

Severing affixes from categories and borrowed derivations

Predrag Kovačević¹ Marko Simonović²

¹University of Novi Sad
predrag.kovacevic@ff.uns.ac.rs

²Universität Graz
marko.simonovic@uni-graz.at

BCGL 17: Categories & categorization
December 12–13, 2024

In a nutshell

- Central question: evidence for verbal structure in two classes of nominalisations related to borrowed verbs in Bosnian/Croatian/Montenegrin/Serbian (BCMS):
 - *edukacija* 'education' vs. *educiranje* 'educating'
- Method: corpus-based quantitative study focusing on 2 diagnostics:
 - genitive complements,
 - plural marking.
- Findings: native *-nje* nominals are more likely to have genitive complements and less likely to form plurals in comparison to Latinate *-cija* nominals.
- Analysis: *-nje* nominals are built from verbs, while *-cija* nominals lack verbal structure.

Derivational affixes as roots?

Prime examples are affixes that are:

- categorially flexible,
- prosodically active,
- semantically unpredictable

Such are:

- English *-ic* and *-al*, Lowenstamm (2014),
- Dutch *-iek* and *-aal*, Creemers et al. (2018)

These affixes:

- originate from contact (Romance/Latinate)
- are flexible between As and Ns

Derivational affixes as roots?

Creemers et al. (2018) offer a tentative solution for the restriction to flexibility between As and Ns in Dutch:

- Dutch has the 'greedy' suffix *-eer*, which always spells out the verbal category, so it does not leave any space for a lexical suffix (i.e. a root) to show up as cross-categorial.
- Issue: restriction to the Latinate domain.

BCMS Latinate verbs

Comparable to Dutch because they have an obligatory derivational affix following Latinate roots:

- in Croatian consistently *-ir-*.
- In Serbian variably *-ir-*, *-is-*, *-ov-*.

Different from Dutch because the derivational suffix is followed by the theme vowel (*-a-* in the examples below), which is arguably the spellout of the verbal category.

Croatian	Serbian	Gloss
imit-ir-a-ti	imit-ir-a-ti	'imitate'
manipul-ir-a-ti	manipul-is-a-ti	'manipulate'
dokument-ir-a-ti	dokument-ov-a-ti	'document'

Table 1: Latinate verbs in BCMS

Nominalizing Latinate verbs in BCMS

Latinate verbs are generally imperfective or bispectual. Such verbs in BCMS productively derive eventive nominalizations in *-nje*.

Serbian verb	Gloss	Serbian nominalization	Gloss
imit-ir-a-ti	'imitate'	imit-ir-a-nje	'imitating'
manipul-is-a-ti	'manipulate'	manipul-is-a-nje	'manipulating'
dokument-ov-a-ti	'document'	dokument-ov-a-nje	'documenting'

Table 2: Latinate deverbal nominalisations in BCMS

These nominalisations often compete with Latinate nominalisation, e.g., *imitacija*, *manipulacija* and *dokumentacija*.

Closer look into *-nje* nominalisations in BCMS

Simonović and Arsenijević (2014): *-nje* nominalisations fall into 2 types:

- paradigmatic
 - productive, compositional, prosodically faithful to the verb
- non-paradigmatic
 - non-productive, carrying unpredictable meaning and shifted prosody.

While the main division line is between imperfective/bisapectual verbs (deriving paradigmatic nominalizations) and perfective verbs (deriving non-paradigmatic nominalizations), there are also some minimal pairs.

Verb	Paradigmatic	Non-paradigmatic
'imáti 'have'	'imá:nje 'having'	i'ma:né 'property'
'obrázovati 'educate'	'obrázova:nje 'educating'	obrazo'ua:né 'education'

Arsenijević (2020); Simonović (2022): Non-paradigmatic items do not contain internal categorizing heads.

Latinate nominalisations are non-paradigmatic

Simonović (2015); Simonović and Arsenijević (2018): Latinate loans like *manipulacija* and *imitacija*

- enter the lexicon as simplex items (despite apparent complexity)
- get analysed within the lexicon, e.g., *manipul-a-c-ij-a*.
- join the non-paradigmatic component of the lexicon.

Latinate nominalisations are non-paradigmatic?

