
Deverbal Nominalisations in the Timok Speech: A Preliminary Corpus Study

Andela Babić¹[\[0000-0002-7472-8198\]](#), Marko Simonović²[\[0000-0002-9651-6399\]](#),
Predrag Kovačević³[\[0000-0003-2545-9240\]](#)

¹ Serbian Academy of Sciences and Arts, Institute for Balkan Studies, Serbia,
andjela.redzic@bi.sanu.ac.rs

² University of Graz, Department of Slavic Studies, Austria,
marko.simonovic@uni-graz.at

³ University of Novi Sad, Faculty of Philosophy, Department of English Studies,
Serbia, predrag.kovacevic@ff.uns.ac.rs

Scientific paper

DOI: 10.18485/judig.2026.2.ch21

Abstract

This paper presents a preliminary corpus-based study of deverbal nominalisations in the Timok speech, a speech of the Prizren-Timok dialect of Serbian. The analysis focuses on two productive morphological classes: nominalisations in *-nje/-će* and nominalisations in *-cija*. The study is based on 606 tokens extracted from the Spoken Torlak Dialect Corpus 1.0 and examines four properties: the aspectual class of the source verb, complement realisation, pluralisation, and correspondence with standard Serbian forms. The results show a strong predominance of *-nje/-će* nominalisations, especially those derived from imperfective verbs, while *-cija* nominalisations are considerably less frequent. Overt complements are rare overall, suggesting limited preservation of verbal argument structure in the corpus data, although they occur more often with *-nje/-će* nominalisations than with *-cija* forms. By contrast, plural forms are proportionally more frequent among *-cija* nominalisations, supporting the view that this class is more strongly lexicalised and more fully integrated into the nominal domain. Finally, most tokens show morphological correspondence with standard Serbian, although the data also leave room for further investigation of dialect-specific derivational patterns. The study contributes to research on deverbal nominalisation in South Slavic by extending the empirical basis to a non-standard Balkan Slavic variety and to spontaneous spoken-language data.

Keywords: deverbal nominalisations, Timok speech, Prizren-Timok dialect, corpus study, aspect, pluralisation, complements

1. Introduction

Deverbal nominalisations occupy an important place in the study of the relation between verbal and nominal structure. Since they are formed from verbs but behave as nouns (Chomsky, 1970; Grimshaw, 1990; Alexiadou, 2001), they provide a useful testing ground for determining how much verbal structure is preserved inside the nominal domain. Previous research has shown that deverbal nominalisations may differ considerably in this respect: some retain properties associated with their verbal bases, such as argument structure, aspectual sensitivity, and event interpretation, whereas others are more lexicalised and behave like ordinary nouns (Alexiadou, 2001, 2025; Kornfilt and Whitman, 2011). In BCMS¹ and other Slavic languages, this contrast is particularly relevant for nominalisations in *-nje/-će* and *-cija*. Nominalisations in *-nje/-će* are usually closely related to verbal bases and have been argued to preserve verbal properties to a significant degree, especially when derived from imperfective verbs (Kovačević, 2021; Simonović and Arsenijević, 2014). By contrast, *-cija* nominalisations are typically more lexicalised and less compositional (Simonović and Arsenijević, 2018). These patterns have been examined in standard BCMS and in broader Slavic comparison (Arsenijević, 2007, 2020, 2021; Simonović et al., 2025b). However, their use in non-standard varieties remains insufficiently explored.

The present study focuses on deverbal nominalisations in the Timok speech, one of the speeches of the Prizren-Timok dialect of Serbian. Two productive morphological classes are examined: nominalisations in *-nje/-će*, illustrated by examples such as *kovanje* ‘forging’ from *kove*, *kovala*² ‘to forge’ and *raspeće* ‘crucifixion’ from *raspe*, *raspela* ‘to crucify’, and nominalisations in *-cija*, illustrated by forms such as *operacija* ‘operation’ related to *operiše*, *operisala* ‘to operate’. The aim of the paper is to provide a preliminary corpus-based analysis of these nominalisations in the Timok speech, focusing on

¹ In this paper, BCMS is used as a cover term for Bosnian, Croatian, Montenegrin and Serbian.

² In line with Milosavljević et al. (2026), we use the present tense and the feminine past participle form to identify Prizren-Timok verbs, as the system lacks the infinitive form.

their morphological properties, aspectual distribution, complement realisation, pluralisation, and correspondence with standard Serbian.

1.1. Speech description

The Timok speech is one of the speeches of the Serbian Prizren-Timok dialect (Ивић, 2001: 146–174). Like the Prizren-Timok dialect more broadly, it belongs to the Balkan Sprachbund, that is, the Balkan linguistic area, and displays a number of features that distinguish it both from standard Serbian and from most other Serbian and, more broadly BCMS dialects (Асенова, 2002; Friedman and Joseph, 2025; Sandfeld, 1930; Trubetzkoy, 1928). These features, commonly referred to as Balkanisms, are found at all levels of linguistic structure, including phonology, morphology, syntax, and the lexicon.

