



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

## ARCHIVIO ISTITUZIONALE DELLA RICERCA

### Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

Under the Magnifying Glass: Dimensions of Variation in the Contemporary Timok Variety

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Vuković, T., Mirić, M., Escher, A., Ćirković, S., Miličević Petrović, M., Sobolev, A.N., et al. (2023). Under the Magnifying Glass: Dimensions of Variation in the Contemporary Timok Variety. ZEITSCHRIFT FUER SLAVISCHE PHILOLOGIE, 79, 153-194.

*Availability:*

This version is available at: <https://hdl.handle.net/11585/926299> since: 2023-05-29

*Published:*

DOI: <http://doi.org/>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Vuković, T., Mirić, M., Escher, A., Ćirković, S., Miličević Petrović, M., Sobolev, A., & Sonnenhauser, B. (2023). Under the Magnifying Glass: Dimensions of Variation in the Contemporary Timok Variety. *Zeitschrift für Slavische Philologie*, Volume 79, Issue 1. pp. 153-194.

The final published version is available online at: <https://zsph.winter-verlag.de/article/ZSPH/2023/1/9>

#### Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

# Under the magnifying glass. Dimensions of variation in the contemporary Timok variety<sup>1, 2</sup>

## 1. Variation in Transition

Torlak varieties are spoken in a geographic area where the spread of Balkan Slavic features has shaped local, genealogically West South Slavic idioms in characteristic ways. As a result, they have been recognized by dialectologists as a distinct group of dialects<sup>3</sup>. The formation of this dialect complex by the diffusion of Balkan Slavic features was facilitated by a particular configuration of political and social boundaries up to the end of the 19<sup>th</sup> century. More recently, socio-political events have been changing the region and, concomitantly, the interactional spaces and communicative habits of its residents, fostering and/or inhibiting social encounters and language contact. The most far-reaching changes have been the demarcation of political boundaries and the establishment of the Serbian and Bulgarian standard language. Both developments contributed to slowing down and eventually reversing formerly convergent processes (see Sikimić et al., this volume).

Consequently, the varieties encountered in this region can be expected to be transitional along two dimensions from a contemporary perspective: horizontally, i.e., in areal respects, by variation in the manifestation of specific structures as ‘Balkan Slavic’ or ‘West South Slavic’, and vertically, i.e., register-based, in the manifestation as ‘dialectal’ or ‘Standard Serbian/Bulgarian’. Focusing on the Serbian part of the region, the present paper aims at assessing the position of the contemporary Timok variety along the areal/horizontal and register/vertical dimensions on the basis of four representative dialect features from nominal and verbal domains: marking of indirect

---

<sup>1</sup> Funded by the SNFS (grant agreement IZRPZO\_177557/1), the Russian Foundation for Basic Research (RFBR) (No. 18-512-76002) and the Serbian Ministry of Education, Science and Technological Development (grant agreement No. 401-00-00642/2018-09), with the Horizon 2020 ERA.Net RUS Plus programme. This support is gratefully acknowledged.

<sup>2</sup> The authors’ responsibilities were as follows. Editor: Teodora Vuković, choice of features and corpus analysis, marking indirect object and possessor: Mirjana Mirić; post-positive demonstratives: Teodora Vuković; the particle *si*: Svetlana Ćirković; AUX omission: Anastasia Escher (née Makarova); statistical analyses: Teodora Vuković, Mirjana Mirić, Maja Miličević Petrović, Anastasia Escher; dialectological and Torlak expertise: Andrej Sobolev; linguistic embedding into the South/Balkan Slavic context: Barbara Sonnenhauser.

<sup>3</sup> The Torlak dialect area is spread across the territories of South-Eastern Serbia, Western Bulgaria and in the northern parts of North Macedonia. As this paper deals with data collected on Serbian territory, a remark on the Serbian dialects included in the Torlak dialect area should be made. The Prizren-Timok dialectal zone, spoken in the south-eastern parts of Serbia, belongs to Torlak and comprises five dialects, one of them being the Timok-Lužnica dialect, which is further divided into Timok and Lužnica subdialects (Belić 1905, Ivić 2009, for a more detailed division to Timok, Lužnica, Bela Palanka and Pirot subdialects see Soboljev 1994; 1995). – Regarding the socio-political circumstances, it should be mentioned that the major part of the Timok region in Eastern Serbia (with the town of Knjaževac as its administrative centre) became part of the Principality of Serbia (Serb. Kneževina Srbija) in 1833 by the third edict (hatisherif) on the autonomy of Serbia (Miličević 2012, re-print). On the other hand, the Lužnica region to the south was included only after the Congress of Berlin in 1878, while its most southern parts were officially under Bulgarian jurisdiction until 1918 (Šantić, Martinović 2007). These circumstances are responsible to a certain extent for the variation between the Timok and Lužnica subdialects.

object and possessor, post-positive demonstratives, particle usage of dative reflexive *si* and auxiliary omission in the perfect tense.

Each of these features can be realized in a ‘Balkan Slavic’ (i.e., dialectal / prototypically Timok) or ‘Serbian’ (i.e., standard Serbian) form. Measuring the usage frequencies of both realizations and their respective ratios reveals the overall degree of variation. Investigating the influence of specific linguistic factors on the respective options will demonstrate whether the choice of options is functionally conditioned; i.e., whether the distributions attest to formal and/or functional differentiations. Analyzing the effect of socio-geographic factors on the distribution of options for each feature gives insight to whether the distribution of ratios between one or the other option relates to the embedding of users in particular geographic and social contexts.

In a larger perspective, the specific case of Timok is representative for the more general challenge in dialectological and areal research: identifying and discriminating the internal and external conditions triggering variation and the features affected by these conditions. As under a magnifying glass, zooming in on this rather small region—in both socio-geographic and linguistic respects—offers insight into the intricate interaction of drivers of variation and eventual change.

The paper is structured as follows: Section 2 places our approach into the tradition of research on Torlak and introduces the corpus used for the present study. The usage frequencies of the diatopic and diastratic variants possible for four morphosyntactic features under consideration and the potential linguistic conditions underlying their distribution are identified in Section 3, while Section 4 is concerned with the impact of extra-linguistic factors. The findings are discussed in Section 5. Section 6 provides a short conclusion.

## 2. Towards a Corpus-based Approach

Previous work on Torlak has highlighted the properties that distinguish this dialect group from the rest of West South Slavic/Serbian and make it closer to Balkan Slavic, among them the features in the Timok variety discussed in the present paper (cf. in particular Belić 1905; Bogdanović 1979; Sobolev 1998). Based on a variety of fieldwork transcriptions, this tradition of research has been pursuing a primarily qualitative approach focusing mainly on the documentation of the dialect in the most prototypical instantiation possible. Accordingly, the features under consideration are presented in their purest manifestations, suggesting that they are used consistently over the whole of the Timok variety’s area. While this strand of research is highly valuable for language documentation purposes, it harbors the danger of glossing over the variation encountered in actual language usage. As revealed by recent fieldwork data, even the most typical Torlak features attested in the Timok variety are not used consistently neither across the speech community nor by single speakers. What is thus still missing are in-depth linguistic analyses and systematic assessments of variation in the region (cf. Sobolev 2020).

Capturing the variation encountered in the Timok region is relevant also beyond the purpose of documenting the current state of this variety. With this variety being located at the fringes of an alleged linguistic area, comparing the ranges of variation, their underlying linguistic

patterns and potential socio-cultural correlates with what is known from the historical developments in the Slavic varieties located at the center of the Balkan area promises insight into the mechanisms and processes facilitating and inhibiting feature diffusion. That is, an empirical corpus-based analysis of feature variation in the Timok variety of the Timok-Lužnica dialect as provided by the present study makes a highly relevant contribution to areal linguistic research.

The basis for this analysis is provided by the corpus of the Timok variety (Vukovic 2020<sup>4</sup>) created within the TraCeBa-Project (<https://traceba.net/>). This corpus encompasses approximately 500,000 tokens of spoken material collected between 2015 and 2018 in 64 villages by a team of 12 researchers in the field using semi-structured interviews on topics concerning local tradition and oral history (on the fieldwork see Ćirković 2018). The corpus includes text produced by 179 speakers: 167 interviewees (89.05% of the text) and 12 researchers (10.95% of the text)<sup>5</sup>. Most of the interviews were conducted with single speakers, 83 interviewees produced between 1,000 and 16,800 words each, making up for 86% of the entire corpus. Questions were asked in Standard Serbian with occasional attempts at dialect use aimed towards lessening the observer's paradox and, more importantly, to avoid eliciting standard forms. The responses comprise various ranges between the Timok variety and Standard Serbian (see Vuković, Escher, Sonnenhauser 2022). The selection of speakers was focused predominantly on the elderly, presumably less influenced by the standard, but it also included a number of speakers whose language clearly shows recurrent characteristics of the standard, in order to gain insight into the variation present in the Timok variety. The interviewees are thereby divided into two categories according to their linguistic production (based on the researcher's knowledge of dialect native intuition): those who are deemed to be representatives of the non-standard variety (68.37% of the entire corpus), and those closer to the standard (21.07% of the entire corpus), the rest of the corpus is the researcher's production. Importantly, most speakers exhibit some degree of standard influence, so there is no 'ideal' speaker of Timok. The transcripts included in the corpus represent the integral speech by the interviewees; i.e., they were not internally segmented according to standard vs. dialect. While that kind of segmentation would reveal potential patterns of code-mixing and code-switching and extract stretches of pure dialect, it is not relevant for the purposes of the present paper. The analysis includes all speakers irrespective of their classification.

The corpus consists of three layers: 1) accentuated speech transcripts, 2) lemmatization and 3) part-of-speech and morpho-syntactic annotation. The PoS annotation and lemmatization were done using a custom ReLDI tagger model trained for this corpus (more on the corpus in Vukovic, forthcoming; Miličević Petrović et al. in this volume).<sup>6</sup>

### 3. Facets of Variation

---

<sup>4</sup> Accessible on <http://hdl.handle.net/11356/1281>

<sup>5</sup> The speech production of the researchers has been excluded from the analysis.

<sup>6</sup> The tagger files can be downloaded from the Github repository: <https://github.com/bravethea/Torlak-ReLDI-Tagger-2019>

In order to gain insight into the morphosyntactic and socio-geographic variation characterizing the contemporary Torlak as spoken in the Timok region along the horizontal and vertical axis, four typical representative dialectal features are chosen for analysis (see Sobolev et al. in this volume for the linguistic description of typical Torlak features). These features are particularly suited for placing this variety into its diatopic and diastratic context:

- Marking of indirect object and possessor
- Post-positive demonstratives
- Dative reflexive *si* as a particle
- Auxiliary omission in the perfect tense

These features are commonly associated with the Torlak group of dialects, not only in the linguistic discourse, but also in the perception of speakers from across the BCMS dialect continuum. The particular ways of marking indirect objects and the possessors, the usage of postpositive demonstratives<sup>7</sup> and the particle *si* are regarded as shibboleths of the southeast varieties. The omission of the auxiliary in the perfect, which is also observed in Serbian vernacular varieties, is perceived as colloquial and expressive speech. The present study uses a quantitative approach to correlate the occurrences of these features with other linguistic and socio-geographic features in search of underlying usage patterns.<sup>8</sup>

### 3.1 Analysis of the linguistics factors

This section illustrates the usage frequencies of the ‘Balkan Slavic’ and ‘dialectal’ vs. ‘Serbian’ variants along with their respective ratios with the aim to identify the potential linguistic conditions underlying their distribution. In order to assess whether the respective options are influenced by specific formal triggers and/or functional differentiations, they have been examined in the context of relevant linguistic factors at the level of morphosyntax.

#### 3.1.1 Marking indirect object and possessor

The reduction of the inflectional case marking on nominals in Balkan Slavic languages has been extensively discussed in the literature as one of the most pertinent features of the Balkan area

---

<sup>7</sup> For a terminological discussion, see footnote 12.

<sup>8</sup> Documentation containing the data and the code of the quantitative analysis are available as an HTML document accessible online on RPubs (<https://rpubs.com/tevuko/636831>), with the files containing examples and the quantitative data used in the study, including the HTML document, available publicly in the GitHub repository ([https://github.com/TraCeBa/Timok\\_variation\\_Documentation](https://github.com/TraCeBa/Timok_variation_Documentation)). Note that each of the analyses presented in the paper involves different parameters, hence the segmentation of the data and the variables may differ from one feature to the other. Individual details will be presented regarding each feature in the text and the documentation. The analysis used R (R Core Team 2020) and the following R packages: ggplot2 (Wickham 2016), glm2 (Marschner 2011), vcd (Meyer et al. 2020), leaflet (Cheng et al., 2019), dplyr (Wickham et al. 2011), reshape (Wickham 2007), tidyverse (Wickham et al. 2019), tinytex (Xie 2019), ggmosaic (Jeppson et al. 2018), stats (R Core Team 2020).