Latinate nominalisations diagnose as clearly non-paradigmatic in terms of:

- Prosody: all items in *-acija* have neutralised prosody: *edukacija* [edu'ka:tsíja], *manipulacija* [manupu'la:tsíja], unlike *-nje* nominalisations, which preserve the prosody of the verb.
- Productivity: *telefonirati* and *programirati* do not produce **telefonacija* and **programacija*.

Latinate nominalisations are non-paradigmatic?

However, it is not obvious that all pairs of *-nje* vs. *-cija* nominalizations have distinct syntax and semantics:

- both allow genitives expressing both internal and external arguments
- both tolerate pluralisation.

stalna **manipuliranja** predstavnika vlade

stalne **manipulacije** predstavnika vlade

'constant manipulations of/by the representatives of the government'

Focus and Research questions

Pairs of *-cija* and *-nje* nouns derived from the same Latinate root:

- *edukacija, manipulacija* etc.
- *educiranje, manipuliranje* etc.

Q: Are *-nje* nominals more verbal?

Method

Diagnostics (per Grimshaw 1990):

- the presence of a genitive complement (verbal)
- the presence of a plural suffix (non-verbal)

Data collection:

- pairs of *-cija* and *-nje* nominals
 - both judged by four native speakers to allow the eventive meaning, genitive complements and pluralisation.
 - both with more than 3000 tokens in the respective CLASSLA-web corpus (Ljubešić et al., 2024)
- random sample of 50 tokens per nominal
- each token annotated for:
 - genitive complement,
 - plural morphology.

Two samples

- Croatian sample from CLASSLA-web.hr
- 11 pairs:
 - promoviranje – promocija
 - kombiniranje – kombinacija
 - provociranje – provokacija
 - komuniciranje – komunikacija
 - educiranje – edukacija
 - registriranje – registracija
 - aktiviranje – aktivacija
 - investiranje – investicija
 - instaliranje – instalacija
 - prezentiranje – prezentacija
 - saniranje – sanacija

Two samples

- Serbian sample from CLASSLA-web.sr (Katić et al., in preparation)
- 20 pairs:
 - aktiviranje – aktivacija
 - regrutovanje – regrutacija
 - projektovanje – projekcija
 - iniciranje – inicijacija
 - regulisanje – regulacija
 - eliminisanje – eliminacija
 - konzerviranje – konzervacija
 - formulisanje – formulacija
 - redukovanje – redukcija
 - identifikovanje – identifikacija
 - imitiranje – imitacija
 - modifikovanje – modifikacija
 - instaliranje – instalacija
 - variranje – varijacija
 - organizovanje – organizacija
 - rezervisanje – rezervacija
 - registrovanje – registracija
 - simuliranje – simulacija
 - reprodukovanje – reprodukcija
 - manipulisanje – manipulacija

Results

Genitive complements are significantly more frequent with 'native' *-nje* nominals

Plural forms are significantly more frequent with Latinate *-cija* nominals

suffix	gen_comp	plural
<i>-acija</i>	190	117
<i>-nje</i>	357	1

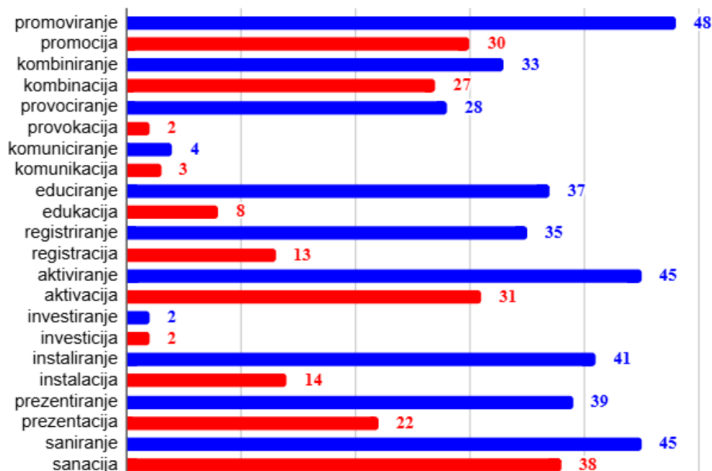
Table 3: Number of tokens with genitive complements and plural forms across the two types of nouns in the Croatian sample

suffix	gen_comp	plural
<i>-acija</i>	578	209
<i>-nje</i>	938	24

Table 4: Number of tokens with genitive complements and plural forms across the two types of nouns in the Serbian sample

Results

Virtually all pairs follow the same tendency, very few ties and no pairs that show the opposite tendency. These are the results for the genitive complement in the Croatian sample.