The features most relevant to the present study are morphological and syntactic: the postposed definite article, the analytic case system, the merger of genitive and dative forms, the analytic comparison, the analytic *velle*-future, object reduplication, the loss of the infinitive and its replacement by a finite complement (*da* + present tense) construction, and innovative verbal paradigms (Friedman and Joseph, 2025). These features are attested, to varying degrees, throughout the Prizren-Timok dialect area, including the Timok speech (Динић, 2008: IX–XXIII; Sobolev et al., 2023). The loss of the infinitive and the restructuring of the case system are especially relevant here, since they make the Timok speech a particularly suitable domain for investigating the interaction between verbal and nominal structure in deverbal nominalisations.

1.2. Available digital resources on Prizren-Timok dialect

Although many speeches of the Prizren-Timok dialect have been described in dialectological research, often on the basis of extensive textual material, digital corpora suitable for systematic investigation remain scarce. A notable exception is the *Spoken Torlak Dialect Corpus 1.0* (Vuković, 2020), from which a more recent database of Torlak verbal forms has also been derived (Milosavljević et al., 2026), (Milosavljević et al., 2025).

The Spoken Torlak Dialect Corpus 1.0 contains a substantial collection of semi-orthographically transcribed recordings from the Timok area. It consists of transcripts of semi-structured interviews

with 167 speakers of the Timok speech, mostly on topics related to traditional culture, history, and everyday life. The interviews were designed to elicit extended stretches of spontaneous speech. The corpus contains approximately 86.5 hours of recordings and 489,021 tokens annotated for stress, lemmatisation and morphosyntactic information (Vuković, 2021).

As noted by Vuković, the recorded speech shows the influence of standard Serbian, which speakers of the Timok speech typically perceive as more prestigious. These influences result from contact with the standard language through media and education, as well as from interaction with speakers of other speeches. The corpus therefore exhibits considerable variability in the use of non-standard features: it includes both prototypical dialect speakers with limited influence from the standard language and speakers whose language combines dialectal and standard features. This makes the corpus particularly relevant for the present study. In addition to examining a Balkan Slavic variety underexplored in the domain of deverbal nominalisations, the analysis is based on spontaneous spoken-language data, whereas previous research on deverbal nominalisations has generally relied on written corpora.

The verb database Mapping out the Torlak verb (Milosavljević et al., 2026), the dataset (Milosavljević et al., 2025) was used to verify the verbal forms recorded in our dataset. It comprises the 5,000 most frequent verbal lemmas extracted from the Spoken Torlak Dialect Corpus 1.0, of which 3,085 have been subjected to detailed morphosyntactic annotation. For each annotated verb, the database provides the third-person present-tense form, the feminine singular past participle, the theme-vowel class, the presence of a derivational suffix, and information on root allomorphy between the present-tense and participial form.

1.3. Nominalisations in BCMS and beyond

Across Slavic languages, nominalisations derived from verbal bases exhibit substantial variation with respect to aspectual restrictions, argument realisation, productivity, semantic transparency, and lexicalisation. As a result, they have played a central role in discussions concerning the amount of verbal structure preserved within nominal expressions and the relationship between morphology, syntax, and lexical semantics.

Nominalisations in *-nje/-će* (and their cognate forms in other Slavic languages) are among the most productive deverbal formations and frequently preserve properties associated with the verbal domain, including event interpretation and argument structure, serving as the Slavic counterpart of the English gerund. Previous research on BCMS (Arsenijević, 2020, 2021; Kovačević, 2021; Simonović and Arsenijević, 2014) has shown that this is especially true of *-nje/-će* nominalisations with imperfective verbal bases, which typically denote processes, activities, or ongoing events and remain relatively transparent both semantically and morphologically. By contrast, BCMS nominalisations derived from perfective verbs are considerably less productive and often display a higher degree of lexicalisation. In many cases, they no longer denote events in a straightforward compositional manner, but instead develop more specialised or resultative meanings. The studies cited above have interpreted this asymmetry as evidence that imperfective-derived nominalisations preserve more verbal structure than those derived from perfective verbs (Arsenijević, 2020; Simonović and Arsenijević, 2014). These differences have also been connected to broader questions concerning event structure and the interaction between aspect and nominal morphology in Slavic languages.

This contrast can be illustrated by the pair *priznavanje* and *priznanje*. The imperfective verb *priznavati* ‘to admit’ gives rise to *priznavanje* ‘admitting’, a nominalisation with a transparent process interpretation. By contrast, the perfective verb *priznati* ‘to admit, recognise’ gives rise to *priznanje*, a more lexicalised form whose meanings include ‘admission’, ‘confession’, and ‘recognition’.

