa; Bara et al 2005: 54)

⁹ Examples (17) and (18) also show that the verb *nvets/əmvéc* means 'to learn/to teach' in Aromanian, unlike in many other Romance varieties.

- (20) *kum fuḍ* *di trāk'olu* *'unā*
 how leave/flee.3SG.PST of Trakol(M).NOM/ACC.SG.INDF INDF.NOM/ACC.SG.F
ho'ar-ā
 village(F)-NOM/ACC.SG.INDF
 'How he left/fled from Trakol to a village.' (Pindean Aromanian, Turia/Kranéa; Bara et al 2005: 54)

As (18) shows, *skulie* can also occur with the preposition *la*. The same is true for *ho'ar-ā* and *filāk'ii*, which can appear with the preposition *tu* in the Turia/Kranéa variety. Thus, this indicates intra-diatopic variation. However, given a small number of occurrences with these common nouns, we cannot make any claims about the possible factors influencing these marking patterns. Nonetheless, the zero-marking patterns do not seem to be as predominant with these common nouns as with the proper nouns.

In addition, no cases of zero-marking were observed with any other common noun, e.g., *kəsəbə* 'city', *nosokom'iu* 'hospital', *plat'eia* 'square', *bise'arkā* 'church'.

3.3 Differential Place Marking with Source/Ablative

There are also occurrences of Differential Place Marking in the marking of Ablative or Source in the Aromanian variety of Turia/Kranéa:

- (21) *ḡambr'o-lu* *īar'a* *di* *tu* *a'estu*
 groom(M)-NOM/ACC.SG.DEF be.3SG.IMP from/of in/to DEM.PROX.M.SG
māhāl-ā.
 district(F)-NOM/ACC.SG.INDF
 'The groom was from this district.' (Pindean, Aromanian, Turia/Kranéa; Bara et al 1984: 114)

- (22) *īci* *duku* *di* *la* *bāse'arik-ā*
 there/then go.1SG.PRS of/from in/to church(F)-NOM/ACC.SG.INDF
ak'asā
 home
 'then I went from church home.' (Pindean Aromanian, Turia/Kranéa; Bara et al 2005: 471)

- (23) *v'in'i* *'alt-u-Ø* *fil'ak'u-Ø*
 come.3SG.PRS other-NOM/ACC.SG.M-INDF prisoner(M)-NOM/ACC.SG-INDF
di *gr'ebin-e*.
 of/from Grevena(F)-NOM/ACC.SG.INDF
 'another prisoner comes from Grevena.' (Pindean Aromanian, Turia/Kranéa; Bara et al 2005: 244)

- (24) [...] *kum fuḍ* *di* *trāk'olu*
 [...] how flee.3SG.PST from/of Trakol(M)-NOM/ACC.SG.INDF
'unā *ho'ar-ā*.
 INDF.F.NOM/ACC.SG school(F)-NOM/ACC.SG.INDF
 '[...] how he fled from Trakol to a village.' (Pindean Aromanian, Turia/Kranéa; Bara et al 2005: 241)

Examples (21) to (24) are not zero-marking patterns *stricto sensu*, but they show that the marking patterns are not identical. Common nouns (21 and 22) use the preposition clustering *di* ‘of/from’ + *tu/la* ‘in/at’ to mark Source. Proper nouns (23 and 24) can only use the preposition *di*. Thus, there is an asymmetry in the marking patterns, with the proper nouns showing ‘a shorter marking pattern’ as predicted (Haspelmath 2019: 319).

In the linguistic data from the Fārshārot variety of Ohrid and Struga, there are only examples of Source with proper nouns. These cases are particularly intriguing as they use only the preposition *di*. A few examples suggest similar tendencies in Kruševo.