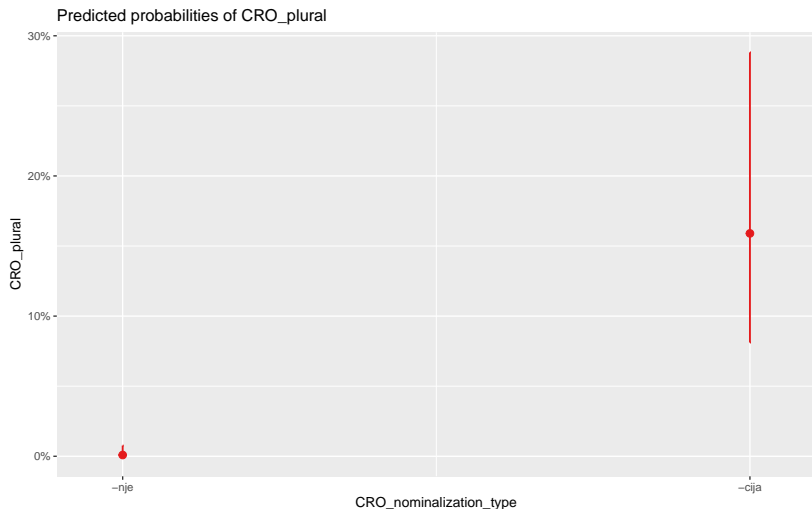


Results: The Likelihood of Plural Marking Based on Nominalization Type (Croatian sample)

Term	Estimate	Std. Error	z Value	p-Value
(Intercept)	-7.01	1.07	-6.54	6.08×10^{-11} ***
loan_type_Latinate	5.34	1.01	-5.31	1.12×10^{-7} ***

Table 5: Binomial Generalized Linear Mixed-Effects Model predicting the likelihood of Plural Marking based on Nominalization Type (Croatian sample).

Results: Predicted probabilities of plural based on nominalization types in the Croatian sample



Results: The Likelihood of Genitive Complement Based on Loan Type

Term	Estimate	Std. Error	z Value	p-Value
(Intercept)	0.76	0.5	1.52	0.13
loan_type_Latinate	-1.79	0.16	-11.08	$< 2 \times 10^{-16}$ ***

Table 6: Binomial Generalized Linear Mixed-Effects Model predicting the likelihood of genitive complement based on loan type in the Croatian Sample.

Results: Predicted probabilities of genitive based on loan types in the Croatian sample