- (1)
- | | | | |
|----|--------------------------|---|---|
| a. | <i>priznavati</i> | → | <i>priznavanje</i> |
| | ‘to admit.IPFV’ | → | ‘admitting’ |
| b. | <i>priznati</i> | → | <i>priznanje</i> |
| | ‘to admit/recognise.PFV’ | → | ‘admission; confession;
recognition’ |

Another robust class of deverbal nominalisations in BCMS consists of Latinate nominals in *-cija*. Unlike *-nje/-će* nominalisations, *-cija* forms are often more lexicalised and less transparently connected to their verbal bases. A relevant contrast is provided by pairs such as

asociranje and *asocijacija*, both related to the verb *asocirati* ‘to associate’. While *asociranje* transparently denotes the process or activity of associating, *asocijacija* may denote not only the act or process of association, but also a mental connection, an organisation, or an established associative relation. The latter is therefore more readily available as a count noun and more easily integrated into the nominal lexicon. Earlier research has therefore proposed that *-nje/-će* and *-cija* nominalisations differ not only morphologically, but also structurally and semantically (Simonović and Arsenijević, 2018).

Research on deverbal nominalisations in BCMS has focused primarily on standard language varieties and written corpora. Much less attention has been devoted to dialectal varieties, despite the fact that they provide important evidence for understanding the interaction between verbal and nominal structure under different morphosyntactic conditions. This issue is particularly relevant in the case of the Prizren-Timok dialect area, whose speeches exhibit numerous Balkan linguistic features, including the restructuring of the case system and the loss of the infinitive. Such features may affect both the formation of deverbal nominalisations and the realization of their complements.

The Timok speech is therefore especially relevant for the study of deverbal nominalisations. On the one hand, it belongs to the wider South Slavic system in which the distinction between *-nje/-će* and *-cija* nominalisations has already been shown to correlate with aspect, productivity, and lexicalisation. On the other hand, its Balkan morphosyntactic profile raises the possibility that deverbal nominalisations may behave differently from those described in standard BCMS. Whether the patterns established for standard varieties also hold in the Timok speech remains an open empirical question. The present study addresses this question through a corpus-based analysis of deverbal nominalisations in spontaneous spoken language.

2. Research questions and hypotheses

The analysis is guided by four interrelated research questions, formulated in light of previous findings on deverbal nominalisations in BCMS and other Slavic varieties. Together, these questions address the distribution, internal structure, and nominal behaviour of the forms under investigation.

2.1. Quantitative relations (IPFV vs. PFV. vs. CIJA)

The first research question concerns the relationship between the nominalisation type and the aspectual class of the source verb. More specifically, it asks whether nominalisations in *-nje/-će* are preferentially derived from imperfective or perfective verbs, and how their distribution compares with that of *-cija* nominalisations. On the basis of previous research on standard BCMS, we expect imperfective-derived *-nje/-će* nominalisations to be the most frequent type in the Timok speech as well, whereas for the relative ordering of the other two types there is no clear expectation. The predicted hierarchy is therefore: imperfective *-nje/-će* nominalisations > perfective *-nje/-će* nominalisations, *-cija* nominalisations.

2.2. Complements

The second research question concerns the syntactic behaviour of deverbal nominalisations, in particular the extent to which they occur with overt complements corresponding to the arguments of their source verbs. The analysis focuses on three complement types: genitive complements, accusative complements in the general case form (*casus generalis*), and prepositional complements introduced by *na* followed by the general case. Since overt complements may indicate the preservation of argument structure, they are expected to occur most frequently with imperfective *-nje/-će* nominalisations. The predicted pattern is therefore: imperfective *-nje/-će* nominalisations > perfective *-nje/-će* nominalisations and *-cija* nominalisations.

2.3. Plural ratios

The third research question concerns number, specifically whether the nominalisations under investigation occur in the plural. The underlying assumption is that pluralisation reflects countability and lexicalisation, both of which indicate stronger integration into the nominal domain. We therefore expect plural forms to be relatively more frequent with *-cija* nominalisations and perfective *-nje/-će* nominalisations. The predicted pattern is therefore: *-cija* nominalisations and perfective *-nje/-će* nominalisations > imperfective *-nje/-će* nominalisations.

2.4. Standard form

The fourth and final research question is motivated by the distance between the Timok speech and standard Serbian, as well as by Vuković’s (2021) observation that the corpus exhibits variation with respect to standard language influences. The study therefore also asks whether the morphological form of each deverbal nominalisation corresponds to its standard Serbian counterpart. This is relevant for determining whether any of the nominalisation classes might reflect standard-language influence rather than dialect-internal formation. The comparison concerns only the morphological form of the nominalisation; prosodic and phonetic differences are disregarded.