(cf. Asenova 2002; Lindstedt 2000; Mišeska-Tomić 2004; Sobolev 2003; Sobolev 2008; Wahlström 2015, among others). The gradual replacement of synthetic (inflectional) marking of syntactic functions on nominals in syntactically dependent positions by analytic (prepositional) marking on the NP has been described as one of the ‘Balkan’ innovations in the Timok-Lužnica dialect, including the Timok variety (cf. Ivić 1998; 2009; Sobolev 2008). In this dialect, the analytic construction with the preposition *na* ‘on’ + general oblique case is used (see (1a) and (1b)) interchangeably with the inflectional dative case (see (1c) and (1d)) to mark indirect nominal and pronominal objects.

- (1)<sup>9</sup>
- a. Já      *na*    *njénu*      *kýrku*                      *ponesém*              *orási*.<sup>10</sup>  
 I.NOM    on    her.OBL.SG    daughter.OBL.SG              bring.1SG.PRES              walnut.M.ACC.PL  
 ‘I bring walnuts to her daughter.’
- b. *daděš*              *tí*              *na*    *njéga*                      *ón*              *na*    *tébe*  
 give.2SG.PRES    you.NOM    on    he.OBL.SG                      he.NOM              on    you.OBL.SG  
 ‘You give to him, he (gives) to you.’
- c. *i*      *u*    *poláko*    *u*    *sébe*                      *molítve*              *čítam*              *bógu*  
 and    in    slowly    in    myself.OBL    prayer.F.ACC.PL              read.1SG.PRES              God.M.DAT.SG  
 ‘And slowly to myself I read prayers to God.’
- d. *takój*              *měni*    *pričáli*  
 that way    I.DAT    read.PPART.M.PL  
 ‘(They) were telling me like that.’

In addition, the analytic strategy is also used to mark the possessor, again attested with nouns (2a), and personal pronouns (2b). This strategy is interchangeable with the use of the dative (see (2c) and (2d)).

- (2)
- a. *pomréše*              *sví*    *ostáde*                      *mi*                      *sámo*              *na*    *bráta*  
 die.3PL.AOR    all    stay.3SG.AOR                      I.DAT.CL                      only              on    brother.M.OBL.SG  
*unúk*.  
 grandson.M.NOM.SG  
 ‘Everybody died, I am only left with (my) brother’s grandson.’
- b. *na*    *méne*              *tétka*                      *umréla*  
 on    I.OBL              aunt.F.NOM.SG                      die.PPART.F.SG  
 ‘My aunt died.’
- c. *ozgór*              *de*                      *bojánu*                      *kúka*                      *bilá*  
 up there    where                      bojan.M.DAT.SG                      house.F.NOM.SG                      be.PPART.F.SG

<sup>9</sup> Examples presented in the paper are extracted from the Timok section of the Torlak corpus (Miličević Petrović et al. this Volume, Vuković, forthcoming), unless stated otherwise.

<sup>10</sup> A note on case syncretism is in order. In the Timok variety plural forms of nouns are used in only one, general case (Belić 1905), but in the glosses we made a syntactic distinction between the nominative, accusative and oblique case (e.g., *orasi* ‘walnuts’ in (1a), *molitve* ‘prayers’ in (1c)). The same holds for those singular forms which do not show distinctive morphological inflection. These are neuter nouns, which have the same form for nominative, accusative and oblique case (e.g., *jedenje* ‘food’ in (6a) and *bratanče* ‘nephew’ in (6b)).

‘Up there, where Bojan’s house was.’  
d. *méni* je májka iz marinóvac  
I.DAT AUX.3SG.PRES mother.F.NOM.SG from Marinovac.M.OBL.SG  
‘My mother is from Marinovac.’

Previous empirical research on Timok has shown a generally higher frequency of analytic marking in comparison to the use of the synthetic dative (cf. Konër, Makarova, Sobolev 2019 for the analysis of a single speech sample from the Timok variety). In addition, theoretical descriptions of nominal morphology in several Prizren-Timok varieties (Sobolev 1991) suggest that certain nominal categories are more prone to analytism than others, such that a ‘declensional hierarchy’ (Sobolev 2008, 724) has been proposed: plural inflection is less stable than singular, inflection in neuters is less stable than in feminine and masculine nouns, inanimate nouns are less stable than animate, and inflection of nouns not referring to persons is less stable than that of nouns that refer to persons.

In assessing the degree of variation between the synthetic and the analytic strategies attested in the corpus of Timok these features are taken into account, complemented by part-of-speech and function (indirect object and possessor). The analysis is thus based on the following variables:

- Dependent variable: type of marking (*na* + general oblique case vs. inflectional dative)
- Independent variables: function (indirect object, possessor), part-of-speech (nouns, personal pronouns), nominal categories: noun type (proper, common nouns), grammatical number (singular, plural), grammatical gender (masculine, feminine, neuter), animacy (animate, inanimate nouns), reference to persons (referring, non-referring).

Judging from previous descriptions, we expect a higher frequency of analytic marking in comparison to the use of the synthetic dative.

In investigating the effects of the linguistic factors on the marking strategy, we measured the absolute and relative frequencies of the dependent and the independent variables. The data were searched for in all of the corpus samples, but the relevant examples of nouns and personal pronouns used as an indirect object and possessor were found with 87 speakers from 59 different locations. As the initial corpus search provided numerous examples of words in the dative or in the construction *na* ‘on’ + general oblique case which were used in functions distinct from those of indirect object and possessor, they were excluded manually<sup>11</sup>. In addition, the examples in which a word form was homophonous with another form were manually excluded as well (e.g., accusative forms of personal pronouns, which in the Timok variety share the same form with the dative). The total number of analyzed examples taken from the corpus is 675. The features for the independent variables were marked manually for each example.

---

<sup>11</sup> The main reason for creating a large initial output was the fact that it was not possible to automatically narrow down the search only to the relevant examples. The only way we may avoid this amount of manual work in the future is by ‘improving’ the tagger and the overall automatization process, including the syntactic annotation.



In this sample, comprising nouns and personal pronouns, the use of the analytic construction prevails significantly,  $N_{\text{analytic}}=567$ ,  $N_{\text{synthetic}}=108$  (84% to 16%, Chi-square goodness-of-fit test  $\chi^2=312.12$ ,  $df=1$ ,  $p<0.001$ ). With regard to the function, the analytic construction *na* ‘on’ + general oblique case is used to mark both the indirect object (AbsFreq=356, RelFreq=52.74%) and possessor (AbsFreq=211, RelFreq=31.26%), while the dative case is mostly used to mark the indirect object (AbsFreq=91, RelFreq=13.48%), while the possessor is marked by the dative to a lesser extent (AbsFreq=17, RelFreq=2.52%); see Figure 1 for a graphical display of the distribution.

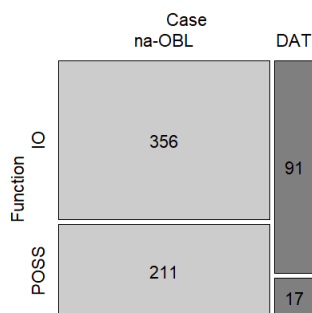


Figure 1: Type of marking: overall and per functions (absolute frequencies)

Regarding the parts-of-speech in the overall sample (547 nouns, 128 personal pronouns), the distribution of the marking strategy is calculated with regard to the overall number of examples (i.e., dative contexts) analyzed. The analytic strategy prevails with nouns (76.44% vs. 4.59%), while personal pronouns are slightly more used in the dative (11.41% vs. 7.56%), see Figure 2 for absolute frequencies. In marking the indirect object, nouns are predominantly used in the analytic construction in comparison to the dative (71.81% vs. 6.71%), while personal pronouns are used slightly more in the dative (13.65%) than in the analytic construction (7.83%; see Figure 3 for absolute frequencies). In marking the possessor, nouns are predominantly used in the analytic construction (85.53% vs. 0.44%), while personal pronouns are used both in analytic construction (7.02%) and in dative (7.02%), see Figure 3 for absolute frequencies.

If the synthetic strategy is used at all, it is strongly preferred with personal pronouns. To check if the impact of the studied factors on the choice of marking strategy is significant, a logistic regression was conducted, taking into account function, part of speech and their interaction (see Table 1); the more frequent variable values were selected as baseline – analytic marking, indirect object and nouns. The model showed a statistically significant effect on the choice of the marking strategy of both function and part of speech, as well as their interaction (see Table 1): the dative case is overall less likely to mark the possessor, but this changes for pronouns. They are almost 19 times more likely than nouns to occur with the synthetic dative (10.5 times more likely when marking a possessor).

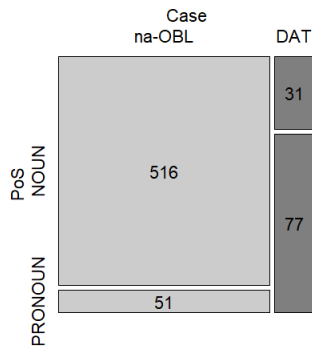


Figure 2: Case marking in nouns and personal pronouns (absolute frequencies)

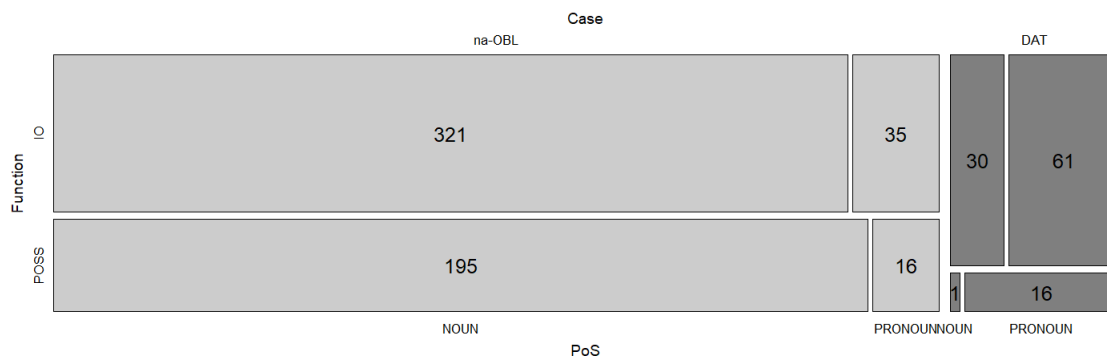


Figure 3: Case marking in IO and POSS function with respect to PoS (absolute frequencies)

	<b>B (SE)</b>	<b>Z value</b>	<b>Odds ratio</b>	<b>P-value</b>
Intercept	-2.37 (0.19) ***	-12.415		p<0.001
Function (POSS)	-2.90 (1.02) **	-2.846	0.05	p<0.01
PoS (Pronoun)	2.93 (0.29) ***	10.254	18.65	p<0.001
Function (POSS) * PoS (Pronoun)	2.35 (1.10) *	2.134	10.46	p<0.05

Table 1: The effects of function and PoS on the marking strategy

As for the nominal categories, see Figure 4 for absolute frequencies. The statistical analysis (logistic regression) was performed in order to investigate the effect of nominal categories on the choice of marking strategy, as well as their interaction. The more frequent variable values were selected as baseline: analytic marking, common nouns, feminine nouns, singular nouns, nouns referring to persons. The analysis did not show that the type of noun had a statistically significant effect. Both proper and common nouns prefer the analytic construction over the synthetic expression (27.24% and 67.09% vs. 1.1% and 4.57%). The analysis showed that gender had a

statistically significant effect: although masculine, feminine and neuter nouns are all more frequently used with the analytic construction than the synthetic expression (M: 33.46% vs. 4.02%, F: 54.84% vs. 1.65%, N: 6.03% vs 0%), the masculine nouns are more likely to occur in the synthetic dative in comparison to feminine nouns (9 times more likely, see odds ratio in Table 2)<sup>12</sup> and the neuter nouns do not appear in the dative at all. Regarding number, the analytic construction dominates in both singular (83.36% vs 4.57%) and plural nouns (10.97% vs 1.1%). The effect of number is also statistically significant, as plural nouns are more likely to be expressed synthetically than singular nouns (24 times more, cf. odds ratio in Table 2). Additionally, the statistical analysis showed significant interaction of gender and number, indicating that masculine plural nouns are more likely to be expressed analytically, which is the effect opposite of the one obtained by observing the gender and number categories separately. With animacy, the analytic construction prevails in both animate (93.42% vs. 5.48%) and inanimate nouns (0.91% vs 0.18%), but due to the low number of examples of inanimate nouns, no statistical analysis was performed. The statistical analysis did not show a statistically significant effect of the reference to person, as the analytic construction prevails in both nouns referring to persons (86.11% vs. 5.48%) and those not referring to persons (8.23% vs 0.18%).