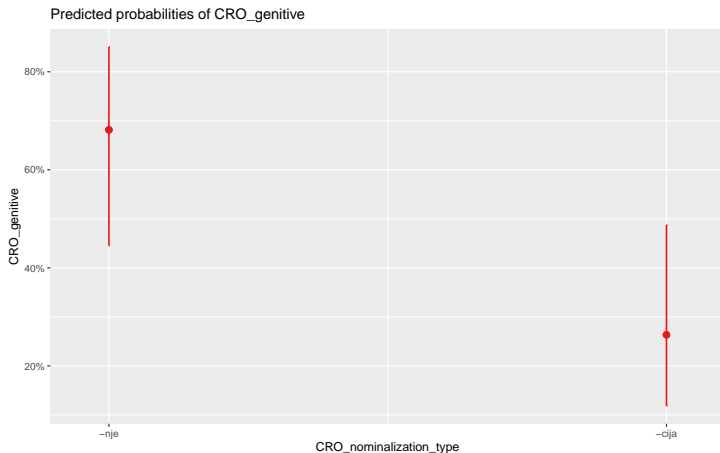


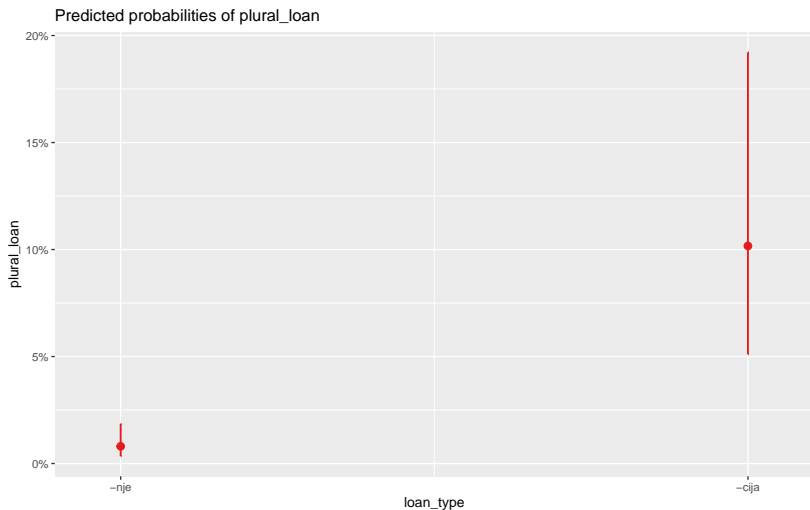
Figure 2: Predicted probabilities of genitive based on loan types in the Croatian sample.

Results: The Likelihood of Plural Marking Based on Loan Type (Serbian sample)

Term	Estimate	Std. Error	z Value	p-Value
(Intercept)	-4.82	0.44	-11.05	$< 2 \times 10^{-16}$ ***
loan_type_Latinate	2.64	0.23	11.43	$< 2 \times 10^{-16}$ ***

Table 7: Binomial Generalized Linear Mixed-Effects Model predicting the likelihood of plural based on loan type (Serbian sample).

Results: Predicted probabilities of plural based on nominalization types in the Serbian sample

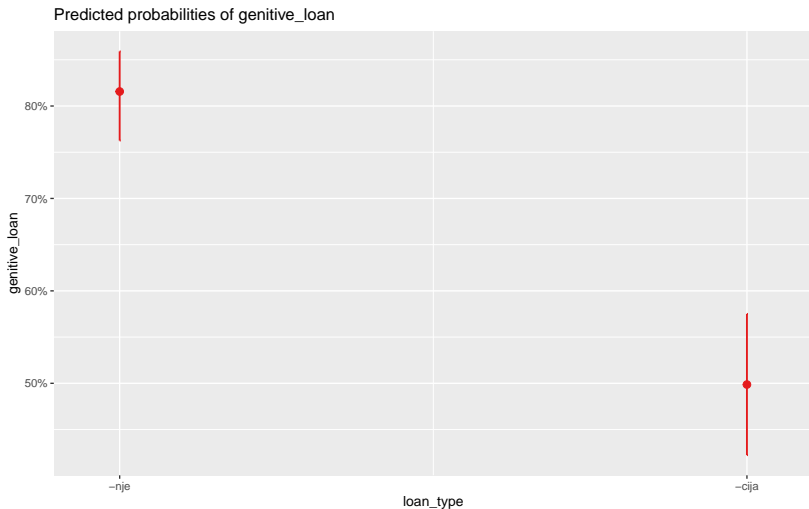


Results: The Likelihood of genitive complement Based on Loan Type (Serbian sample)

Term	Estimate	Std. Error	z Value	p-Value
(Intercept)	1.49	0.16	9.07	$< 2 \times 10^{-16}$ ***
loan_type_Latinate	-1.49	0.09	-15.35	$< 2 \times 10^{-16}$ ***

Table 8: Binomial Generalized Linear Mixed-Effects Model predicting the likelihood of genitive complement based on loan type (Serbian sample).

Results: Predicted probabilities of genitive based on the nominalization type in the Serbian sample



Native contrast?

If the contrast between *-nje* and *-cija* nominals truly mirrors the native contrast between paradigmatic and non-paradigmatic nominalisations, we expect pairs of native *-nje* nominalisations to show the same tendencies. In the native domain, the main contrast is between imperfective and perfective bases.