3. Materials and Methods

The analysis is based on a database of deverbal nominalisations extracted from the *Spoken Torlak Dialect Corpus 1.0* (Vuković, 2020).³ The search included all speakers from the Timok area, while portions produced by the researchers were excluded. The target forms were extracted by means of regular expressions designed to capture all the inflectional forms of nominalisations in *-nje/-će* and *-cija* in the Timok speech, as given below.

Regular expressions for PT forms:

```
[word=".*(a|e|é|á)nj(a|e|é|á|ata|eto|éto|áta)"]
```

```
[word=".*(e|é|i|í|ú|u)(ć|č|q|k)(e|é|á|ata|eto|éto|áta)"]
```

```
[word=".*(c|s|z)(i|í)j(a|ata|á|áta|e|ete|é|éte|u|utu|ú|útu)"]
```

After the initial extraction, additional searches were carried out for standard Serbian case forms not captured by the original queries; these forms were checked, but not included in the analysis.

Regular expressions for additional standard forms:

```
[word=".*(a|e|é|á)nj(u|em|ima)"]
```

```
[word=".*(e|é|i|í|ú|u)(ć|č|q|k)(u|em|ima)"]
```

```
[word=".*(c|s|z)(i|í)j(i|om|ama)"]
```

All extracted tokens were entered into a table together with their corpus references. Each token was assigned a unique identifier and provided with contextual information in the form of left context, KWIC, and right context. The tokens were then manually filtered according to strict inclusion criteria. All tokens in case forms which are

³ The corpus is available on the NoSketch Engine platform: <https://www.clarin.si/ske/#dashboard?corpname=torlak>. Accessed: 28 May 2026.

not part of the core case system (nominative and general case) were excluded. This also applies to items retrieved due to accidental morphological overlap. Moreover, only nominalisations belonging to the target classes (*-nje/-će* and *-cija*) and showing a clearly identifiable relation to a source verb were retained. The final dataset consists of 606 target tokens and has been published as [Babić et al. \(2026\)](#).

The dataset was annotated for the properties relevant to the research questions. For each nominalisation, we recorded its morphological class, lemma, correspondence with standard Serbian, number, and the presence or absence of the complement types under investigation. In addition, the related verb was identified and annotated for aspect. Since the Timok speech lacks the infinitive form, source verbs are given in the feminine singular past participle form.

4. Results

The results are presented in the order of the research questions formulated above. The dataset contains 606 target tokens, i.e. *-nje/-će* and *-cija* nominalisations produced in the Timok variety, as well as 22 tokens which were annotated as standard due to their case form. Of 606 target tokens, 570 tokens belong to the *-nje/-će* class, while only 36 tokens belong to the *-cija* class. The dataset therefore shows a pronounced asymmetry in favour of *-nje/-će* nominalisations, which clearly constitute the dominant type in the corpus.

4.1. Aspectual distribution and nominalisation types

The first hypothesis concerned the quantitative relation between nominalisation type and the aspectual class of the source verb. More specifically, it predicted that imperfective *-nje/-će* nominalisations would be the most frequent type, while no clear ordering was assumed between perfective *-nje/-će* nominalisations and *-cija* nominalisations. This hypothesis is confirmed by both token and type frequencies.

At the token level, the dataset contains 606 deverbal nominalisations. Of these, 493 are imperfective *-nje/-će* nominalisations, corresponding to 81.3% of all tokens. Perfective *-nje/-će* nominalisations account for 77 tokens, or 12.7%, while *-cija* nominalisations account for 36 tokens, or 6%. The predicted dominance of imperfective *-nje/-će* nominalisations is therefore very clear.

The same pattern is found at the type level. Among the 173 classified types, 134 are imperfective *-nje/-će* nominalisations, corresponding to 77.5%; 22 are perfective *-nje/-će* nominalisations, corresponding to 12.7%; and 17 are *-cija* nominalisations, corresponding to 9.8%. Thus, imperfective *-nje/-će* nominalisations are not only the most frequent in actual corpus use, but also the most diverse lexically. At the same time, the token-level results show that perfective *-nje/-će* nominalisations are almost twice as frequent as *-cija* nominalisations, although this ordering was not part of the original prediction. This further underlines the marginal status of the *-cija* class in the corpus.

4.2. Complement realization

The second hypothesis concerned complement realisation. It predicted that overt complements, insofar as they occur, would be most frequent with imperfective *-nje/-će* nominalisations, since these forms are expected to preserve verbal argument structure to the greatest extent.

Overall, overt complements are rare in the dataset. They are illustrated in example (2).