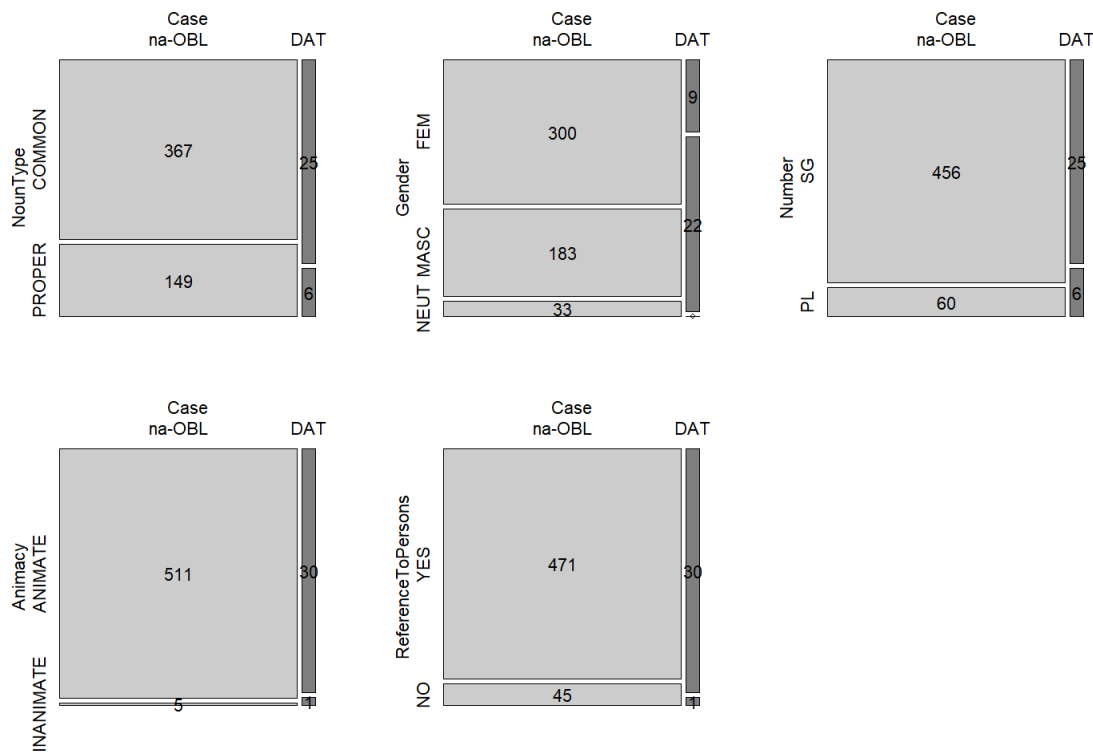


Figure 4: Case marking with regard to nominal categories

<sup>12</sup> It should be noted that in the sample of the nouns used in the dative, 13 out of 32 are the lexicalized dative form of the masculine noun *bog* ('God') in the constructions such as *moliti se bogu* 'pray to God'.

	<b>B (SE)</b>	<b>Z value</b>	<b>Odds ratio</b>	<b>P-value</b>
Intercept	-4.04 (0.58) ***	-6.932		p<0.001
GenderMASC	2.24 (0.64) ***	3.477	9.44	p<0.001
NumberPL	3.19 (0.9) ***	3.533	24.28	p<0.001
GenderMASC : NumberPL	-3.67 (1.12) **	-3.265	0.025	p<0.01

Table 2: The significant effects of nominal categories on marking strategy (logistic regression)

Summing up, regardless of the differences in the applied methodology (such as sample size), the corpus-based analysis confirms previous findings for older data from the early 1990s (Sobolev 1991; Konër, Makarova, Sobolev 2019) concerning the general prevalence of analytic vs. synthetic marking in the Timok variety. On a more detailed level, the analysis revealed that variation is influenced by part-of-speech (synthetic strategy is more likely to occur with personal pronouns), grammatical gender of nouns (synthetic strategy is more likely to occur with masculine nouns, while it is not used with neuter nouns and to a small degree with feminine nouns), and grammatical number of nouns (synthetic strategy more likely to occur with plural nouns). In addition, the analysis revealed a connection between the type of case marking and the function, in the sense that dative case marking is highly infrequent for possessors, whereas analytic marking appears for both possessor and indirect object. Regarding the declensional hierarchy (Sobolev 1991; 2008), results show higher percentages of analytic marking in all nominal categories. More specifically, when it comes to gender, the analytic strategy is overall preferred, for neuter nouns the analytic strategy is the only option, and masculine nouns are more likely to occur with synthetic marking than feminine nouns. For number, the analytic strategy is overall preferred, but plural nouns are more likely to occur with synthetic marking than singular nouns. For animacy, only animate nouns exhibit (small) numbers of dative marking. For type of noun and reference to persons, analytic marking is overall preferred. Given that the declensional hierarchy was proposed for data from the late 20<sup>th</sup> century, this suggests a leveling towards the analytic strategy and the Timok variety seems to follow Balkan Slavic development (see Asenova 2002; Mišeska-Tomić 2004; Sobolev 2003, among others). However, further research is required, especially keeping in mind that the hierarchy refers not only to the indirect object and possessor, but to other functions as well.

### 3.1.2 Post-positive Demonstratives

Post-positive demonstratives<sup>13</sup> (PPD) are a shared feature of the Balkan Slavic varieties. They are part of the Bulgarian and Macedonian standard and also used in the Timok variety of Torlak (Joseph 1992; Belić 1905; Ivić 1985). Etymologically, they originate from demonstrative pronouns, but while demonstrative pronouns precede the noun and are accentuated (3a), PPDs are un-accentuated and attach to the right of their host (3b). They both agree with the noun in grammatical gender, number and case (3b, 3c, 3d).

- |     |    |                             |                |    |                                 |
|-----|----|-----------------------------|----------------|----|---------------------------------|
| (3) | a. | tá                          | žená           | b. | žená-ta                         |
|     |    | that.F.SG.NOM               | woman.F.SG.NOM |    | woman.F.SG.NOM-<br>DEM.F.SG.NOM |
|     |    | ‘that woman’                |                |    | ‘the woman’                     |
|     | c. | ženú-tu                     |                | d. | žené-te                         |
|     |    | woman.F.SG.ACC-DEM.F.SG.ACC |                |    | woman.F.PL-DEM.F.PL             |
|     |    | ‘the woman’                 |                |    | ‘the women’                     |

Just as the demonstrative pronouns from which they derive, post-positive demonstratives make a three-way distance distinction indicated by the *-v*, *-t*, *-n* stems, resulting in forms like *žena-va*, ‘this/the woman’, *žena-ta*, ‘that/the woman’, *žena-na*, ‘that over there/the woman’. They adhere to grammatical agreement, in most cases phonetically, based on the noun’s ending<sup>14</sup>. Typically, masculine nominals ending in a consonant take a nominative singular PPD *-at/ət*, the form for nouns of grammatical feminine gender ending in *-a* is *-ta*, while neuter nouns show the form *-to* (the forms for the other two demonstrative stems follow the same morphological and agreement pattern; the paradigm of PPD involves a small degree of phonological variation). In the noun phrase with modifiers to the left, PPDs appear on the left-most modifier of the head noun, see example 4.

- |     |    |                          |                |
|-----|----|--------------------------|----------------|
| (4) | a. | moja-tá                  | žená           |
|     |    | my.F.SG.NOM-DEM.F.SG.NOM | woman.F.SG.NOM |
|     |    | ‘that woman’             |                |

<sup>13</sup> We choose to use the term post-positive demonstratives instead of the term articles, which is commonly used for the analogous category in Bulgarian and Macedonian for several reasons. This category is largely underinvestigated in Timok, current descriptions are rather vague, stating that in Timok, it is being used with mixed meanings of demonstratives and articles (Belić 1905; Ivić 1985), while Tomić 2006 gives only a few sketchy examples, which give an incomplete description. These do not give enough reason to assume that the meaning and use of these morphemes is the same as in Bulgarian and Macedonian just because they are morphologically similar. Therefore, in order to actually empirically begin to prove this functional similarity and to find the actual status of this category in Timok, we choose to be agnostic and avoid presuming that it acts as an article. The analysis in this article, and some other ones (Vuković, forthcoming) deals with this question precisely taking into account different linguistic aspects.

<sup>14</sup> For instance, in Torlak, certain nouns denoting masculine gender, such as *deda*, ‘grandfather’, are of feminine gender, ending in *-a*, a typical feminine marker. Such nouns have a phonetic agreement following their phonetic form and not semantic agreement, based on their meaning.

Note that there is variation in the plural forms across the Torlak dialect area, and some of them display the absence of phonetic agreement. Forms such as *voloveti*, *volovete*, *volovite*, *voloveto* ‘the cattle’ each have a distinct areal distribution (cf. Sobolev 1998).



the nominal) in nominals hosting the PPD can be used as an indicator of its status with respect to the grammaticalization from adnominal demonstrative pronouns into definite articles.

In order to identify the distribution of different forms of PPD (nominative/unmarked vs. accusative/oblique, as well the distribution of masculine, feminine and neuter gender), nouns containing PPD were compared against bare nouns<sup>17</sup>. The comparison regarding gender includes all nouns without modifiers<sup>18</sup>, while the comparison concerning case takes into account only feminine nouns ending in *-a* and masculine animate nouns ending in a consonant (regardless of the syntactic position). The following variables were used:

- Dependent variable: PPD status (+PPD or -PPD)
- Independent variables: gender (masculine ending in consonant, feminine ending in *-a*, neuter), case (nominative/unmarked and oblique/accusative singular)

The data were extracted semi-automatically. Since the analysis focuses only on certain nouns, all nouns were first extracted using PoS tags and word endings, and lemmas for nouns of each gender were extracted manually (see Documentation referenced in Footnote 8 for the lists of lemmas). Further searches were based on those lemmas. In order to inspect the distribution of case, the search retrieved masculine singular animate nouns ending in consonant/*-at* in the nominative/unmarked and *-a/-atoga* for the oblique/accusative, and feminine singular nouns ending in *-a/-ata* in the nominative and in *u/-utu* in the oblique/accusative (and analogous *-v* and *-n* forms).

Our data reveal that PPDs are generally rare in the corpus: only 1.55% of all nouns in the corpus carry one (AbsFreq<sub>bare</sub>=78565, AbsFreq<sub>N+PPD</sub>=1225). The results obtained for bare nouns show that masculine and feminine nouns appear with similar frequency, while neuter nouns are less frequent. Nouns hosting PPDs show a different distribution, feminine nouns appear more frequently with PPDs than masculine nouns, while neuter nouns are as frequent as in case of bare nouns (Figure 5). Logistic regression revealed a significant effect of gender on the likelihood of a PPD being used (B(SE)=0.50(0.06) for feminine, and (B(SE)=0.44(0.09) for neuter nouns, compared to masculine nouns as baseline, both significant at the  $p < 0.001$  level). All three genders are overall more likely to occur without PPD, but the likelihood is lower for feminine and neuter than for masculine nouns; feminine nouns are 1.64 times more likely to carry a PPD than masculine nouns.

---

<sup>17</sup> Only nouns were taken into account, for several reasons. They occur most frequently with PPDs than other part-of-speech categories, and therefore provide more material. The morphological properties attested in nouns are expected to recur across the other categories that can host a PPD. Categories such as personal pronouns, which are known to have a different case paradigm are not affected by PPDs.

<sup>18</sup> There are inconsistencies with how PPDs are distributed on noun modifiers. PPDs appearing on a modifier are rare: out of 1313 occurrences of NPs with PPD, only 52 instances have a PPD on a left modifier.