Gloss	IPFV	Nominal	PFV	Nominal
'destroy'	uništavati	uništavanje	uništiti	uništenje
'liberate'	oslobađati	oslobađanje	osloboditi	oslobođenje

Data collection (Katić et al., in preparation):

- 20 pairs of *-nje* nominals with IPFV and PFV bases
 - both judged by four native speakers to allow the eventive meaning, genitive complements and pluralisation.
 - both with more than 4000 tokens in CLASSLA-web.rs (Ljubešić et al., 2024)
- random sample of 50 tokens per nominal
- each token annotated for: genitive and plural.

Results

Genitive complements are significantly more frequent with imperfective native *-nje* nominals.

Plural forms are significantly more frequent with perfective native *-nje* nominals.

suffix	gen_comp	plural
<i>-perfective</i>	667	103
<i>-imperfective</i>	986	14

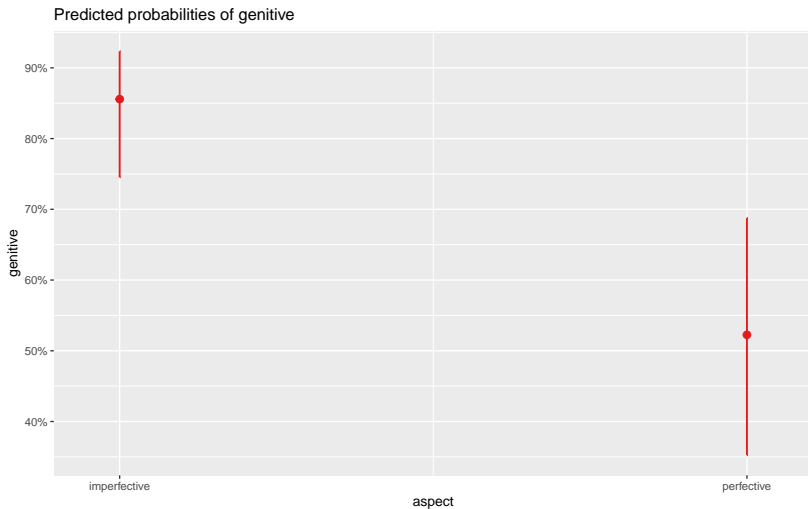
Table 9: No. of tokens of genitive complements and plurals across the two types of native *-nje* nominals (Serbian data only)

Results: The Likelihood of Plural Based on Aspect of the Base

Term	Estimate	Std. Error	z Value	p-Value
(Intercept)	-7.39	1.01	-7.29	3.11×10^{-13} ***
aspect_perfective	2.59	0.31	8.26	$< 2 \times 10^{-16}$ ***

Table 10: Binomial Generalized Linear Mixed-Effects Model predicting the likelihood of plural based on the aspect of the base.

Results: Predicted probabilities of genitive based on Aspect (Serbian sample)

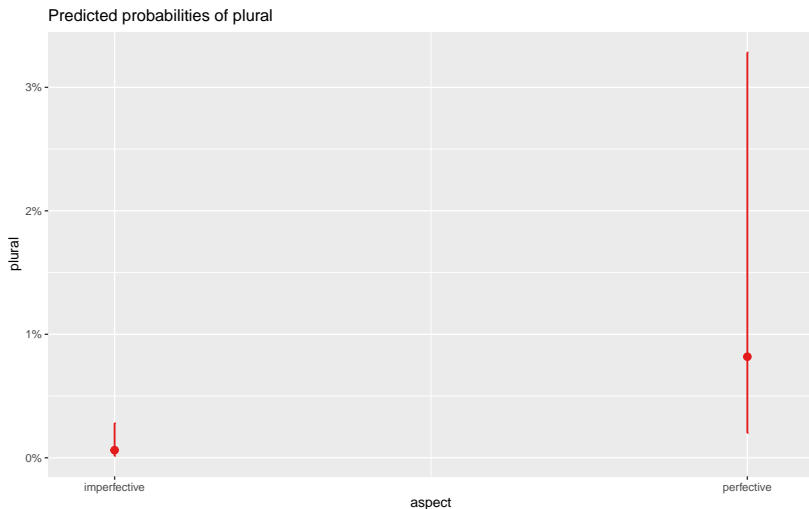


Results: The Likelihood of Genitive Based on Aspect of the Base

Term	Estimate	Std. Error	z Value	p-Value
(Intercept)	1.78	0.36	4.93	8.02×10^{-7} ***
aspect_perfective	-1.69	0.11	-14.14	$< 2 \times 10^{-16}$ ***

Table 11: Binomial Generalized Linear Mixed-Effects Model predicting the likelihood of genitive based on the aspect of the base.