(2)

- | | | |
|----------------------------|-----------------------------|-------------------|
| a. genitive complements: | <i>prázňovanje</i> | <i>velíkden-a</i> |
| | celebrating | Easter-GEN |
| | ‘the celebration of Easter’ | |
| b. accusative complements: | <i>prskánje</i> | <i>vinógrad-i</i> |
| | spraying | vineyards-ACC |
| | ‘spraying vineyards’ | |
| c. <i>na</i> -complement: | <i>kršténje na</i> | <i>déc-u</i> |
| | baptising na | children-ACC |
| | ‘christening of children’ | |

Genitive complements are attested in 12 tokens, corresponding to 2% of the full dataset. All of them occur with *-nje/-će* nominalisations: 5 with imperfective-derived forms and 7 with perfective-derived forms. No genitive complements are attested with *-cija* nominalisations. Split by class, genitive complements occur in 9.1% of perfective

-*nje/-će* tokens, 1% of imperfective -*nje/-će* tokens, while no genitive complements were found with -*cija* tokens.

At the type level, the pattern still remains marginal. Genitive complements are associated with 4.8% of deverbal nominalisation types overall, whereas the proportion remains higher among perfective -*nje/-će* types, at 17.3%, than among imperfective -*nje/-će* types, at 3.4%. Again, no genitive complements are attested with -*cija* types.

The much higher frequency of genitive complements with perfective -*nje/-će* nominalisations is clearly not in line with our hypothesis. However, the genitive data should be interpreted with particular caution. The genitive case is not a feature of the core Prizren-Timok case system, and quite a few of the attested examples appear to be borrowed or fixed collocations rather than freely generated realisations of argument structure. This is especially clear in cases such as *oproštenje gréhova* ‘forgiveness of sins’, *poznávanje prírode* ‘nature study’ (a school subject), and medical terms such as *zapalénje plúka* ‘pneumonia’ and *zapalénje mózga* ‘brain inflammation’. These examples should therefore not be treated straightforwardly as evidence for regular preservation of verbal argument structure inside the nominalisation. In that sense, they might actually serve as an indicator that the perfective nominalisations are a much more lexicalised domain containing many conventionalised terms, some of which contain other fixed parts.

Accusative complements are similarly rare. They are attested in 9 tokens, corresponding to 1.5% of the full dataset. Of these, 7 occur with imperfective -*nje/-će* nominalisations, 1 with a perfective -*nje/-će* nominalisation and 1 with a -*cija* nominalisation. Within each class, accusative complements occur in 1.4% of imperfective -*nje/-će* tokens, 1.3% of perfective -*nje/-će* tokens and 2.8% of -*cija* tokens.

At the type level, accusative complements are associated with 2.6% of deverbal nominalisation types overall. They occur with 3.1% of imperfective -*nje/-će* types, 0.9% of perfective -*nje/-će* types and 0.8% of -*cija* types. Notably, this distribution does follow the predicted ordering.

Finally, a prepositional complement introduced by *na* followed by the general case is attested only once, in the example *krštenje na décu* ‘christening of children’. This example occurs with an imperfec-

tive *-nje/-će* nominalisation and is particularly relevant because *na* + general case represents a construction more in line with the analytic structure of the Prizren-Timok case system than the genitive examples discussed above. However, since it is represented by only one token, it cannot support any broader quantitative conclusion.

Taken together, the complement data only partly support the second hypothesis. Imperfective *-nje/-će* nominalisations show the highest absolute number of complements and the broadest range of complement realisation. However, the proportional picture is less straightforward: genitive complements are proportionally more frequent with perfective *-nje/-će* nominalisations, while accusative and *na*-complements are proportionally highest with *-nje/-će* nominalisations (taking the type count into consideration). Importantly, however, a sizable portion of genitive examples appear to be fixed, borrowed, or terminological collocations rather than regular argument-structure realisations. The safest conclusion is therefore that overt complement realisation is rare overall, and that the evidence for preserved argument structure is limited, though it is found primarily with *-nje/-će* nominalisations, especially imperfective ones in absolute terms.

4.3. Pluralisation patterns

The third hypothesis concerned pluralisation. It predicted that plural forms (including the paucal forms) would be more frequent with *-cija* nominalisations and perfective *-nje/-će* nominalisations than with imperfective *-nje/-će* nominalisations. This prediction follows from the assumption that pluralisation reflects countability and lexicalisation, both of which indicate stronger integration into the nominal domain.

Plural forms are rare in the dataset as a whole. At the token level, they are attested in 25 out of 606 deverbal nominalisations, corresponding to 4.1% of the full dataset. Pluralisation is proportionally most frequent with *-cija* nominalisations: 8 of the 36 *-cija* tokens occur in the plural, corresponding to 22.2% of this class. By contrast, pluralisation is attested in only 2 of the 77 perfective *-nje/-će* tokens, corresponding to 2.6%, and in 15 of the 493 imperfective *-nje/-će* tokens, corresponding to 3%.