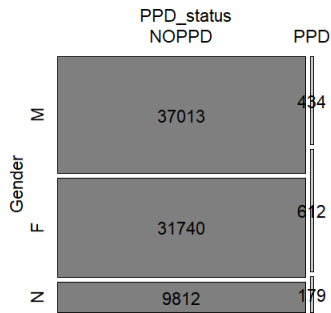


Figure 5: PPD and gender of nouns

The data reveals that nouns hosting a PPD take oblique/accusative case inflection less often than bare nouns, as show in Figure 6. A closer look at masculine and feminine nouns reveals that feminine nouns appear with PPDs more frequently than masculine nouns. However, for this subsample of feminine and masculine nouns, a logistic regression showed that there is no significant effect of either gender or case.

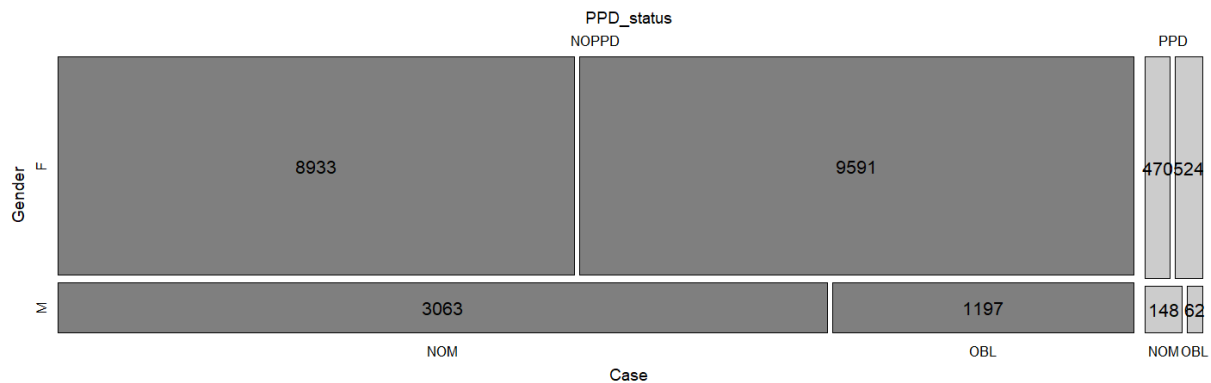


Figure 6: Proportions of nominative/unmarked and oblique/accusative case forms in nouns with and without PPD<sup>19</sup>

The results show that PPDs tends to attach more often to feminine and neuter nouns than to masculines. Although there is a slight difference in case marking between bare nouns and those carrying a PPD, we have not found statistically significant evidence for a diachronic process of change where inflectional markings are gradually lost in the presence of PPD. It must be noted that such tendencies indeed might not be due to the PPD, since feminine nouns seem to appear more frequently with accusative/oblique case marking than masculine nouns. The inflectional properties of nouns containing a PPD are affected also by their syntactic function. Nouns containing a PPD might appear more frequently in positions which require nominative marking

<sup>19</sup> The mosaic plot indicates only the difference between different case forms. The vertical line divides the field in half and does not indicate the proportion between nouns with and without PPD.



(anaphoric/given referents might prefer the topic/subject position). However, at present it is not possible to check for this as the corpus does not have syntactic annotation. The results indicate that the Timok variety displays similar characteristics to the stages of the emerging article in Bulgarian, although such an analogy would require a diachronic analysis which is beyond the scope of the present paper.

### 3.1.3 Particle *si*

The form *si* is morphologically a dative clitic form of the reflexive pronoun, and as such can be used either as an indirect object reflexive (7a) or as a possessive (7b). Dialectologists note that in the Timok variety and the wider Prizren-Timok dialectal zone, the enclitic form of the pronoun *si* may be delexicalized and develop into a verbal particle in addition to its inherent meaning, see (7c) (Ćirić 1983; Bogdanović 1979; Ćirić 1999; Jurišić 2009; Pavlović 1970). It is common in Balkan Slavic<sup>20</sup> (Petrova, 2014; Savova, 2017), its usage in Standard Serbian is not considered by prescriptive grammars (for example, it is not mentioned in Klajn and Piper’s (2013) grammar).

- (7) a. Ponéli                    smo                    *si*                    jédenje.  
bring.PPART.M.PL    AUX.1PL.PRES    REFL.DAT.CL    food.N.ACC.SG  
‘We brought food for us.’
- b. Ság    *si*                    žalim                    brátanče.  
now    REFL.DAT.CL                    mourn.1SG.PRES                    nephew.N.ACC.SG  
‘Now I am mourning after my nephew.’
- c. Ón        *si*                    dójde                    pa    *si*                    me                    čúva.  
he.NOM    REFL.DAT.CL    come.3SG.PRES    and    REFL.DAT.CL    I.ACC.CL    guard.3SG.PRES  
‘He comes and takes care of me.’

While the Serbian dialectological literature on the Prizren-Timok dialectal zone registers *si* as part of the paradigm of the pronoun, its use as a particle is in the most part not clearly defined. It is described in a vague sense as having a “dative meaning” (Ćirić 1983; Bogdanović 1979; Ćirić 1999; Pavlović 1970) and a “pronominal meaning” (Vukadinović 1996; Bogdanović 1987), as being “pleonastic” (Stanojević 1911; Ćirić 1983), as “euphemistic use of the first person” (Ćirić 1983), or simply as a “common expression” (Vukadinović 1996). In the general linguistic literature, the function of this particle has been described as evaluating the eventuality indicated by the verb in the predicate (see Arsenijević 2013 and Milosavljević 2019, who both refer to it as ‘evaluative dative reflexive’). In addition, it is also analysed as a middle voice marker (Vuković, Escher, Sonnenhauser, 2022), highlighting its function to indicate that the action is being made in the interests of the subject, whereby the opposition between the agent and patient is neutralized

---

<sup>20</sup> It is used in Croatian as well, but since here we only deal with distinctions between typically Serbian and Balkan Slavic feature realizations, we do not take this fact into consideration.

(Kemmer 1993). The present analysis focuses on structural properties, leaving in-depth semantic investigations for further research.

In order to assess the non-pronominal use of *si* in the Timok corpus and thereby arrive at a clearer formal description, the most frequent syntactic patterns of *si* used as a particle are identified, taking into account linguistic factors such as the properties of the verb, position of the particle *si* in relation to the verb and other constituents in the clause. In addition, the most frequent verbs in their context are registered in order to identify the eventualities marked with the clitic *si*. The following variables are included into the analysis:

- Subsample for analysis: the analysis included a subsample of all examples containing the particle *si*<sup>21</sup>
- Independent variables: person, number of verb, animacy, reflexivity, voice. Although not treated as a separate variable, lexical type variation in the syntactic patterns in the contact position between *si* and the verb is also analysed descriptively.

A total of 1,364 relevant examples of the use of *si* were extracted from the corpus. The selection was first done automatically and further filtered manually.<sup>22</sup> The features for the independent variables were marked manually for each example. The variables analyzed mainly point to relevant morpho-syntactic properties related to the linguistic trait under scrutiny. At the same time, it is possible to search for them in the corpus by using their formal surface expression. This is not the case for some other potentially relevant, though implicit properties, such as semantic or pragmatic traits. Other variables, such as agentivity, aktionsart, among others which could be included in the analysis, fall within the domain of semantics, which was not the focus of this particular analysis.

The statistical analysis (Chi-square goodness-of-fit test) showed a significant difference between different person and number combinations ( $\chi^2=957.65$ ,  $df=5$ ,  $p < 0.0001$ ). The highest frequency is observed with the 3<sup>rd</sup> person singular (AbsFreq=600; RelFreq=43.99%), the least frequent is the 2<sup>nd</sup> person plural (AbsFreq=13; RelFreq=0.95%), see Figure 7 for the full distribution.

---

<sup>21</sup> The analysis involved only the examples containing the particle *si* instead of all verbal contexts because the corpus does not contain explicit information on the features analysed. In the quantitative framework, this is not a dependent variable, but rather a sub-sample where other parameters are observed. Since the corpus did not contain an annotation on the verbal categories included in the analysis, except for number, the data would have to be annotated manually. The manual processing and labelling of all verbs for the features analyzed would be too difficult and time-consuming.

<sup>22</sup> See footnote 8.

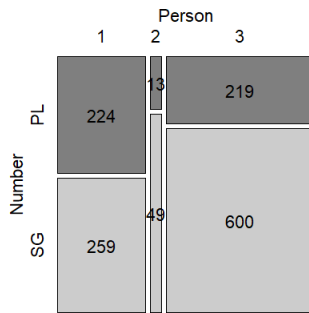


Figure 7: Frequency of the particle *si*: Person and number of the verb (absolute frequencies)

The particle *si* is used significantly more with animate than inanimate subjects, with non-reflexive than reflexive verbs, and with active than passive verbs, as shown by means of Chi-square goodness-of-fit tests. The number of examples involving an animate subject (whether overt or covert) is 1,152 (84.46%), while inanimate subjects are used in 212 examples (15.54%) ( $\chi^2=647.8$ ,  $df=1$ ,  $p<0.001$ ). Furthermore, the particle *si* is used more often with non-reflexive verbs and constructions<sup>23</sup> (91.2%) ( $\chi^2=926.23$ ,  $df=1$ ,  $p<0.001$ ), as well as the active verb forms<sup>24</sup> (96.04%) ( $\chi^2=1156.6$ ,  $df=1$ ,  $p<0.001$ ), see Figure 8 for absolute frequencies.

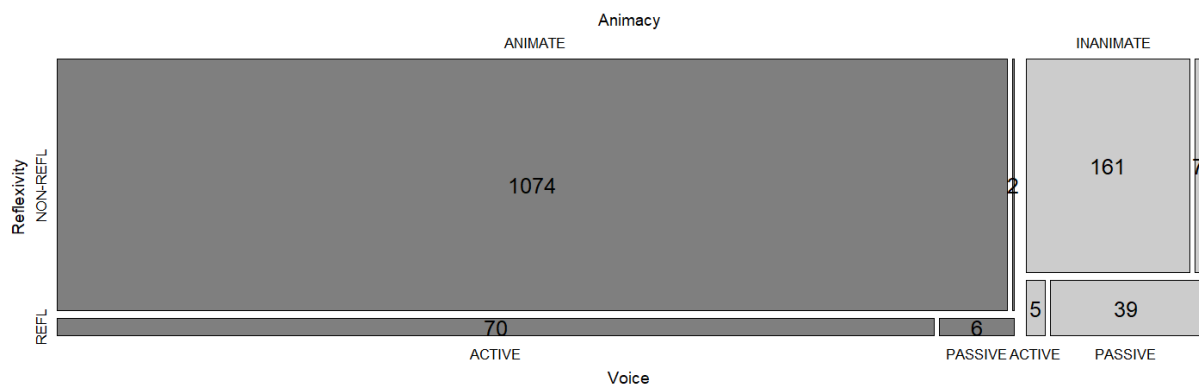


Figure 8: Frequency of the particle *si*: Animacy, reflexivity, voice

Concerning the word order patterns, in 927 (67.96%) cases out of 1,364 examples, the particle *si* is positioned next to a verb. The absolute frequencies of each contact pattern between the verb and the particle *si* are given in Figure 9. The most frequent among them is the direct juxtaposition

<sup>23</sup> Reflexive verbs and constructions are those containing the reflexive pronoun *se*, i.e., lexical verbs containing *se* and the reflexive passive.

<sup>24</sup> Passive forms include passive constructions with the passive participle and the reflexive passive. All the rest are taken as active, including existential sentences with the verbs *imati* ‘have’ and *biti* ‘be’ as they are used in the active voice. As for adjectival participles, they have been observed as an adjectival part of the predicative and the sentences they were used in were therefore treated as active, regardless of their morphological form, which would suggest that the sentence is passive.

between the particle and the lexical verb, the inflected form or past participle (891 examples of *si* + V / PPART or V / PPART + *si*).