Results: Predicted probabilities of plural based on the aspect of the base in the Serbian sample



Discussion

Emergent picture from the corpus analysis:

- Genitive complements are much more frequent with *-nje* nouns,
- Plural forms are a lot more frequent with *-cija* nouns.
- **Neither genitives nor plurals are completely excluded with either type of nominals.**

Explanation:

- *-nje* nominals include a categorizing *v* below the suffix
- *-cija* nouns lack internal verbal structure,
- morphemes that *-cija* nominals consist of are not categorizers but roots (Lowenstamm, 2014)

Discussion: Genitive complements

- The genitive case on complements of nouns is not structurally licensed by *v*.
- The genitive case is assigned to an argument of a noun iff a noun has an argument (McIntyre, 2014).
- Since *-nje* nouns are built from verbs, they always inherit the verb's eventive semantics and its internal argument.
- If the argument of the underlying verb is not overtly present, it is implied.

Discussion: Plural forms

- *-nje* nominals derived from imperfective/biaspectual verbs are atelic allowing progressive and iterative readings, which makes pluralization redundant (Iordăchioaia and Soare, 2009; Alexiadou, 2010) .
 - Hence, rare plurals.
- *-cija* nominals are telic if they denote events or they denote entities.
 - In both cases, a plural form is needed to signal pluralities. Hence, frequent plurals.

How borrowing introduced complex root nominalisations

- In a system that has a productive native pattern for eventive nominalisations, borrowed nominals will be pushed towards other meanings.
- Obligatory verbalising morphology decreased the morphological overlap between Latinate verbs and Latinate nominals.

Thanks!

References I

- Alexiadou, A. (2010). On the morphosyntax of (anti)-causative verbs. In M. R. Hovav, E. Doron, and I. Sichel (Eds.), *Lexical semantics, syntax, and event structure*. Oxford Scholarship Online.
- Arsenijević, B. (2020). Deverbal nouns in -ie and their variation across the south slavic area. *Linguistica* 60(1), 7–29.
- Creemers, A., J. Don, and P. Fenger (2018). Some affixes are roots, others are heads. *Natural Language Linguistic Theory* 36, 45–84.
- Grimshaw, J. (1990). *Argument structure*. MIT Press.
- lordăchioaia, G. and E. Soare (2009). Structural patterns blocking plural in Romance nominalizations. In E. O. Aboh, E. van der Linden, J. Quer, and P. Sleeman (Eds.), *Romance Languages and Linguistic Theory: Selected Papers from 'Going Romance' Amsterdam 2007*, pp. 145–160. John Benjamins Publishing Company.

References II

- Ljubešić, N., P. Rupnik, and T. Kuzman (2024). Croatian web corpus classla-web.hr 1.0. Slovenian language resource repository CLARIN.SI.
- Lowenstamm, J. (2014). Derivational affixes as roots. In A. Alexiadou, H. Borer, and F. Schäfer (Eds.), *The Syntax of Roots and the Roots of Syntax*, pp. 230–259. Oxford University Press.
- McIntyre, A. (2014, Mar). Constraining argument structure in nominalizations: The case of english -er. *Lingua* 141, 121–138.
- Simonović, M. (2022). Derivational affixes as roots across categories. *Journal of Slavic linguistics* 30(2), 195–233.
- Simonović, M. (2015). *Lexicon immigration service - Prolegomena to a theory of loanword integration*. Ph. D. thesis, Utrecht University.
- Simonović, M. and B. Arsenijević (2014). Regular and honorary membership: On two kinds of deverbal nouns in Serbo-Croatian. *Lingue e Linguaggio* 13(2), 185–210.

References III

Simonović, M. and B. Arsenijević (2018). The importance of not belonging: Paradigmaticity and loan nominalizations in serbo-croatian. *Open Linguistics* 4(1), 418–437.