- (25) *Inši* *di=tu* *ód-ə* [...] .
 exit.3sg.pst from.at=to room(f)-nom/acc.sg.indf [...]
 ‘S/he gets out of a room [...].’ (Gramostean Aromanian, Kruševo; Gołąb 1984: 165)

- (26) *N’-aveám* *tát-ə=n’u* *din*
 I.DAT=have.1SG.IMPF father(M)-NOM/ACC.SG.INDF=1SG.DAT from=at/to
 Kəstúr-Ø.
 Kastoria(M)-NOM/ACC.SG.INDF
 ‘I had a father from Kastoria.’ (Gramostean Aromanian, Kruševo; Gołąb 1984: 165)

- (27) *N’-aveám* *dad-ə,* *laiu,*
 I.DAT=have.1SG.IMPF mother(F)-NOM/ACC.SG.INDF black.NOM/ACC.SG.M.INDF
di *Nevésk-a.*
 from Neveska/Nymfaio(F)-NOM/ACC.SG.DEF
 ‘I had a mother from Neveska/Nymfaio.’ (Gramostean Aromanian, Kruševo; Gołąb 1984: 165)

Example (26) shows a different preposition cluster compared to the common noun in (25). In (27), there is only the Source preposition *di*. However, these examples should also be approached with caution, as both can be interpreted not only as Source, but also as Possessive from a semantic perspective.

4 Discussion

The results from section 3 suggest that phenomena of Differential Place Marking or zero-marking occur in the three Aromanian varieties. However, the three varieties seem to show inter- as well as intra-diatopic variation, as seen in Table 1.

Table 1. Patterns of Differential Place Marking in the Aromanian varieties.

	Toponyms proximal cities	Toponyms proximal villages	Toponyms coun-tries/regions	Common nouns	Source with cities
Fārshārot (Ohrid, Struga)	Yes	No	No	No	No
Gramostean (Kruševo)	Yes	Yes	No	Some	Not clear
Pindean (Turia/Kranéa)	Yes	Yes	No	Some	Yes

Zero-marking of spatial relations (Location and Direction) with local toponyms can be observed in the transcripts of all three varieties. In the Fārshārot variety of Ohrid and Struga, it is restricted to the names of proximal towns, while in Gramostean of Kruševo and in Pindean from Turia/Kranéa it can be also attested with village names. None of the varieties use the zero-marked pattern with names of larger regions or countries.

In the Aromanian variety of Kruševo, zero-marking is clearly predominant for proximal city names; for the proper names of Bitola, Kruševo, and Skopje, only zero-marked forms were found. With more distant toponyms such as Belgrade, Vienna, Istanbul, and Paris, there was variation. Morphosyntactic features of the nouns did not provide a clear explanation for these differences. Geographic proximity might be the most likely factor, though more data would be needed to make any claims.¹⁰ In the Pindean variety of Turia/Kranéa, we observe free variation between zero-marking and the preposition, although the zero-marking is clearly predominant with proper names. Factors such as Definiteness or the distinction between Location and Direction do not appear to play a significant role, neither in the variety of Kruševo, nor in that of Turia/Kranéa.

As discussed in the previous section, we have also encountered cases of zero marking for Location and Direction with a restricted number of common nouns in the Gramostean variety of Kruševo and in the Pindean variety of Turia/Kranéa. The common nouns that showed such zero-marking with Location and Direction were not entirely unexpected. Many of them, such as *skulīe* ‘school’ in Kruševo and *filāk* ‘ii’ ‘prison’, *ho'arā* in Turia/Kranéa, have also been noted in other Balkan varieties, including Balkan Slavic and Greek (Sandfeld 1930; Holton et al 1997; Vidoeski 1999). What is crucial is that zero-marking in the Aromanian varieties does not seem to be restricted to only one preposition *tu* or *la*. In neither of the two varieties is zero-marking predominant; and these findings should be treated with caution as the absolute number of occurrences with common nouns remains low. Nonetheless, the existence of these zero-marking patterns already points to a more complex situation in the Aromanian varieties than was suggested by Stolz et al. (2014).

Since zero-marking patterns can be observed in all the synchronic Aromanian varieties included in the study, one might assume that such patterns were also present in earlier linguistic documents of Aromanian. However, we did not observe any occurrence in Kristophson's (1974) critical edition of the *Tetraglosson* of Daniel Moscopolites from 1794, neither in the Aromanian passages (28 and 29) nor in the Balkan Slavic text (30).

