СОВРЕМЕНА ФИЛОЛОГИЈА **JOURNAL OF CONTEMPORARY PHILOLOGY**
821.133.1.09 JCP 2025, 8 (1), 41-66.
82.0

CONTACT-INDUCED CHANGES IN THE ORDER OF MODIFYING ADJECTIVES AND NOUNS IN SLAVIC DIALECTS IN ALBANIA[1]

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

djal-i	*i*	*Mire*
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), Markoviḱ (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 (Markoviḱ 2007: 87).

ADJN > NADJ[8]

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) dialect[9] (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 Macedonian[10] (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] Vraka 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 (Bërxholi 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.

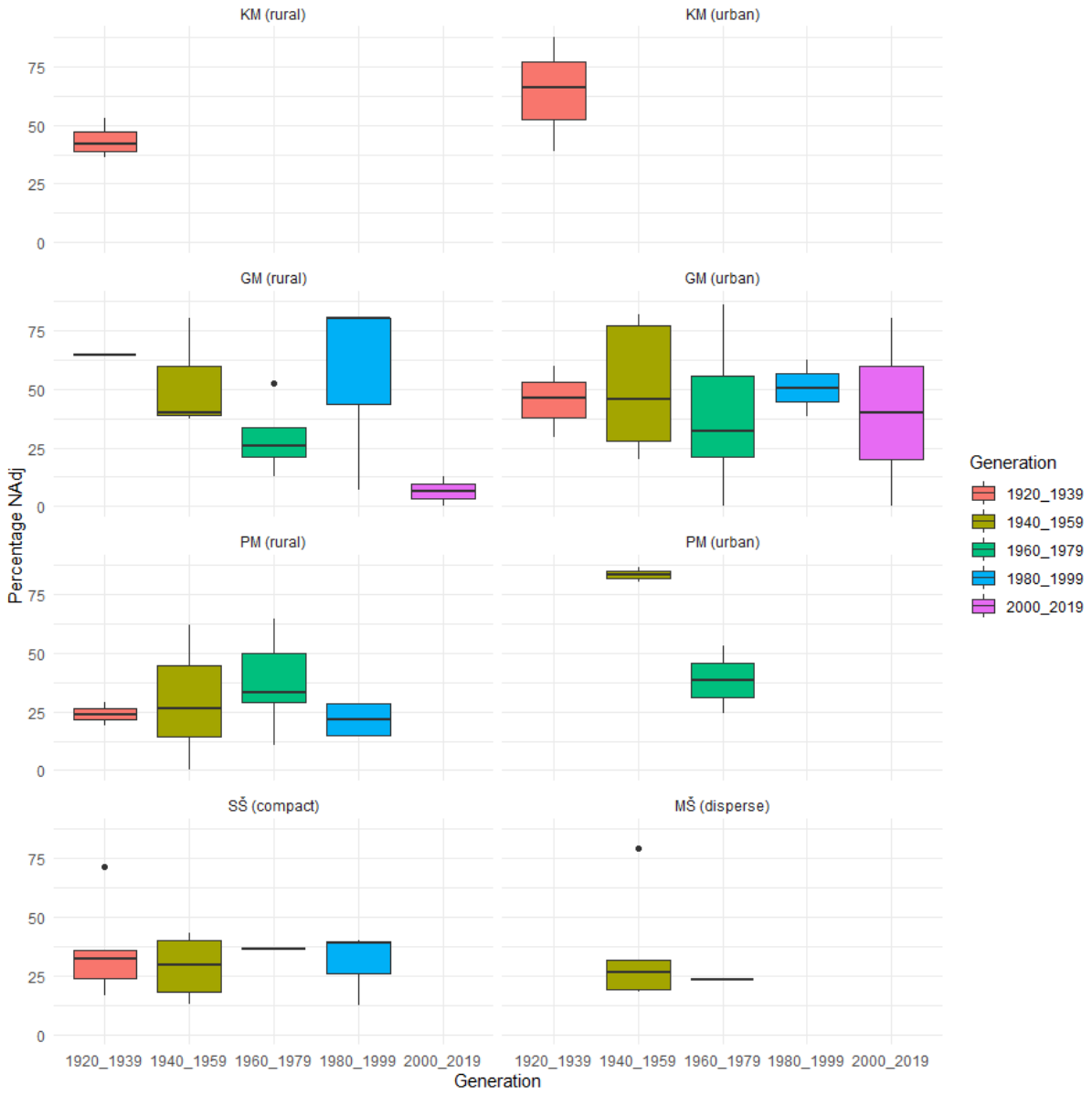


Figure 1. Distribution of NADJ percentage, speaker-wise, by generation, dialect, and residence type