The type-level data show the same general tendency. Plural forms are associated with 4.3% of deverbal nominalisation types

overall. They are again most frequent among *-cija* types, at 21.4%. However, perfective *-nje/-će* types show the lowest proportion of pluralisation, at 1%, while imperfective *-nje/-će* types show a higher proportion, at 2.7%.

The results therefore only partially support the predicted pattern. The fact that *-cija* nominalisations show the highest rate of pluralisation at both token and type level is compatible with the view that they are more lexicalised, more readily countable, and more strongly integrated into the nominal domain. However, the second part of the prediction is not confirmed: perfective *-nje/-će* nominalisations are not more frequently pluralised than imperfective *-nje/-će* nominalisations. Instead, they show the lowest pluralisation rate in the dataset.

4.4. Standard influence

The fourth domain explored here concerned morphological correspondence with standard Serbian. The aim was not to test a directional prediction, but rather to determine to what extent the deverbal nominalisations attested in the corpus coincide morphologically with their standard Serbian counterparts, and whether any differences emerge across nominalisation classes.

Morphological correspondence with standard Serbian is high across the dataset. At the token level, 490 out of 606 nominalisations, corresponding to 80.9% of the dataset, have a form identical to the corresponding form in standard Serbian. By class, standard-identical morphology is attested in 61 perfective *-nje/-će* nominalisations, 400 imperfective *-nje/-će* nominalisations, and 29 *-cija* nominalisations. When considered proportionally within each class, correspondence with standard Serbian is highest among imperfective *-nje/-će* nominalisations, where it reaches 81.1%. It is slightly lower among *-cija* nominalisations, at 80.6%, and lowest among perfective *-nje/-će* nominalisations, at 79.2%.

The type-level data show somewhat greater differentiation. Overall, 72.4% of deverbal nominalisation types correspond morphologically to standard Serbian forms. The highest proportion is found among *-cija* types, at 81.5%, perfective *-nje/-će* types follow with 81.1%, while imperfective *-nje/-će* types show the lowest proportion, at 69.9%.

These results show that all three classes contain a high proportion of forms that are morphologically identical to their standard Serbian counterparts. At the same time, the somewhat lower type-level correspondence among imperfective *-nje/-će* nominalisations may reflect their greater productivity and lexical diversity in the dialect material. However, the available data do not allow for strong conclusions about the extent or direction of standard-language influence. The figures should therefore be treated as descriptive rather than explanatory. Moreover, the measure refers only to morphological correspondence in tokens assessed as potentially belonging to the Timok speech system; case forms interpreted as direct effects of standard Serbian influence were excluded from the calculation.

5. Discussion

The results largely confirm the hypotheses formulated above, especially with respect to the relationship between aspect and nominalisation type. Nominalisations in *-nje/-će* are predominantly derived from imperfective verbs, while perfective-derived forms are much less frequent. The *-cija* class is even smaller. The Timok data therefore follow the general pattern known from standard BCMS and other Slavic varieties: imperfective *-nje/-će* nominalisations constitute the central and most productive type as discussed in previous studies (Arsenijević, 2020, 2021; Kovačević, 2021; Simonović and Arsenijević, 2014, 2018).

The complement data provide a less straightforward picture. Overt complements are rare across the dataset, which is not surprising given the spontaneous spoken nature of the corpus. When they do occur, they are found primarily with *-nje/-će* nominalisations, especially imperfective-derived ones. This is compatible with the idea that these nominalisations preserve verbal argument structure to the greatest extent. However, the low number of examples prevents stronger conclusions. The occurrence of genitive complements also requires caution, since the genitive case is not part of the core Prizren-Timok case system and arguably reflects standard Serbian influence.

The pluralisation results are somewhat clearer than those concerning complements. Plural forms are proportionally most frequent with *-cija* nominalisations, which supports the hypothesis that this class is more lexicalised, more readily countable, and therefore more strongly integrated into the nominal domain. By contrast, plurali-

sation remains rare among both perfective and imperfective *-nje/-će* nominalisations, which suggests that these forms retain a stronger connection to eventive and process-related interpretation.

Finally, we found a high degree of morphological correspondence with standard Serbian, i.e. that many forms in the dataset are identical to their standard Serbian counterparts. However, the available material does not allow us to determine whether this reflects standard-language influence, shared inherited morphology, or independent dialect-internal formation. This question remains open and requires more targeted investigation.

6. Conclusion

This paper has presented a preliminary corpus-based study of deverbal nominalisations in the Timok speech, a non-standard Balkan Slavic variety whose morphosyntactic profile differs substantially from that of standard Serbian. The analysis focused on four properties: aspect, complement realisation, pluralisation, and morphological correspondence with standard Serbian.

The results show that *-nje/-će* nominalisations clearly dominate the dataset, especially those derived from imperfective verbs. This confirms that the Timok speech follows the general pattern known from standard BCMS, where imperfective-based nominalisations constitute the central type. At the same time, the relatively low frequency of *-cija* nominalisations suggests that this class is less prominent in the spoken dialectal material analysed here.