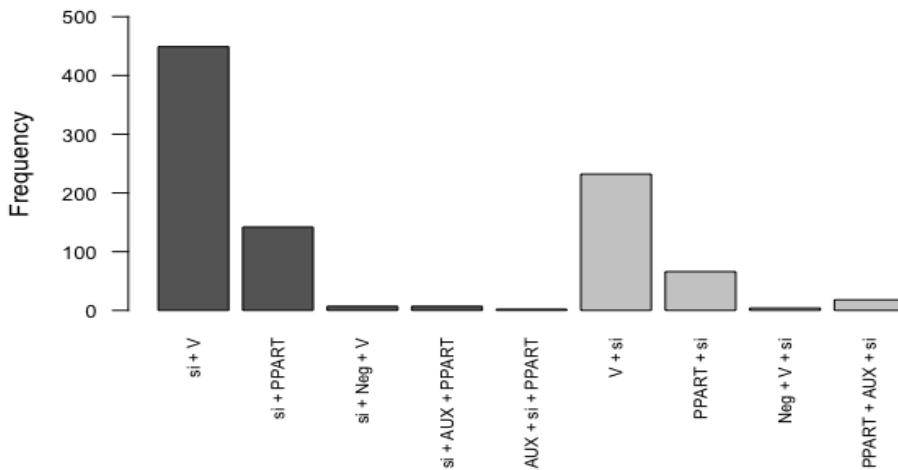


Figure 9: The proportions of contact patterns between *si* and the verb

The particle *si* behaves like a clausal Wackernagel element, always taking the second position. At the same time, it most frequently occurs next to the verb. The placement of *si* in front of the verb can be conditioned by the presence of the subject, adverbial, conjunction, a subordinate clause or another element in the clause-initial position, cf. (8a). The syntactic pattern where the verb precedes the particle *si* (V + *si*, PPART (AUX) + *si*, NEG V + *si*) occurs when the subject is omitted or located in another, non-initial position in the clause and when there are no other constituents in the clause-initial position (8b).<sup>25</sup>

- (8) a. Ón si čéka.  
 he.NOM PART wait.3SG.PRES  
 ‘He is waiting.’
- b. Sédi si onám.  
 sit.3SG.PRES PART there  
 ‘He is sitting there.’

Using the automatic lemmatization of the verbs occurring with the particle *si* revealed that it is used with 329 different verb lemmas. The most frequent are the copulative verb *biti*<sup>26</sup> ‘to be’ (150 examples, 10.99%), and the verb *imati* ‘to have’ with an existential and possessive meaning (109

<sup>25</sup> Although very rare, there are 10 occurrences in the corpus of patterns where V + *si* or PPART + *si* are preceded by a clause-initial pronoun, functioning as subject or object. There exist also rare cases when the pair is preceded by an adverb or a conjunction.

<sup>26</sup> In general, the Timok dialect does not have infinitive forms, however they are kept in the lemmatization for practical purposes. On how lemmatization was performed, see Miličević Petrović et al. in this volume.

examples, 7.99%). Among lexical verbs, the most frequent ones are the motion verb *ići* ‘to go’ and its derivatives *otići* ‘to go away’ and *doći* ‘to come’ (204 examples, 14.95%)<sup>27</sup>.

In 10.99% of the cases, *si* is used in a clause with a copulative predicate (in present tense forms 65.33%, in perfect tense forms 32.67%, rarely in other verbal forms). In some Timok verbal constructions, the copula, as well as the auxiliary, can be omitted (for AUX omission see Section 3.1.4). By analyzing the examples with *si*, we observe no pattern in the omission due to the presence of *si*.

To sum up, the particle *si* in Timok tends to occupy the second position in the clause and occurs most frequently juxtaposed to the verb. The interaction of these two placement preferences makes this phenomenon another feature showing the ‘transitional’ character of Timok between the Bulgarian and the Serbian systems (see Pancheva 2005 for the diachronic development in Bulgarian from 2nd position to verb adjacency). The position of *si* relative to the verb is determined by the presence of another constituent in the clause-initial position. The question whether the *si* + VERB cluster is preceded by another constituent requires further investigation from the perspective of information structure s (cf. Halupka-Rešetar 2009; Halupka-Rešetar 2011; Féry, Krifka 2008).

When it comes to the morpho-syntactic and lexical properties, the distribution of *si* in the corpus shows that it tends to occur with non-reflexive verbs in the active voice and with animate subjects. It co-occurs most often with the verbs in the 3<sup>rd</sup> person singular. It is used frequently with the copulative verb ‘to be’, ‘to have’ used existentially, as well as with motion verbs. Some of these tendencies might be due to the content of the interviews (e.g., speakers tend to refer to the third persons rather than to the addressee).

### 3.1.4 Auxiliary Omission in the Perfect Tense

The omission of the 3<sup>rd</sup> person auxiliary in the perfect tense is characteristic of the Balkan Slavic verb system. However, the functional properties of this omission vary across dialects and the corresponding standards. In the Macedonian standard, auxiliaries are always omitted in the third person singular and plural, but this does not carry any functional or semantic load (Friedman 2002). In Bulgarian dialects, auxiliary omission is often associated with the category of evidentiality, that is the ‘short’ forms of the perfect (Stankov 1967 speaks of a paradigmatic opposition between ‘long’ and ‘short’ perfect forms) are assumed to display several functions related to epistemic distancing, e.g., the indication of a change in perspective (Meerman and Sonnenhauser 2017, 85–86).

The Timok corpus provides many instances of the ‘short’ forms, see (9):

---

<sup>27</sup> Other verbs, such as *pričati* ‘to speak’, *znati* ‘to know’, *živeti* ‘to live’, *sedeti* ‘to sit’ occur between 22 and 31 times. The rest of the verbs with a lower frequency belong to verbs signifying actions such as *raditi* ‘to work’, *jesti* ‘to eat’, *čuvati* ‘to keep’, etc. It is important to note that out of the verbs mentioned in relation with the particle *si*, the verbs *biti*, *imati*, *ići*, *znati* are the four most frequent in the corpus, the verb *doći* takes the eighth place while the rest of the verbs mentioned here are less frequent.

- (9) a. Mój stárc išál tám.  
 my.M.SG.NOM old man.M.SG.NOM go.PPART.M.SG there  
 ‘My old man went there.’

This feature differs from the other dialectal characteristics of the Timok variety described in this paper since it is neither entirely dialectal nor quite standard. The phenomenon is widely distributed in colloquial registers in practically all the regions of Serbia and has become increasingly widespread in the language of the media (Levin-Steinmann 2004: 15). Nonetheless, it appears notably more frequently in rural varieties (Meerman and Sonnenhauser 2017). Our data show that it is especially present in Timok varieties, thus it posits itself as a relevant feature.

In order to exploratively address the distribution of ‘short’ and ‘long’ forms in the whole corpus on the one hand, and the linguistic factors that influence the omission on the other, the following variables are taken into consideration:

- Dependent variable: AUX status in 3<sup>rd</sup> person (-AUX / +AUX)
- Independent variables: aspect, transitivity, lexical group (modal verbs, the verb *ima(ti)* ‘to have’ in existential (intransitive) function (as opposed to the initial possessive meaning), the verb *je (biti)* ‘to be’ vs. verbs of other lexical groups of different semantic classes).

Searching for relevant examples in the annotated Timok corpus required identifying all clauses where the perfect participle tense is used. In a next step, the examples retrieved were automatically divided into three groups, using text and annotation: clauses with -AUX perfect forms, clauses with +AUX perfect forms and clauses in the conditional mood (the latter group was excluded from the analysis). From the total number of 13,233 examples with perfect tense, 8,343 (63.05%) are -AUX forms, 4,890 (36.95%) are +AUX forms. The difference is statistically significant ( $\chi^2=58.039$ ,  $df=1$ ,  $p\text{-value}=0.001$ ). The plot in Figure 10 depicts the distribution across all speakers; it shows that the median frequency of the -AUX forms is greater than that of the +AUX forms,

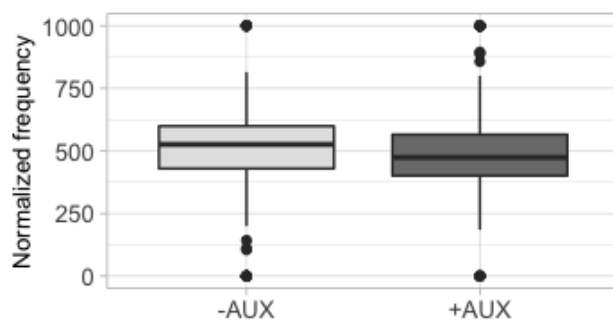


Figure 10: +AUX and -AUX frequencies across speakers in the overall sample

In order to identify linguistic factors that potentially play a role in the choice of either option (aspect, transitivity, lexical class), the sample was reduced by selecting several frequent verbs since the initial overall sample was too large for manual coding. The verbs were chosen to belong

to different semantic classes (motion, sensory perception, etc.) in order to control for the potential effect of the semantic type of the verb.<sup>28</sup> The resulting subsample contains 3171 examples: 1800 +AUX (57%) and 1371 -AUX (43%) constructions. As there is no explicit morphological annotation in the corpus related to the linguistic factors under scrutiny, the subsampling enabled manual categorization. The correlation between the linguistic variables and the frequencies of +AUX and -AUX forms does show a significant effect. The tendencies for each linguistic variable are illustrated in the bar plot in Figure 11.

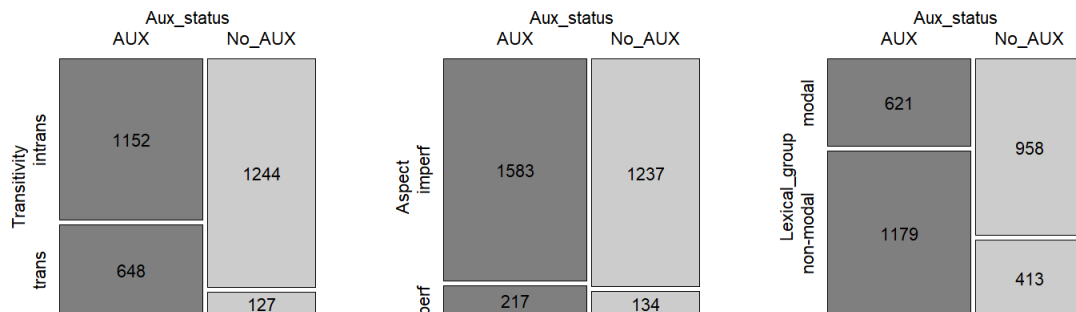


Figure 11: Linguistic properties of -AUX and +AUX forms in Timok corpus (proportions).

In order to control for the statistical significance of the linguistic factors—transitivity, aspect and lexical group—we apply a logistic regression analysis to the reduced dataset. The aspect variable contains two values: perfective and “not perfective”, the latter combining imperfective and biaspectual (e. g. *čuti* ‘to hear’, *videti* ‘to see’) verbs. We ran two models in order to avoid the quasi-complete separation of the predictor variables aspect and lexical group, since there are no perfective verbs in the group of “modal” verbs. It could have affected the reliability of the regression, if we had included both of them in one model.

	modal	non_modal
not_perfective	1872	1231
perfective	0	367

Table 3. Frequencies of different modal and aspect classes

<sup>28</sup> The thus selected verbs comprise: (1) *razumeti* / *razumevati* ‘to understand’, (2) *čuti* (biaspectual) / *slušati* ‘to hear’ (*čuti* is biaspectual in Standard Serbian), (3) *videti* / *gledati* ‘to see’, (4) *doći* / *dolaziti* ‘to arrive’, (5) *živeti* (imperfective only) ‘to live’, (6) *naći* / *nalaziti* ‘to find’, (7) *ići* (biaspectual) ‘to go’, (8) *umreti* / *umirati* ‘to die’. Apart from these eight verbs, several other frequent verbs were selected: *biti* ‘to be’, *imati* ‘to have’, *moći* ‘to be able to, to be allowed to’, *hteti* ‘to want’, *morati* ‘to must’, *treba* ‘to need’. The latter form a group conventionally called “modal” (as opposed to the previous group called “lexical”) even if it includes not only modal verbs but also existential (*biti* and *imati*) and possessive (‘to have’) ones.

For the first model (with Aspect and Transitivity as predictors) the reference value of the response variable is the presence of the auxiliary (+AUX) and for the grammatical variables the baseline values are “intransitive” and “not perfective”. The coefficients of the model and corresponding odds ratios for the predictors are presented in Table 4. The intercept coefficient shows the log odds of occurring of -AUX form with intransitive and non-perfective verbs. The odds are greater than 1.18 and this is statistically significant, which means that the grammatical context “imperfective and non-perfective” is the most favorable for the occurrence of -AUX. Having a closer look at the model coefficients we can conclude that the “transitive” value of the transitivity-variable reduces the likelihood of -AUX (odds ratio = 0.17), which means that the odds of -AUX with intransitive verbs is 5.8 times higher than with transitive verbs. As for the aspect, the odds of -AUX with imperfective verbs are 1.7 times higher than with the “non-perfective” ones (odds ratio = 0.56).