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:

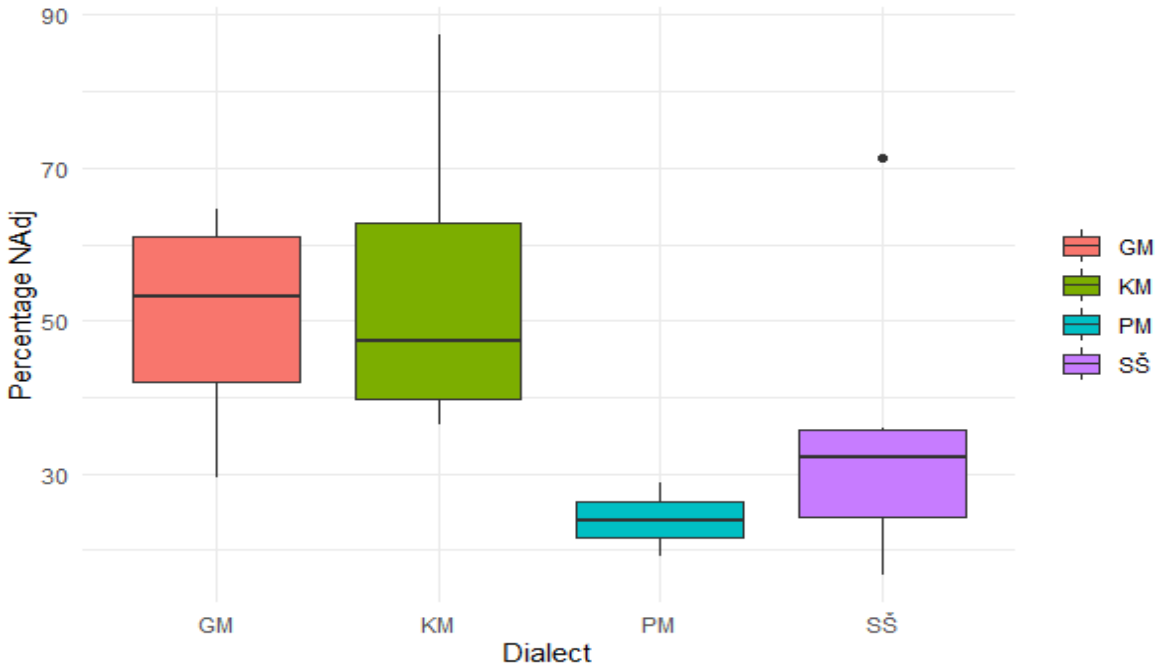


Figure 2. Distribution of NADJ percentage, speaker-wise (generation 1920–1939)

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:

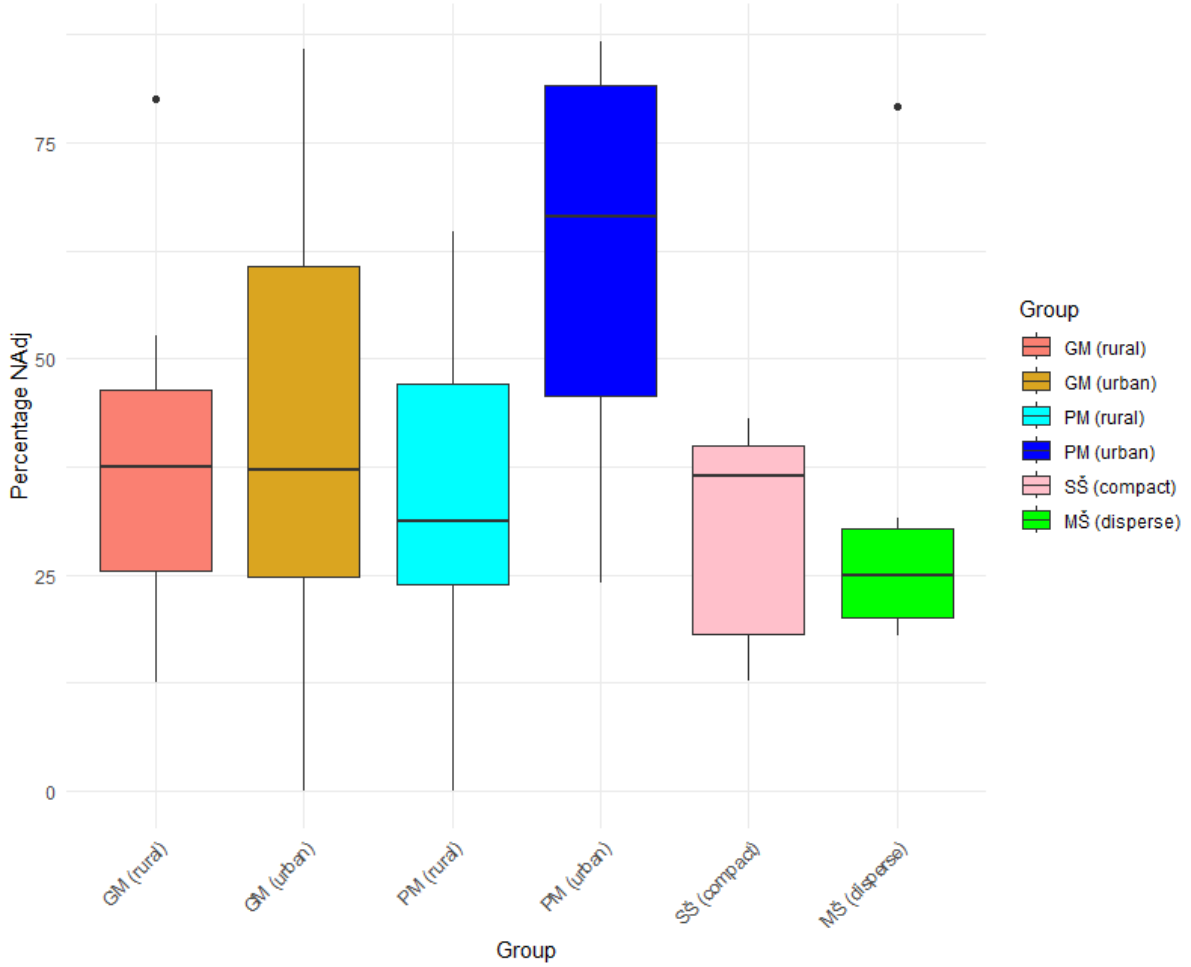


Figure 3. Distribution of NADJ percentages, speaker-wise (generations 1940–1979)

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

grouping	*n*	Mean	Median	*SD*	Min	Max	*IQR*
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-Witney text (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.