The study also shows that overt complements are rare, although they occur primarily with *-nje/-će* nominalisations. Pluralisation, by contrast, is proportionally most frequent with *-cija* nominalisations, supporting the view that this class is more lexicalised and more readily countable. Finally, the high degree of morphological correspondence with standard Serbian indicates that many forms are identical to their standard counterparts, but the available material does not allow for strong conclusions about the extent or direction of standard-language influence.

While our study has brought a number of relevant findings, it also has certain limitations. In several domains, we worked with very limited numbers, especially with regard to argument realisation, where, as we have seen, no form of argument realisation exceeds the 5% threshold. It should also be noted that our selection of the target

data allowed nominalisation tokens followed by case forms characteristic of standard Serbian, but excluded nominalisation tokens that were themselves in such a case form. Thus, we included *zapalénje mózga* ‘brain inflammation’, where the complement is in the genitive, but excluded, for example, *datum rođenja* ‘date of birth’, where the *-nje* nominalisation itself is in the genitive form. While this was necessary, it also limited the size and scope of our dataset.

Several directions for future research follow from these findings. First, the analysis should be extended to other speeches of the Prizren-Timok dialect. Second, a parallel study of deverbal nominalisations in spoken standard Serbian would help determine whether the patterns observed here are dialect-specific or characteristic of spontaneous speech more generally. Third, more detailed semantic and syntactic annotation would make it possible to investigate lexicalisation, complement realisation, information structure, and implicit arguments in greater depth. A diachronic perspective, based on older recorded materials, would also be valuable. Finally, given that the Timok speech is a Balkan Slavic variety, it would be interesting to investigate the properties of deverbal nominalisations in relation to Bulgarian and Macedonian (for the discussion of nominalisations in these two languages, see (Vulchanova-Dimitrova and Mitkovska, 2009)).

Acknowledgements

This paper is the result of the ÖAD WTZ project *What’s in a Verb? Mapping Serbian Verbs Borrowed into Romani (2024–2026)*. It also results from the research carried out at the Institute for Balkan Studies at the Serbian Academy of Sciences and Arts, funded by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, based on the Agreement on the Implementation and Financing of Scientific Research Work in 2026, no. 451-03-33/2026-03 dated February 5th 2026. Work on this paper was also supported by the Austrian Science Fund (FWF), project *Multifunctionality in morphology* (Grant DOI: 10.55776/I6258; PI: Marko Simonović; University of Graz).

Language Resources

- Babić, Anđela and Simonovic, Marko and Miric, Mirjana. 2026. *Annotated database of deverbil nominalisations in Timok Torlak*. Zenodo. Dataset.
- Milosavljević, Stefan and Babić, Anđela and Simonovic, Marko. 2025. *Mapping out the Torlak verb*. Zenodo. Dataset.
- Simonović, Marko and Kovačević, Predrag and Milićev, Tanja. 2025b. *Deverbil nominalisations and passive participles in Slavonic Serbian*. Zenodo. Dataset.
- Vuković, Teodora. 2020. *Spoken Torlak dialect corpus 1.0 (transcription)*. Slovenian Language Resource Repository CLARIN.SI. Language resource, ISSN 2820-4042.

References

- Artemis Alexiadou. 2001. *Functional Structure in Nominals*. John Benjamins, Amsterdam.
- Artemis Alexiadou. 2025. All Nominalizations, Big and Small: Towards the Typology of Complex Event Nominals. In *Proceedings of the Sixtieth Annual Meeting of the Chicago Linguistic Society (CLS 60)*, pages 17–36, Chicago, IL. Chicago Linguistic Society.
- Boban Arsenijević. 2007. Značenje srpskog imenskog sufiksa -je. In *Jezik, književnost, umetnost, knj. 1: Jezik i društvena kretanja*, pages 171–176, Kragujevac. FILUM.
- Boban Arsenijević. 2020. *Deverbil Nouns in -ie and their Variation across the South Slavic Area*. *Linguistica*, 60(1):7–29.
- Boban Arsenijević. 2021. Perfective Deverbil -je Nouns in Slavonic-Serbian. Paper presented at INTERSLAV, University of Graz, 24–26 February 2021.
- Noam Chomsky. 1970. Remarks on Nominalization. In *Readings in English Transformational Grammar*, pages 184–221. Ginn and Company, Waltham, MA.
- Victor A. Friedman and Brian D. Joseph. 2025. *The Balkan Languages*. Cambridge University Press, Cambridge.