	<b>B(SE)</b>	<b>Odds ratio</b>	<b>P-value</b>
<b>Intercept</b>	0.16(0.04)		p<0.001
<b>Transitivity=trans</b>	-1.77(0.106)	0.16	p<0.001
<b>Aspect=perf</b>	-0.56(0.11)	0.567	p<0.001

Table 4. Effects of transitivity and aspect on the AUX status.

For the second model (with Lexical group and Transitivity as predictors) the reference value of the dependent variable is the presence of the auxiliary (+AUX) and for the grammatical variables the baseline values are “intransitive” and “non-modal”. The coefficients of the model and corresponding odds ratios are given in the Table 5. The most favorable grammatical context for the -AUX forms are intransitive and modal verbs (intercept = 0.44, odds =1.5). As in the previous model, the odds of -AUX with transitive verbs are lower than with intransitive ones (odds ratio = 0.36). As for the variable “Lexical group”, the non-modal value reduces the likelihood of -AUX (odds ratio = 0.34), so the odds of -AUX with “modal” verbs are 2.8 times higher than with non-model ones).

	<b>B(SE)</b>	<b>Odds ratio</b>	<b>P-value</b>
<b>Intercept</b>	0.44(0.05)		p<0.001
<b>Transitivity = transitive</b>	-1.01(0.12)	0.36	p<0.001
<b>Lexical group = non-modal</b>	-1.02(0.08)	0.34	p<0.001

Table 5. Effects of transitivity and lexical group on AUX status

In sum, the preference for omitting the auxiliary in Timok depends on the linguistic context. Given that auxiliary omission is not uncommon in the development of the Slavic languages (see Dickey



2013 for an overview), it remains to be investigated whether the observations for Timok also reflect a more general diachronic tendency. This is also interesting in functional terms: in Russian, for instance, the -AUX general past has replaced the +AUX perfect (e.g., *sotvorilʹ esi* ‘you have created’). Typically, this diachronic process was preceded by a period of coexistence and competition of both forms (see a short review of these processes in Sichinava 2014) with no obvious semantic differentiation. In Czech, on the contrary, this period of coexistence also displayed an opposition of current relevance/emphasis vs. neutral preterit (Dickey 2013). A functional differentiation associated with the usage / omission of the auxiliary concerning the relation between the author and the text also emerged in Balkan Slavic (cf., e.g., Fielder 1998), resulting in a contrast in confirmativity in some contemporary varieties (cf., e.g., Friedman 2004). As a consequence, the fact that in the Timok variety -AUX frequency is higher than +AUX frequency (fig. 10) might reflect either a diachronic change along the general Slavic lines or a functional differentiation of +AUX and -AUX forms characteristic of Balkan Slavic. A more detailed qualitative analysis is needed in order to state whether this distribution is semantic or rather formal. An analysis of various linguistic variables which revealed that the -AUX forms are preferred by imperfective and intransitive verbs of the “modal” lexical group could be evidence for the latter (more details in Escher 2021).

### 3.2 Analysis of the Socio-geographic Factors

The Timok area is situated in the continuum between non-Balkan-Slavic Serbian varieties to the North and the West, and Balkan Slavic Bulgarian, other Torlak varieties and Macedonian (see Sobolev et al. in this volume for details on areal embedding)<sup>29</sup>. Based on the Timok variety’s geographic position, Balkan Slavic features can be expected to vary across the Timok area and show higher frequency in the territories close to the ones where other balkanized varieties are spoken, especially to the East and South. This can be verified by correlating the values of geographic coordinates longitude and latitude with the linguistic variables.

In addition to the geographic coordinates, further geographic factors have been shown to affect the spread of language properties and linguistic features (see Axelsen, Manrubia 2014; Nichols 2015). Distance and geographical accessibility can determine the extent of contact between people and therefore affect their language production (Jeszensky et al. 2017), in particular the choice among diatopically and/or diastratically different manifestations of a specific structure. In the case of Timok, some villages located at higher altitudes are remote and sparsely inhabited, while others, located at lower altitudes, are well connected by a regional road (and were well connected in the recent past). It can be expected that this difference is reflected in the degree of Standard Serbian influence. The administrative and educational center of the area, Knjaževac, an urban center and representative of the prestigious culture of the capital, can be assumed to play a

---

<sup>29</sup> It is noteworthy that the Timok variety directly borders with the Lužnica and Svrlijig-Zaplanje dialects (also belonging to the Prizren-Timok dialect zone) to the west and south, while Vlach (Romance) varieties are positioned to its north.

role in the spread of Standard Serbian influence in the area. Villages at higher altitudes and those further from the city are hence expected to show a higher quantity of features that are non-standard.

Demographic factors such as age or gender are often associated with linguistic variation (Chambers 2004, Chapter 3; Chambers & Trudgill 1998, 57–81). In particular the former can be regarded as particularly important in the case of Torlak, considering its endangered status and the danger of obsolescence among the last speakers (Salminen 2010, 37; Penev, Marinković 2012; RZS 2014). Young people are more aware of the prestige of the standard (over southern dialects in general; Petrović 2015), and are also in closer contact with the standard variety through education and the media, and by living in the city. Older people who remain in the area are therefore expected to use a more dialectal language, with lesser influence of the standard variety. The factor of gender can be expected to influence variation, given that the in the context of the Slavic dialectology women preserve non-standard features, being on the whole less travelled and educated<sup>30</sup> than men (Belić 1905, XXXIII; Ivić 1986, 92-93; Petrović 2015, 33).

In order to statistically determine the influence of the extra-linguistic factors underlying the variation encountered in our sample of the Timok region, the use of linguistic features was correlated with socio-geographic information. The conditions taken into account as independent variables are the geographic factors of longitude and latitude, altitude and the distance from the city, as well as the demographic factors age and gender. In investigating the influence of both kinds of factors, the data was segmented into locations for the analysis of geographic factors, and into speakers for the study of demographic factors. The frequencies of the features discussed in Section 3 were analyzed, each with a different reference value<sup>31</sup>:

- Analytic or synthetic marking strategy proportion per overall number of analysed examples of indirect object and possessor
- Post-positive demonstratives normalized per 1,000 nouns
- Particle *si* per normalized 1,000 verbs
- AUX omission normalized per 1,000 occurrences of the perfect tense

Since each of the analyses involved different parameters, the sampling of the data differs slightly. The number of locations and speakers for each analysis is presented in Table 6.

	N of locations	AGE			GENDER		
		Older	Younger	Total	F	M	Total
Marking of the IO and POSS	60	57	7	64	67	18	85
Post-positive demonstrative	59	76	11	87	68	18	86

<sup>30</sup> The corpus metadata on education is not sufficient to test the correlation between education and the use of linguistic features.

<sup>31</sup> See the documentation for details.

Particle <i>si</i>	59	62	6	68	15	76	91
AUX omission	68	71	11	81	60	22	82

Table 6. The number of locations and speakers included in each analysis

Given that the distribution of the dependent variables was not normal, Kendall's rank correlation was used for the analysis of the geographic factors, while a Mann-Whitney/Wilcoxon test was used for the socio-demographic factors.

The analysis presented here revealed a statistically significant correlation between the frequency of PPD and longitude and latitude and the distance from the administrative centre, showing that they are more commonly used in the eastern and southern parts of Timok, as well as closer to the city (PPD ~ LONG:  $\tau=0.446$ ,  $p<0.001$ ; PPD ~ LAT:  $\tau=-0.219$ ,  $p<0.02$ , PPD ~ Dist\_city:  $\tau=0.19$ ,  $p<0.05$ ). A previous analysis by Vuković and Samardžić (2018) revealed a correlation between geographic factors such as altitude and distance from the city and the use of PPDs. The statistical analysis of the case marking strategy revealed a marginally significant effect of latitude in the sample of nouns only, indicating that the analytic strategy is used more in the southern parts of Timok ( $\tau=-0.202$ ,  $p\text{-value} = 0.04885$ ). For the features 'auxiliary omission' and 'particle *si*', no correlation with geographic factors could be observed. Although there are differences in frequencies between the individual locations, they do not reveal areal patterns.

For the demographic factors of age and gender, it can be expected that the older speakers use the dialectal features more frequently than younger ones, and that women do so more than men. Moreover, it can be expected that these parallels will show different patterns for each of the features, depending in particular on how strongly they are regarded as shibboleths of Torlak. Analytic marking of indirect objects and possessors, PPDs and the particle *si* are expected to be used by the older/feminine population. On the other hand, AUX omission is not as sensitive, even though it is frequent in older speakers living in remote locations (Vuković, Escher, Sonnenhauser 2022), it is a part of the colloquial language in other regions as well (Meermann 2015).

In order to analyze the effect of age, the informants for which the necessary information was available (see the documentation for details) were divided into two groups: older and younger; in the older group there were no informants younger than 55, while the younger group included only high-school students. Exact values for the year of birth were not used, because the age gap between the older and the younger group is large and looking at specific ages did not seem necessary for the given context.

The effect of age differs across the four features investigated, see Figure 12. An effect can be seen in the marking strategy for indirect object/possessor, i.e., analytic/synthetic case (regardless of the function) (overall sample:  $W=304.5$ ,  $p<0.002$ ; the sample of nouns:  $W=134$ ,  $p<0.002$ ; the sample of pronouns:  $W=125$ ,  $p<0.002$ ), indicating that the older speakers use the analytic strategy more frequently. PPDs are also used more frequently by older speakers ( $W=369$ ,  $p<0.002$ ). No effect of age is visible for the use of the particle *si* and the AUX omission.

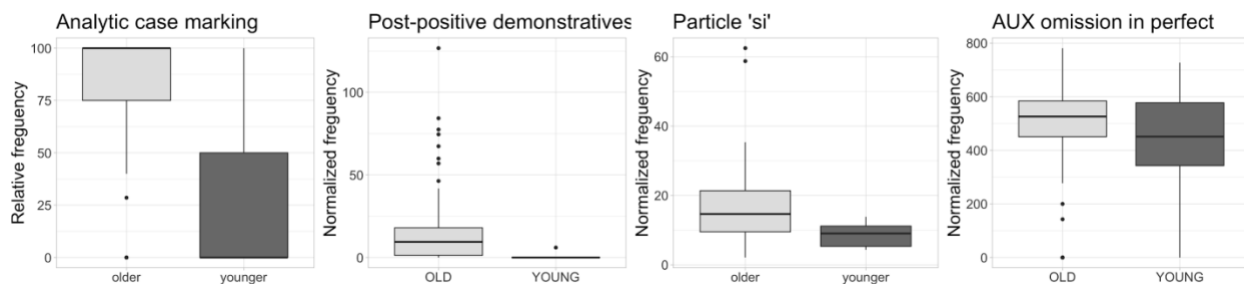


Figure 12: The distribution of features among older and younger speakers

When it comes to gender, all of the features studied are used more frequently by women, see Figure 13. The analysis showed a statistically significant difference for the analytic marking of possessor and indirect object (regardless of the function) (overall sample:  $W=875.5$ ,  $p<0.002$ , sample of pronouns:  $W=356$ ,  $p<0.002$ ; however the effect of gender was not significant for the sample of nouns), the post-positive demonstratives ( $W=260$ ,  $p<0.005$ ), as well as the particle *si* ( $W=841.5$ ,  $p<0.004$ ) and AUX omission ( $W=470.5$ ,  $p<0.005$ ). However, the effect of gender might be the consequence of the interrelationship between the age and the gender, as the younger group included mostly male speakers.

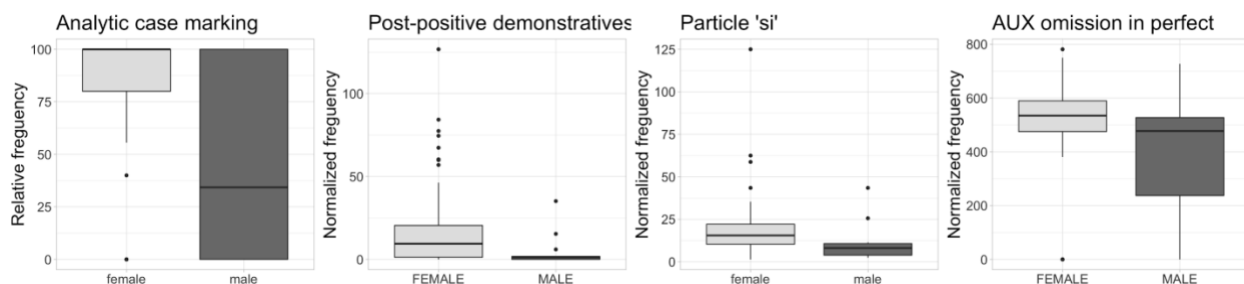


Figure 13: The distribution of features among female and male speakers

#### 4. Discussion

The analyses presented in this paper show that the Timok variety is not homogeneous with respect to the features investigated; instead, each feature displays characteristic preferences between the poles of dialectal/Balkan Slavic and standard Serbian/non-Balkan Slavic.