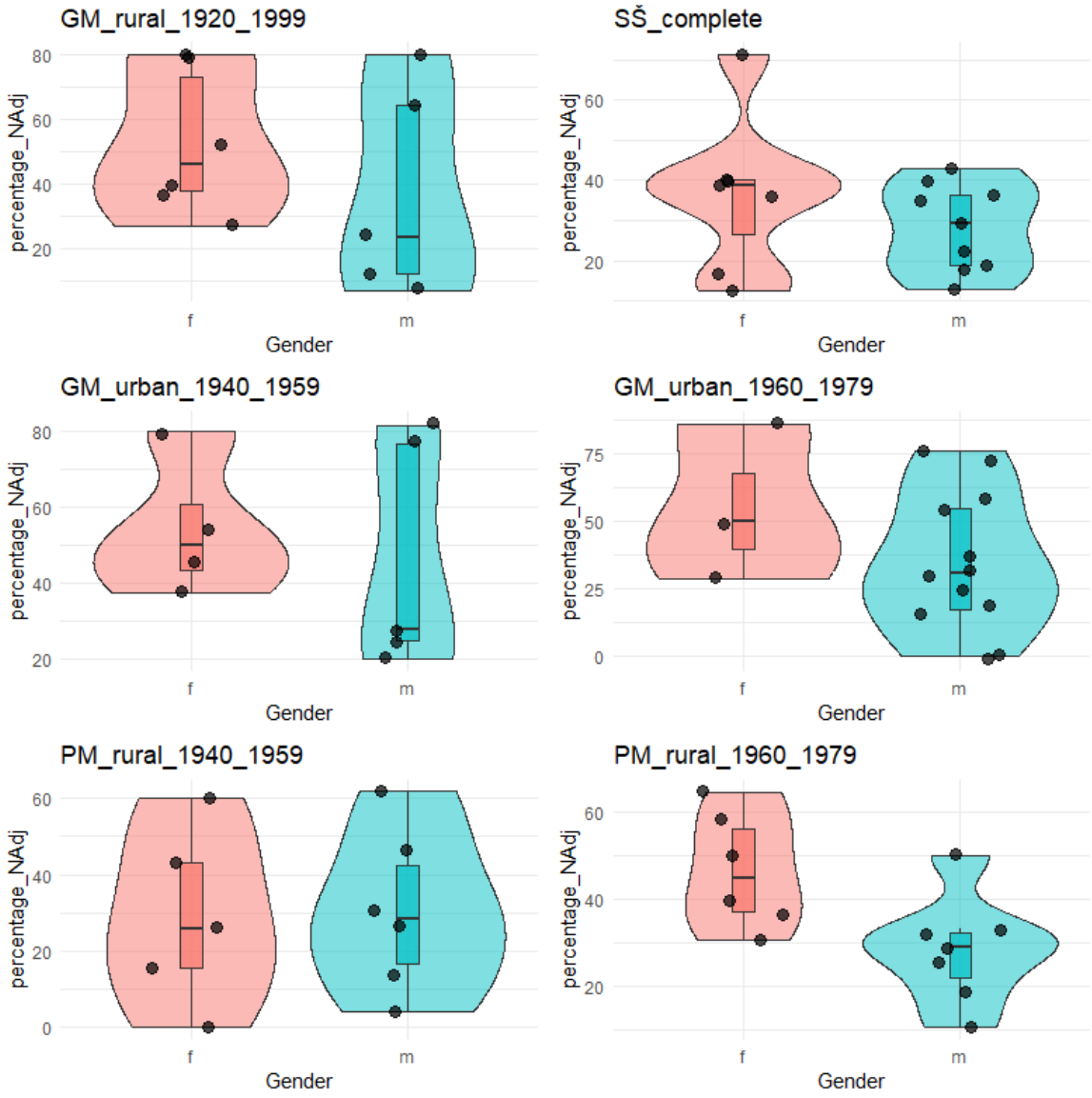


Figure 4. Distribution of NADJ percentage per gender, speaker-wise (selected groupings)

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

gender	Median	Mean	*SD*	*IQR*	Min	Max	Count	Group	*W*	*p*-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.