¹⁰ Proximity as a factor in the zero-marking patterns was also mentioned during a private conversation with Afrodita Totsili, a native speaker of the local variety of Kruševo, who unfortunately passed away in 2024. A future, systematic study with a larger number of native speakers of the Kruševo variety might shed some light on this issue.

whether multi-lingual individuals use and accept zero-marking patterns to the same extent across the different varieties. Finally, it is worth investigating whether there are differences between monolingual and multilingual speakers.

The cases of Differential Place Marking for the comparative concept of Source/Ablative in the Aromanian variety of Turia/Kranéa are also intriguing. They are not zero-marking *stricto sensu* as the preposition *di* ‘from’ is still present to mark Source. Nonetheless, there is again the opposition between proper nouns and common nouns, as the former do not show the prepositional clusters. Hence, their marking patterns are asymmetric, and the place name is indeed shorter, as predicted by Haspelmath (2019: 319). Similar prepositional clusters have also been attested with common nouns in Macedonian (Ganenkova 2015: 197):

- (32) *ist-o taka mi se sluč-i vrvii*
 same-N.SG like.that 1SG.DAT REFL.ACC happen.3SG.AOR pass.3SG.PRS
ed-en dečk-o-Ø od vo kol-a-Ø [...]
 one-M.SG boy(M)-SG-INDF from in/to car(F)-SG-INDF [...]
 ‘This also happened to me, a guy came out of a car [...].’ (Macedonian,
<https://forum.femina.mk>)

Preposition clustering with common nouns as in (32)¹² is possible in spoken Macedonian. At the same time, zero-marking with toponyms (and common nouns) is also widely attested in the Macedonian varieties (Vidoeski 1999: 25). Thus, the situation is comparable to the patterns in the Aromanian variety of Turia/Kranéa. This raises the question of whether the tendencies regarding preposition clustering would also be similar, i.e., that preposition clusters are only possible with common nouns or they also occur with proper nouns.

5 Conclusion

The results of this preliminary study suggest that patterns of Differential Place Marking/zero-marking are frequent across different branches of Aromanian and not only in the Gramostean variety of Kruševo. However, the analyses of Fārshārot in Ohrid and Struga, Gramostean in Kruševo, and Pindean in Turia/Kranéa provide a more complex picture with differences in the individual zero-marking patterns.

The zero-marking patterns in the varieties are not random. Rather, they seem to follow the hierarchies established in linguistic typology (Stolz et al 2014; Haspelmath 2019): they are predominant with proper nouns for local toponyms (cities and to a certain extent smaller localities) in all three varieties considered. They can even be observed with common nouns like *skuliie* ‘school’ in Kruševo, *filāk’ii* ‘prison’ and *ho’arā* ‘village’ in Turia/Kranéa, although to a lesser extent. There are no zero-marking patterns with Source *stricto sensu*, i.e., without any preposition.

Nonetheless, our analysis of the linguistic data suggests that there is more intra-diatopic variation with common nouns, while the zero-marking patterns are less predominant in both varieties (Kruševo and Turia/Kranéa). The low absolute numbers of occurrences are also to be taken into account.

Therefore, a systematic study of the phenomenon of zero-marking pattern – taking into account the inter- as well as intra-diatopic variation in Aromanian – would be necessary. A combination of tasks from language production and grammaticality judgment tests could

¹² We are aware of the typo in *vrvi* by the author, but opted to stick to their original writing.

provide new interesting insights into the boundaries and possible grey zone of such patterns: for instance, distinctions might emerge between proximal and distant punctual toponyms, or between small-scale toponyms for mountains/hills/regions and common nouns referring to geographical entities (e.g., “district”, “village”). A future study could also include other Balkan varieties where these phenomena have been attested, e.g., Bugurdži Romani, Greek, Macedonian (Stolz et al 2014: 76). The last two varieties seem to be the most promising as many native speakers of Aromanian are also proficient in Greek and/or Macedonian. To state possible contact-induced interference, fragments of speech of bilingual speakers can be compared to those of monolingual speakers. In addition, one could also include diachronic data from different Aromanian varieties, although it might be problematic to consider such texts as they mainly come from a literary tradition and are often translations. The *Tetraglosson* by Daniel Moscopolites (1794) exemplifies this issue in Aromanian, Balkan Slavic, and Greek.