Two types of factors potentially influence the patterns of variation: linguistic conditions and socio-geographic factors. As to the former, the single analyses have shown that, for all features, the choice of options is influenced by feature-specific linguistic conditions. This may be interpreted in various ways: as indicating a functional differentiation, as reflecting the diffusion of Balkan features into West South Slavic along the lines established for the diachrony of Balkan Slavic, or as attesting to general Slavic trends. The data from our corpus seem to suggest a functional differentiation for the synthetic vs. analytic marking of indirect objects and possessor.

However, this may be a mere coincidence attesting to a particular stage of an ongoing diachronic process. The possibility of omitting the third person auxiliary seems to conform to a general Slavic trend; whether the options of using or omitting the third person auxiliary with perfect forms is functionally relevant in a way conforming to the Balkan Slavic requires further qualitative analyses. For the usage of both PPDs and the prevalence of the analytic marking strategy for indirect object and possessor, Balkan Slavic influence seems likely, while the usage of *si* as a particle might very well be in line with the features typical of vernacular South Slavic. Concerning the particle *si*, its particular placement preferences – 2<sup>nd</sup> position but verb adjacent – seem to follow a development that has been observed for older stages of Bulgarian, where it apparently relates to the language internal restructuring of the verbal complex (Pancheva 2005). Gaining more insight into these processes requires the inclusion of older data in order to add the necessary historical dimension, and of data from the broader Slavic (cf., e.g., Sobolev 1998 or Šimko 2020)—or even more general—context to assess the degree to which the Timok situation is indeed peculiar.

The variation encountered for the structures discussed may also have a vertical, i.e., register-based dimension, with a preference for the West South Slavic realization of the features resulting from Standard Serbian influence. This is confirmed by the effect of the socio-demographic factors of age and gender on the use of some of the features under scrutiny, even though it needs to be kept in mind that the corpus is not balanced in this respect (it includes more women than men and more older than younger speakers, and usually only one speaker per location). Similarly, the fact that younger speakers typically do not use the analytic case to mark indirect objects and possessors but use the dative instead, that they avoid PPDs and rarely use the particle *si* and auxiliary omission, might suggest an influence of Standard Serbian. Though only one young speaker used PPDs, a closer look at the transcripts indicated that he does so only when citing older speakers or giving typical dialect examples. This shows that younger speakers are aware of that option but consider it as diastatically marked and restricted to specific communicative contexts. The preference for the analytic marking of possessor/indirect object, usage of PPDs and the particle *si* by women, and the fact that gender is not correlated with the omission of AUX, all point in the same direction. This also suggests that the particle *si* occupies an intermediate position between Serbian vernacular varieties and Balkan Slavic. Whether, and to what extent, the Serbian influence is indeed vertical, and hence conditioned by education and register, or horizontal, and hence due to dialect contact, needs to be further investigated.

Based on the results of our analysis, we interpret the variation encountered for the features under consideration as being conditioned by an interplay of horizontal and vertical language contact: the availability of the dialectal patterns reflects—in varying degrees—contact with the balkanized South Slavic varieties, whereas their instability and—in some cases—vanishing (as shown by younger speakers) attests to the growing influence of the standard language.

The quantitative and corpus-based approach taken in the study allows us to interpret the obtained data on a large-scale basis and, based on statistically significant results, to make generalizations regarding the variation of the investigated features across locations and speakers included in the corpus (for a similar ‘aggregate’ methodology see papers in Szmrecsanyi, Wälchli

2014 and Nerbonne, Kretzschmar 2012). This approach differs from traditional approaches pursued within linguistic geography and dialectology which do not allow for the comparisons within different categories and typically treat all elements (distinct word forms, constructions, etc.) separately for each location.

## 5. Conclusion

As under a magnifying glass, the variation encountered in contemporary data from speakers across the Timok region uncovers the challenges involved in identifying the driving forces that interact in shaping the manifestations of linguistic patterns. As has been shown for the coding of indirect object and possessor, the employment of PPDs, the distribution of the particle *si* and the usage vs. omission of the 3<sup>rd</sup> person auxiliary, diatopic and diachronic contacts seem to have been crucial for the patterns of variation encountered in contemporary Timok. However, whether these patterns are indeed contact-induced, whether they reflect a particular stage in a family-internal, i.e., Slavic, development or conform to more general (e.g., Standard Average European or even universal) tendencies can only be judged by including historical data and comparative data from the South Slavic dialect continuum and beyond. As the example of Timok shows as well, the various driving forces may play a different role for the different patterns (e.g., PPDs are less likely to conform to a more general Slavic pattern than the drop of the 3<sup>rd</sup> person auxiliary). In addition, this variation and change can only be assessed when looking deeper into complex categories, both in formal (such as focusing on finer-grained argument coding patterns instead of ‘case’ vs. ‘preposition’) and functional (such as identifying potential functional differences associated with differences in usages frequencies as observed for the particle *si*) terms. In these respects, the case of Timok as a manifold transitional variety emerges as highly relevant for dealing with the core questions of linguistic variation, contact and change.

University of Zurich

Institute of Balkan Studies, Serbian Academy of Sciences and Arts

University of Zurich

Institute for Balkan Studies, Serbian Academy of Sciences and Arts

University of Belgrade and University of Bologna

Russian Academy of Sciences, Institute for Linguistic Studies

University of Zurich

Vuković Teodora  
teodora.vukovic2@uzh.ch

Mirić Mirjana  
mirjana.miric@bi.sanu.ac.rs

Escher Anastasia  
anastasia.escher@uzh.ch

Svetlana Ćirković  
[svetlana.cirkovic@bi.sanu.ac.rs](mailto:svetlana.cirkovic@bi.sanu.ac.rs)

Miličević Petrović Maja  
maja.milicevic2@unibo.it

Andrey N. Sobolev  
[sobolev@staff.uni-marburg.de](mailto:sobolev@staff.uni-marburg.de)

Barbara Sonnenhauser  
barbara.sonnenhauser@uzh.ch

## References

- Arsenijević, Boban. 2013. "Evaluative Reflexions: Evaluative Dative Reflexive in Southeast Serbo-Croatian." In Beatriz Fernandez; Ricardo Etxepare (eds). *Variations in Datives: A Microcomparative Perspective*, 1-21. Oxford: Oxford University Press.
- Asenova, Petya. 2002. *Balkansko ezikoznanie*. Veliko Tärnovo.
- Axelsen, Jacob B.; Manrubia, Susanna. 2014. " ". *Proceedings of the Royal society B* 281.1784: 20141179.
- Belić, Aleksandar. 1905. "Dijalekti istočne i južne Srbije." *Srpski dijalektološki zbornik* 1: 1–674.
- Bogdanović, Nedeljko. 1987. "Govor Aleksinačkog Pomoravlja." *Srpski dijalektološki zbornik* 33:7–302.
- Bogdanović, Nedeljko. 1979. "Govori Bučuma i Belog Potoka." *Srpski dijalektološki zbornik* 25: 1-178.
- Catasso, Nicholas. 2011. "The Grammaticalization of Demonstratives: A Comparative Analysis." *Journal of Universal Language* 12.1: 7-46.
- Chambers, Jack K. 2004. "Patterns of Variation including Change." In Jack K. Chambers; Peter Trudgill; Natalie Schilling-Estes (eds). *The Handbook of Language Variation and Change*, 349-372. Malden MA et al.: Blackwell Publishing.
- Chambers, Jack K.; Trudgill, Peter; Schilling-Estes, Natalie (eds.). 2004. *The Handbook of Language Variation and Change*. Malden MA et al.: Blackwell Publishing.
- Chambers, Jack K.; Trudgill, Peter. 1998. *Dialectology*. Cambridge: Cambridge University Press.
- Cheng, Joe; Karambelkar, Bhaskar; Xie, Yihui. 2019. leaflet: "Create Interactive Web Maps with JavaScript 'Leaflet' Library." R package version 2.0.3.  
<https://CRAN.R-project.org/package=leaflet>, accessed February 20, 2020.
- Ćirić, Ljubisav. 1983. "Govor Lužnice." *Srpski dijalektološki zbornik* 29: 7–119.
- Ćirić, Ljubisav. 1999. "Govori Ponišavlja," *Srpski dijalektološki zbornik* 46: 7–262.
- Ćirković, Svetlana. 2018a. "Neverbalna komunikacija u antropološko-lingvističkom intervjuu: analiza multimodalnih transkripata na temu gajenja i prerade konoplje." In Svetlana Ćirković (ed.). *Timok. Folkloristička i lingvistička terenska istraživanja 2015–2017*, 219–238. Knjaževac: Narodna biblioteka „Njegoš“; Beograd: Udruženje folklorista Srbije.
- Escher, Anastasia. 2021. "Auxiliary Omission in the Perfect tense in Timok." *Balkanistica* 34: 41–63.
- Féry, Caroline; Krifka, Manfred. 2008. "Information Structure: Notional Distinctions, Ways of Expression." In Piet van Sterkenburg (ed.). *Unity and Diversity of Languages*, 123–135. Amsterdam, Philadelphia: John Benjamins.
- Halupka-Rešetar, Sabina. 2009. "Red reči u službi izražavanja informacionog fokusa u srpskom jeziku." *Zbornik Matice srpske za filologiju i lingvistiku* 52.2: 169-186.
- Halupka-Rešetar, Sabina. 2011. *Rečenični fokus u engleskom i srpskom jeziku*. Novi Sad: Filozofski fakultet Univerziteta u Novom Sadu.
- Fielder, Grace. 1998. "Discourse Function of Past Tenses in Pre-modern Balkan Slavic Prose." In Alan Timberlake; Michael Flier (eds). *American Contributions to the Twelfth International Congress of Slavists. Cracow, August--September 1998*, 344-361. Indiana: Slavica.
- Friedman, Victor A. 2002. *Macedonian*. München: Lincom Europa.

- Friedman, Victor A. 2004. "The Typology of Balkan Evidentiality and Areal Linguistics." In Olga Tomić (ed). *Balkan Syntax and Semantics*, 101-134. Amsterdam, Philadelphia: John Benjamins.
- Ivić, Pavle. 1985. *Dijalektologija srpskohrvatskog jezika: Uvod i štokavsko narečje*. Novi Sad: Matica srpska.
- Ivić, Pavle. 1986. *Srpski narod i njegov jezik*. Beograd: Srpska književna zadruga.
- Ivić, Pavle. 1998. Srpski dijalekti i njihova klasifikacija. *Zbornik Matice srpske za filologiju i lingvistiku* 41.2: 113–132.
- Ivić, Pavle. 2009. *Srpski dijalekti i njihova klasifikacija*. Ed. by Slobodan Remetić. Novi Sad, Sremski Karlovci: Izdavačka knjižarnica Zorana Stojanovića.
- Jeppson, Haley; Hofmann, Heike; Cook, Di. 2018. ggmosaic: Mosaic Plots in the 'ggplot2' Framework. R package version 0.2.0.  
<https://CRAN.R-project.org/package=ggmosaic>, accessed February 20, 2020.
- Jeszszky, Péter; Stoeckle, Philipp; Glaser, Elvira; Weibel, Robert. 2017. "Exploring Global and Local Patterns in the Correlation of Geographic Distances and Morphosyntactic Variation in Swiss German." *Journal of Linguistic Geography* 5: 1-23.
- Joseph, Brian D. 1992. "The Balkan languages." In William Bright (ed.). *Oxford International Encyclopedia of Linguistics*, 135-155. Oxford: Oxford University Press.
- Jurišić, Marina. 2009. *Govor Gornje Pčinje*, Beograd: Institut za srpski jezik.
- Kemmer, Suzanne. 1993. *The Middle Voice*. Amsterdam, Philadelphia: John Benjamins.
- Klajn, Ivan; Piper, Predrag. 2013. *Normativna gramatika srpskog jezika*. Novi Sad: Matica Srpska.
- Koněr, Darija V.; Makarova, Anastasia L.; Sobolev, Andrej N. 2019. "Statističeskij metod jazykovogo profilirovanija nositelja dialekta: na materiale vostočnoserbškogo idioma sela Berčinovac." *Vestnik Tomškogo gosudarstvennogo universiteta. Filologija* 58: 17–33.
- Levin-Steinmann, A. 2004. *Die Legende vom Bulgarischen Renarrativ: Bedeutung und Funktionen der kopulalosen l-Periphrase*. München: Otto Sagner.
- Lindstedt, Jouko. 2000. "Linguistic Balkanization: Contact-induced Change by Mutual Reinforcement." In Dicky Giblers; John Nerbonne, Jos Schaecken (eds). *Languages in Contact*, 231–246. Amsterdam, Atlanta: Editions Rodopi B.V.
- Marschner, Ian C. 2011. "glm2: Fitting Generalized Linear Models with Convergence Problems." *The R Journal* 3.2: 12-15.
- Meermann, Anastasia. 2015. "Truncated Perfect in Serbian: A Marker of Distance?" In Barbara Sonnenhauser; Anastasia Meermann (eds). *Distance in language: Grounding a metaphor*, 95-116. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Meermann, Anastasia; Sonnenhauser, Barbara. 2016. "Das Perfekt im Serbischen zwischen Slavischer und Balkanslavischer Entwicklung." In Alena Bazhutkina; Barbara Sonnenhauser (eds). *Linguistische Beiträge zur Slavistik. XXII: JungslavistInnen-Treffen in München. 12. bis 14. September 2013*, 83-110. Leipzig: Bibliion Media.
- Meyer, David; Zeileis, Achim; Hornik, Kurt. 2020. vcd: Visualizing Categorical Data. R package version 1.4-5.
- Milićević, Milan 2012 [1876]. *Knjaževački okrug*. Knjaževac: Narodna biblioteka „Njegoš” (Reprint).
- Miloradović, Sofija. 2003. *Upotreba padežnih oblika u govoru Paraćinskog Pomoravlja: balkanistički i etnomigracioni aspekt*. Beograd: Etnografski institut SANU.
- Milosavljević, Stefan. 2019. "Semantika i pragmatika evaluativnog reflektivnog dativa u govoru jablaničkog kraja: u svetlu srodnih fenomena u drugim jezicima/ govorima." In Miloš