- Jane Grimshaw. 1990. *Argument Structure*. MIT Press, Cambridge, MA.
- Jaklin Kornfilt and John Whitman. 2011. Afterword: Nominalizations in Syntactic Theory. *Lingua*, 121(7):1297–1313.
- Predrag Kovačević. 2021. [On the Internal Structure of Serbian -\(n\)je Nominalizations](#). *Acta Linguistica Academica*, 68(4):426–453.
- Stefan Milosavljević, Marko Simonović, Anđela Babić, Mirjana Mirić, and Jelena Stojković. 2026. [Torlak Theme Vowels in the Verbal Domain: A Corpus-based Study](#). In *Proceedings of the 1st ReLDI Conference on Language Science and Technology*, pages 48–57.
- Kristian Sandfeld. 1930. *Linguistique Balkanique: Problèmes et Résultats*. Number 31 in Collection linguistique publiée par la Société de Linguistique de Paris. Librairie ancienne Honoré Champion, Paris.
- Marko Simonović and Boban Arsenijević. 2014. [Regular and honorary membership: On two kinds of deverbal nouns in Serbo-Croatian](#). *Lingue e Linguaggio*, 13(2):185–210.
- Marko Simonović and Boban Arsenijević. 2018. [The Importance of not Belonging: Paradigmaticity and Loan Nominalizations in Serbo-Croatian](#). *Open Linguistics*, 4(1):418–437.
- Andrey N. Sobolev, Mirjana Mirić, Daria V. Konior, and Svetlana Ćirković. 2023. Torlak: Areal Embedding and Linguistic Characteristics. *Zeitschrift für Slavische Philologie*, 79(1):23–46.
- Nikolay S. Trubetzkoy. 1928. Proposition 16. In *Actes du Premier Congrès International des Linguistes*, pages 17–18, La Haye.
- Teodora Vuković. 2021. [Representing Variation in a Spoken Corpus of an Endangered Dialect: The Case of Torlak](#). *Language Resources and Evaluation*, 55(3):731–756.
- Mila Vulchanova-Dimitrova and Liljana Mitkovska. 2009. Nominalizations in Bulgarian and Macedonian. In *Investigations in the Bulgarian and Macedonian Nominal Expression*, pages 165–194. Tapir Academic Press, Trondheim.
- Петя Асенова. 2002. *Балканско езикознание: Основни проблеми на балканския езиков съюз*. Фабер, Велико Търново.
- Јакша Динић. 2008. *Тимочки дијалекатски речник*. Институт за српски језик САНУ, Београд.

Павле Ивић. 2001. *Дијалектологија српскохрватског језика: Увод и штокавско наречје*. Издавачка књижарница Зорана Стојановића, Сремски Карловци; Нови Сад.

Glagolske imenice u timočkom govoru: preliminarna korpusna studija

Anđela Babić, Marko Simonović, Predrag Kovačević

Sažetak

U radu se predstavlja preliminarna korpusna analiza glagolskih imenica u timočkom govoru, jednom od govora prizrensko-timočkog dijalekta srpskog jezika. Analiza je zasnovana na 606 primera dobijenih iz korpusa *Spoken Torlak Dialect Corpus 1.0* i fokusira se na dve produktivne morfološke klase: imenice na *-nje/-će* i imenice na *-cija*. Ispituju se četiri svojstva: vid glagola od kojeg je nominalizacija izvedena, realizacija dopuna, upotreba množine i morfološko poklapanje sa standardnim srpskim jezikom.

Rezultati pokazuju izrazitu dominaciju imenica na *-nje/-će*, naročito onih izvedenih od nesvršenih glagola. Imenice na *-cija* znatno su ređe u analiziranom materijalu. Izražene dopune pojavljuju se veoma retko, što upućuje na ograničeno očuvanje glagolske argumentske strukture u spontanoj govornoj građi. Ipak, kada su dopune potvrđene, one se uglavnom javljaju uz imenice na *-nje/-će*. Nasuprot tome, oblici množine proporcionalno su najčešći kod imenica na *-cija*, što potvrđuje pretpostavku da je ova klasa leksikalizovanja, brojiva i čvršće integrisana u imenički domen.

Većina potvrđenih oblika morfološki se poklapa sa odgovarajućim oblicima u standardnom srpskom jeziku. Ipak, raspoloživa građa ne omogućava pouzdane zaključke o tome da li je reč o uticaju standardnog jezika, zajedničkom nasleđenom obrascu ili unutardijalekatskoj tvorbi. Rad time proširuje istraživanje glagolskih imenica na nestandardni balkanoslovenski govor i na spontani govorni korpus, otvarajući prostor za dalja poređenja sa drugim prizrensko-timočkim govorima i sa govornim standardnim srpskim jezikom.

Ključne reči: glagolske imenice, timočki govor, prizrensko-timočki dijalekat, korpusna studija, vid, množina, dopune

Works are licensed under the terms and conditions of the Creative Commons Attribution 4.0 International License.

<https://creativecommons.org/licenses/by/4.0/>