Therefore, paying further attention to the patterns of zero-marking in both synchronic and diachronic varieties of the Balkan varieties can yield relevant results – not only for the comparative research of the Balkans, but also for general linguistic typology and studies of language contact.

Acknowledgment

The research for this article was supported by the Swiss National Science Foundation (SNSF) grant Nr. 203935. This support is gratefully acknowledged.

List of Abbreviations

ACC – Accusative; AOR – Aorist tense; AUX – Auxiliary verb; DAT – Dative; DEF – Definiteness; DEM – Demonstrative; F – Feminine; FUT – Future tense; GEN – Genitive; IMPF – Imperfect tense; INDF – Indefiniteness; INF – Infinitive; M – Masculine; N – Neuter; NOM – Nominative; PL – Plural; PRS – Present tense; PST – Past tense; PTCP – Participle; REFL – Reflexive; SBJV – Subjunctive; SG – Singular.

Bibliography

- Bara, M., Kahl, T., and Sobolev, A. (2005). *Die suedaromunische Mundart von Turia (Pindos): Syntax, Lexik, Ethnolinguistik, Texte*. München: Biblion.
- Caragiu-Marioțeanu, M. (1975). *Compendiu de dialectologie română:(nord și sud-dunăreană)*. București: Editura științifică și enciclopedică.
- Cunia, T. (2010). *Dictsiunar a limbaljei armanasca*. Costanța: Editura Cartea Aromana.
- Ganenkova, T. (2015). Semanticheskaya struktura predlogov v sovremennom literaturnom makedonskom yazyke. PhD dissertation. Saint Petersburg University.
- Goľab, Z. (1984). *The Arumanian dialect of Kruševo in SR Macedonia, SFR Yugoslavia*. Skopje: MANU.
- Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, 8 (3): 663–687.
- Haspelmath, M. (2019). Differential place marking and differential object marking. *STUF-Language Typology and Universals*, 72 (3): 313–334.
- Holton, D., Mackridge, P., and Philippaki-Warbuton, I. (1997). *Greek: A comprehensive grammar of the modern language*. London: Routledge. ISBN: 978-0-415-10001-4.
- Kahl, T. (2007). *Hirten in Kontakt: Sprach- und Kulturwandel ehemaliger Wanderhirten (Albanisch, Aromunisch, Griechisch)*. Vienna: LIT Verlag. ISBN: 978-3-8258-0944-7.
- Koneski, B. (1965). *Istorija na makedonskiot jazik*. Skopje: Kultura.

- Kramer, J. (1981). Griechische Konstruktionen im Aromunischen. In W. Dahmen and J. Kramer (eds.), *Balkan-Archiv* 6, 95–105. Hamburg: Helmut Buske.
- Kristophson, J. (1974). Das Lexikon Tetraglosson des Daniil Moschopolitis. *Zeitschrift für Balkanologie*, 10: 11–128.
- Luraghi, S. (2017). Differential goal marking vs. differential source marking in Ancient Greek. In S. Luraghi, T. Nikitina and C. Zanchi (eds.). *Space in diachrony*, 119–145. Amsterdam: John Benjamins.
- Markovik, M. (2007). *Aromanskiot i makedonskiot govor od ohridsko-struškiot region* (vo balkanski kontekst). Skopje: MANU.
- Sandfeld, Kristian. (1930). *Linguistique balkanique: problèmes et résultats*. Collection linguistique, 31. Paris: Champion.
- Saramandu, N. (2014). *Atlasul lingvistic al dialectului aromân*. Bucharest: Editura Academiei Române.
- Stolz, T, Lestrade, S. and Stolz, C. (2014). *The crosslinguistics of zero-marking of spatial relations*. Vol. 15. Berlin/Boston: Walter de Gruyter.
- Vidoeski, B. (1999). *Dijalektite na makedonskiot jazik*. Vol. 1. Skopje: MANU.
- Wilkinson, M., Dumontier, M., Aalsberg, I., Appleton, G., Axton, M., Baak, A., Bloomberg, N., Boiten, J.-W., da Silva Santos, L., Bourne, P. (2016). The FAIR Guiding Principles for scientific data management and stewardship”. *Scientific data*, 3: 31–