- Kovačević; Jelena Petković (eds.). *Savremena proučavanja jezika i književnosti* 10.1, 45–56. Kragujevac: FILUM.
- Mišeska-Tomić, Olga. 2004. The Balkan Sprachbund Morpho-Syntactic Properties: An introduction. In Olga Mišeska Tomić (ed.). *Balkan Syntax and Semantics*, 1-55. Amsterdam: John Benjamins.
- Mišeska-Tomić, Olga. 2006. *Balkan Sprachbund Morpho-Syntactic Features*. Dordrecht: Springer.
- Mladenova, Olga. 2007. *Definiteness in Bulgarian. Modelling the Processes of Language Change*. Berlin, Boston: De Gruyter Mouton.
- Nerbonne, John; Kretzschmar Jr, William A. 2013. „Dialectometry.” *Literary and Linguistic Computing* 28.1: 2-12.
- Nichols, Johanna. 2015. “Types of spread zones, Open and closed, horizontal and vertical.” In Rik de Busser; Randy J. LaPolla (eds). *Language Structure and Environment: Social, Cultural, and Natural Factors*, 261–286. Amsterdam: John Benjamins.
- Pancheva, Roumyana. 2005. “The rise and fall of second-position clitics.” *Natural Language & Linguistic Theory* 23: 103–167
- Pavlović, Milivoj. 1970. *Govor Janjeva*. Novi Sad: Matica srpska.
- Penev, Goran; Marinković, Ivan. 2012. “Prvi rezultati popisa stanovništva 2011. S posebnim osvrtom na promenu broja stanovnika jugoistočne Srbije.” In (ed.) *Stanovništvo jugoistočne Srbije: uticaj demografskih promena u jugoistočnoj Srbiji na društveni razvoj i bezbednost*, 22–43. Niš: Centar za naučnoistraživački rad SAU i Univerziteta u Nišu.
- Petrova, Galina. 2014. “Medialni glagoli s refleksivna semantika.” *Naučni trudove na Rusenskyja universitet Serija 6.3*. 53: 36–40.
- Petrović, Tanja. 2015. *Srbija i njen jug. “Južnjački dijalekti” između jezika, kulture i politike*. Beograd: Fabrika knjiga.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.  
<https://www.R-project.org/>, accessed February 20, 2020.
- RZS. 2014. “Popisni atlas 2011 Census atlas.” In *Popis stanovništva, domaćinstava i stanova 2011. u Republici Srbiji*. Beograd: Republički zavod za statistiku.
- Salminen, Tapani. 2010. “Europe and the Caucasus.” In Christopher Moseley (ed.). *Atlas of the World's Languages in Danger*, 32-42. Paris: UNESCO.
- Sikimić, Biljana; Sobolev, Andrey N.; Sonnenhauser Barbara. (this volume). Introduction: (Dis)entangling Traditions in the Central Balkans: Performance and Perception. The Case of Torlak.
- Savova, Dimka. 2017. “Glagoli s elementa *si/sobie* v bālgarskija i v polskija ezik.” *Zeszyty Cyrylo-Metodiańskie* 6: 38–56.
- Sobolev, Andrej. 1991. “Kategorija padeža na periferii balkanoslavjanskogo areala.” *Zbornik Matice srpske za filologiju i lingvistiku* 34.1: 93–139.
- Sobolev, Andrej. 1994. *Timočko-lužnički govori u južnoslovenskoj Č, DŽ-zoni: Govori prizrensko-timočke oblasti i susednih dijalekata*, 85–106. Niš: Filozofski fakultet Univerziteta u Nišu – Beograd: Institut za srpski jezik SANU, Centar za naučna istraživanja SANU i Univerziteta u Nišu.
- Sobolev, Andrej. 1995. “O pirotskom govoru u svetlu najnovijih istraživanja.” *Pirotski zbornik* 21: 195–214.
- Sobolev, Andrej. 1998. *Sprachatlas Ostserbiens und Westbulgariens*. München: Biblion.

- Sobolev, Andrej. 2003. *Malyj dialektologičeskij atlas balkanskix jazykov: Probnyj vypusk*. München: Biblion.
- Sobolev, Andrej. 2008. From Synthetic to Analytic Case: Variation in South-Slavic Dialects. In Andrej Malchukov; Andrew Spencer (eds). *The Oxford Handbook of Case*, 716–729. Oxford: Oxford University Press.
- Sobolev, Andrej. 2020. “Dialect coherence of the speech production in Timok patoi of Serbian: on the data from the Petruša idiom in Knjaževac municipality.” In *Proceedings of the 24th Conference on Indo-European Linguistics and Classical Philology*. St. Petersburg: RAN.
- Sobolev, Andrej N.; Mirić, Mirjana; Konior, Daria V.; Ćirković, Svetlana. (this volume). “Torlak: Areal Embedding and Linguistic Characteristics.”
- Stankov, Valentin. 1967. “Kategorii na indikativa v sãvremennija bãlgarski ezik.” *Bãlgarski ezik* 17: 330–344.
- Stanojević, Marinko. 1911. “Severno-timočki dijalekat.” *Srpski dijalektološki zbornik* 2: 360–463.
- Szmrecsanyi, Benedikt; Wälchli, Bernhard (eds). 2014. *Aggregating Dialectology, Typology, and Register Analysis: Linguistic Variation in Text and Speech*. Berlin, Boston: de Gruyter Mouton.
- Šantić, Danica; Martinović, Marija. 2007. Naselja Lužnice – geografsko-istorijska i prostorno-demografska transformacija. *Glasnik Srpskog geografskog društva* 87.2: 115–124.
- Šimko, Ivan. 2020. Definiteness Markers in the Life of St Petka. *Zeitschrift für Slawistik* 65.2: 272-307.
- Topolinjska, Zuzana. 2006. “Are there three variants of the definite article in Macedonian?” *Južnoslovenski filolog* 62: 7-15.
- Vukadinović, Vilotije. 1996. “Govor Crne Trave i Vlasine.” *Srpski dijalektološki zbornik* 42: 1-317.
- Vuković, Teodora. 2020. Spoken Torlak dialect corpus 1.0 (transcription). CLARIN. <http://hdl.handle.net/11356/1281>, accessed December 2, 2021.
- Vuković, Teodora. 2021. “Representing Variation in a Spoken Corpus of an Endangered Dialect: The Case of Torlak.” *Language Resources and Evaluation* 55.1: 731-756.
- Vuković, Teodora; Anastasia Escher; Barbara Sonnenhauser. 2022. “Degrees of Non-standardness: Feature-based Analysis of Variation in a Torlak Dialect Corpus.” *International Journal of Corpus Linguistics* 27.2: 220-247.
- Vuković, Teodora; Samardžić, Tanja. 2018. “Prostorna raspodela frekvencije postpozitivnog člana u timočkom govoru.” In Svetlana Ćirković (ed). *Timok: folkloristička i lingvistička terenska istraživanja 2015-2017*. Knjaževac: Narodna biblioteka "Njegoš"; Beograd: UFS.
- Vuković, Teodora. (forthcoming). An Empirical Corpus-Based Analysis of the Grammatical Status of Short Demonstratives in the Timok Dialect. *Journal of Slavic Linguistics*.
- Wahlström, Max. 2015. *The Loss of Case Inflection in Bulgarian and Macedonian*. Helsinki: University of Helsinki.
- Wickham, Hadley. 2007. “Reshaping data with the reshape package.” *Journal of Statistical Software* 21.12: 1-20.
- Wickham, Hadley. 2011. “The Split-Apply-Combine Strategy for Data Analysis.” *Journal of Statistical Software* 40.1: 1-29. <http://www.jstatsoft.org/v40/i01/>, accessed February 20, 2020.
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

- Wickham, Hadley; Averick, Mara; Bryan, Jennifer; Chang, Winston; D'Agostino McGowan, Lucy; François, Romain; Grolemund, Garrett; Hayes, Alex; Henry, Lionel; Hester, Jim; Kuhn, Max; Lin Pedersen, Thomas; Miller, Evan; Milton Bache, Stephan; Müller, Kirill; Takahashi, Kohske; Vaughan, Davis; Wilke, Claus; Woo, Kara; Yutani, Hiroaki. 2019. "Welcome to the tidyverse." *Journal of Open Source Software*, 4(43). URL: <https://doi.org/10.21105/joss.01686>, accessed February 20, 2020.
- Wickham, Hadley; François, Romain; Henry, Lionel; Müller, Kirill. 2019. "dplyr: A Grammar of Data Manipulation." R package version 0.8.3. <https://CRAN.R-project.org/package=dplyr>, accessed February 20, 2020.
- Xie, Yihui. 2019. tinytex: Helper Functions to Install and Maintain 'TeX Live', and Compile 'LaTeX' Documents. R package version 0.18.

#### Под увеличительным стеклом. Явление вариативности в торлакском диалектном комплексе

В статье рассматривается языковая вариативность, свойственная тимокским говорам Юго-Восточной Сербии в настоящее время. Диалектный материал, записанный в интервью с местными жителями, демонстрирует значительную вариативность в реализации различных языковых черт. Он отображает частично совпадающие друг с другом паттерны, принадлежащие к разным регистрам и формам существования языка, и возникшие в результате контакта между западно-южнославянскими и балканизированными южнославянскими говорами. В ходе анализа основное внимание уделяется взаимодействию между паттернами, приписываемыми стандартному сербскому языку (в основе которого лежат западно-южнославянские диалекты), и моделями, характерными для балканизированных говоров. Анализируемые релевантные в этом отношении признаки включают: маркирование косвенного объекта и посессора, постпозитивные артикли, частицу *si* как показатель дательного падежа в возвратном залоге, а также опущение вспомогательного глагола в перфекте. В первой части статьи рассматриваются относительные лингвистические факторы, а основное внимание уделяется морфосинтаксису с целью определения, какие языковые структуры препятствуют или облегчают использование тех или иных форм. Во второй части исследуется корреляция между выбранными языковыми явлениями и географическими и социальными параметрами с целью найти потенциальные пространственные закономерности распространения признаков и изучить влияние ландшафта (высоты над уровнем моря) или физических расстояний между населёнными пунктами (расстояния от административного центра). Что касается социальных факторов, возраст и пол рассматриваются в корреляции с лингвистическими данными с целью выявления различий в речи мужчин и женщин, или старших и младших носителей диалекта.