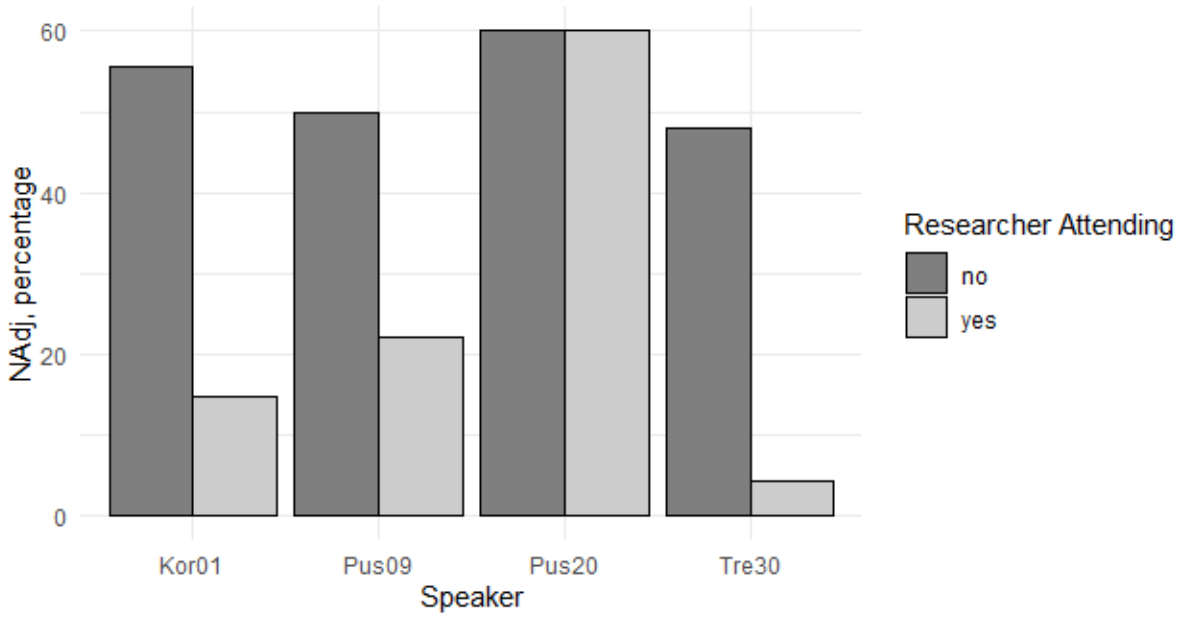


Figure 5. Percentage of NADJ speaker-wise by researcher presence

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

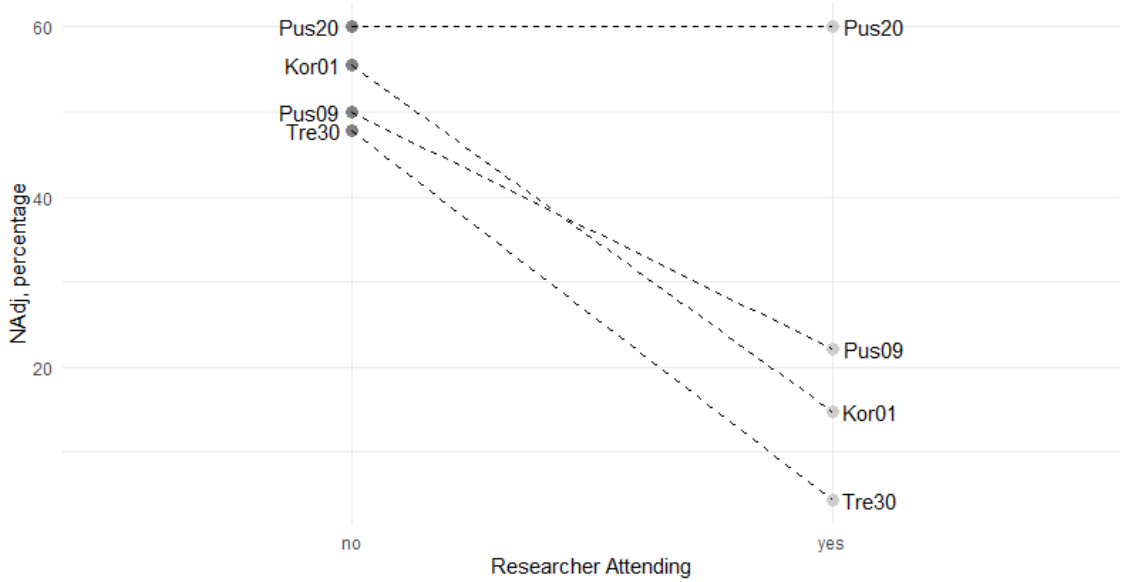


Figure 6. Speaker-wise differences in the percentage of NADJ

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:

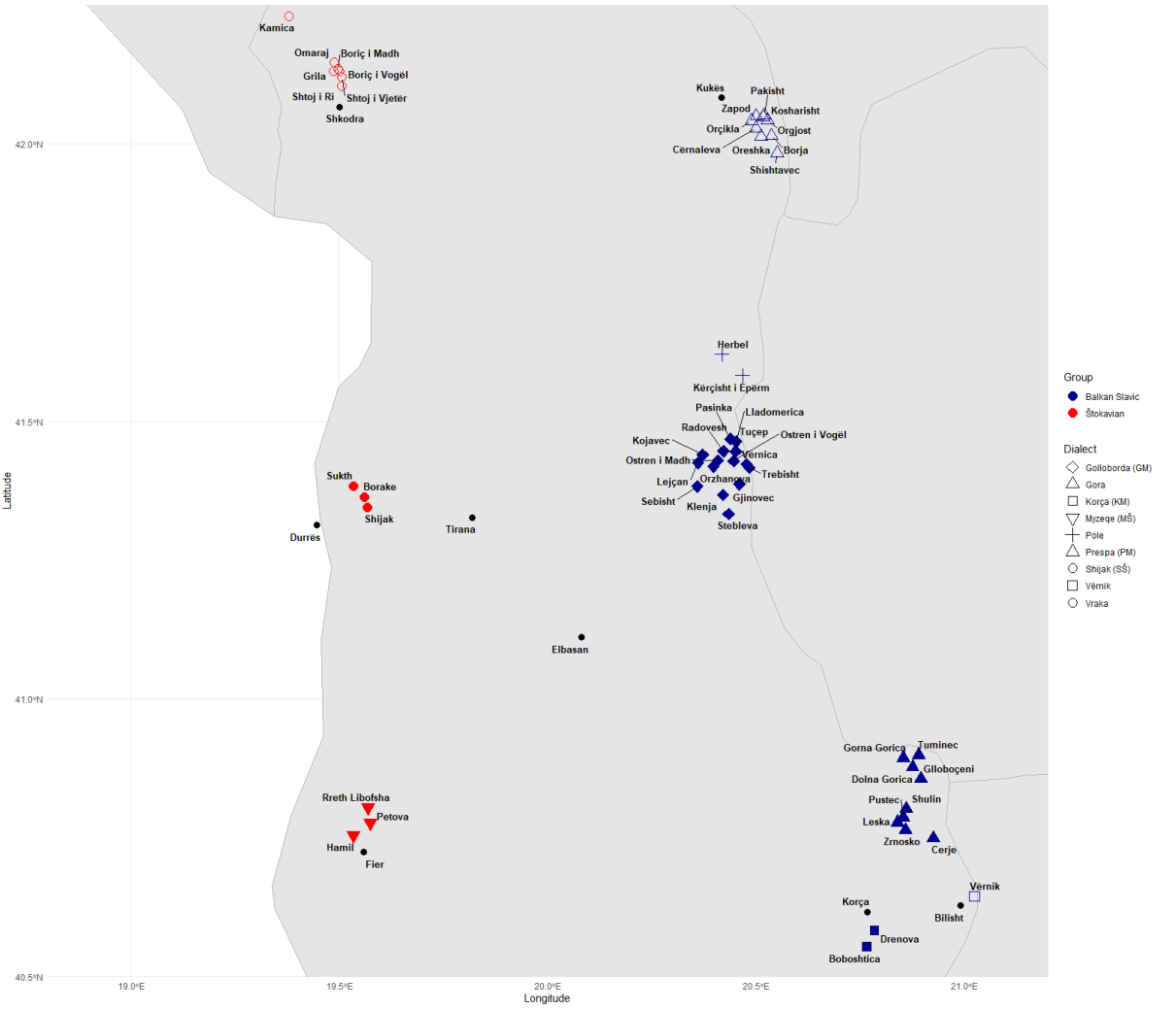
Table 6. Supplementary materials

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 Albania[22]



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 südaromunische Mundart von Turia (Pindos) = Južnoarumynskij govor sela Tur'ja (Pind): Sintaksis, leksika, ėtnolingvistika, teksty*. München: Biblion.

Bërxholi, 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 “Krste 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 Carașova 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.

Markoviḱ, M. (2007). *Aromanskiot i makedonskiot govor od ohridsko-struškiot region (vo balkanski kontekst)*. Skopje: MANU.

Markoviḱ, 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: MAĖ 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 săjuz. *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	***W***	***p*-value**	**adjusted *p***
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	***W***	***p***
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	***W***	***p***
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 senseful 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					