volume 32, number 1

winter–spring 2024

Journal of the Slavic Linguistics Society

Slavic Linguistics Society

The Slavic Linguistics Society is an organization devoted to the systematic and scholarly study of the Slavic languages. Membership in the Society is open to all linguists regardless of field, theory, framework, or school. Members are entitled to present papers at annual meetings and receive the *Journal of Slavic Linguistics*, which is the official journal of the Slavic Linguistics Society. Individuals receive a subscription to *JSL* with their annual membership in the Slavic Linguistics Society. To join, go to http://www.slaviclinguistics.org.

SLS is governed by an Executive Board consisting of six members elected at the annual meeting for three-year staggered terms. The Executive Board negotiates arrangements for the annual meetings with host institutions. Meetings typically alternate between North America and Europe.

Executive Board

Guidelines for Contributors

JSL is intended to address issues in the description and analysis of Slavic languages of general interest to linguists, regardless of theoretical orientation. Papers dealing with any aspect of synchronic or diachronic Slavic phonetics, phonology, morphology, syntax, semantics, or pragmatics will be considered, as long as they raise substantive problems of broad theoretical concern or propose significant descriptive generalizations. Comparative studies and formal analyses are especially encouraged. Submissions ranging from full-length articles to shorter remarks and replies will be considered, along with substantive reviews.

Manuscripts should be submitted through the journal's website, where submission guidelines are available, as well as online versions of volumes starting with volume 19 (2011).

JSL's World Wide Web page: c

The *Journal of Slavic Linguistics* publishes two regular issues per year, totaling approximately 360 pages: Winter–Spring and Summer–Fall. Each year *JSL* also publishes an extra open-access, online-only issue that contains the proceedings from the Formal Approaches to Slavic Linguistics (FASL) conference. For regular issues, library and institutional subscriptions are available from Slavica Publishers at the following rate:

Institutions $60.00

Individuals should join the Slavic Linguistic Society (http://www.slaviclinguistics.org), which carries with it an annual subscription to *JSL*.

Airmail supplement for institutions outside of North America: $11.00/year.

A limited number of back issues are available for $24.00 apiece.

Indexing and Abstracting: *American Bibliography of Slavic and East European Studies, ERIH (European Reference Index for the Humanities and Social Sciences), Humanities International Index, IBZ (Internationale Bibliographie der Geistes- und Sozialwissenschaftlichen Zeitschriftenliteratur), MLA International Bibliography (Modern Language Association), OCLC ArticleFirst, and Web of Science Emerging Sources Citation Index*

This issue was published in May 2025.

ISSN: 1068-2090

Technical Editors: Jordan Hussey-Andersen and Ellen Tamura

Cover design: Nikki Life

Slavica Publishers
Indiana University
1430 N. Willis Drive
Bloomington, Indiana USA
47404-2146

[Tel.] 1-812-856-4186
[Fax] 1-812-856-4187
[Toll-free] 1-877-SLAVICA
slavica@iu.edu
https://slavica.indiana.edu/

journal of slavic linguistics

volume 32, number 1

winter–spring 2024

From the Editors vii

articles

Varvara Magomedova, Kirill Chuprinko, and Natalia Slioussar
Modeling Gender Variation in Russian Indeclinable Nouns: Optimality over Structuralism, Hierarchical MaxEnt, and Degrees of Idiosyncrasy 1

Maria Martynova, Yulia Zuban, Luka Szucsich, and Natalia Gagarina
OV/VO Word Order in Heritage Russian: Is Transfer at Play? 27

Igor Mel'čuk and Jasmina Milićević
The Clausative-Presentative Construction in Russian and Serbian 61

Zorica Puškar-Gallien
Disassembling and Reassembling Pronouns: A Case Study of Bosnian/Croatian/Montenegrin/Serbian 103

reviews

Julia Bacskai-Atkari
Ludmila Veselovská
Wh-Questions: A Case Study in Czech 145

Darya Kavitskaya
Janina Mołczanow
Interactions of Vowel Quality and Prosody in East Slavic 151

Zrinka Kolaković
Jasmina Milićević
Serbian Clitics 159

From the Editors

This is the first issue of volume 32, the volume of 2024.

We are excited to announce that JSL is introducing an "Online First" section. Articles that have been accepted but have not yet appeared as part of an issue will appear, typically in a pre-typeset version, on our website. Once the issue featuring the article is published, the article will be removed from the "Online First" section; the article URL will remain the same.

This issue also marks a change in the editorial team: we welcome Jaye Padgett of the University of California, Santa Cruz, as a new Associate Editor.

Franc Marušič and Rok Žaucer
University of Nova Gorica
franc.marusic@ung.si & rok.zaucer@ung.si

Modeling Gender Variation in Russian Indeclinable Nouns: Optimality over Structuralism, Hierarchical MaxEnt, and Degrees of Idiosyncrasy

Varvara Magomedova, Kirill Chuprinko, and Natalia Slioussar

Abstract: In this paper, we provide an analysis of the grammatical gender of 131 inanimate indeclinable Russian nouns based on the data from the General Internet Corpus of the Russian Language. We demonstrate that most nouns show substantial variation, being used in two or even in all three genders: masculine, feminine, and neuter. We identify several factors affecting this, primarily the gender of the semantic analogy noun and the root-final vowel. We argue that these data can be used to compare several major morphological frameworks and conclude that some approaches, namely optimality-theoretic probabilistic ones, are better suited to account for them. We also compare different models within the chosen set of approaches and show that the hierarchical Maximum Entropy (MaxEnt) models are superior to the classical MaxEnt models.

1. Introduction*

Russian nouns are inflected for number and case and are divided into several declensions or inflectional classes based on their affixes. They are also divided into three genders: masculine, feminine, and neuter (M, F, and N). The gender of a noun closely correlates with its inflectional class, especially in the case of inanimate nouns.

However, there is a small group that does not conform to this generalization: indeclinable inanimate nouns. In this paper, we aim to analyze their grammatical gender based on corpus data. As a result, we reveal extensive variation both in the group as a whole and in individual nouns, and the factors affecting this variation. Our second goal is to show that these data can be used to compare several major morphological frameworks; we argue that some of

* The work on this project was carried out in the framework of the Basic Research Program at the HSE University, Russia. Kirill Chuprinko's research was also partially supported by the Slovenian Research Agency grant no. J6-4615.

them, namely optimality-theoretic probabilistic approaches, are better suited to account for such data than the others. Our third goal is to compare different models within the chosen set of approaches. We argue that the hierarchical Maximum Entropy (MaxEnt) model is superior to the classical MaxEnt models if some idiosyncrasies in the data are properly taken into account.

The paper has the following structure. After providing more information on genders, declensions, and indeclinability in Russian, we briefly formulate the theoretical questions we aim to address, introduce the corpus data we collected, and proceed to discuss them in light of different frameworks. Finally, we compare several optimality-theoretic models of our data.

2. Genders, Declensions, and Indeclinability in Russian

There are several approaches to Russian nominal declensions. Our data do not bear on this question, so we will rely on a widely accepted system with three classes, presented in Table 1 (e.g., Aronoff 1994; Halle 1994; and Shvedova, 1980). Inflections in the classes IIa and IIb are the same in all cases except for nominative and accusative, which is why some authors unite them, as in Table 1, and others do not. On the other hand, inflections in the classes IIa and III are the same only in nominative and accusative (in inanimate nouns). Classes that have distinct inflections in the singular share the same inflections in the plural, and there is no gender agreement, so we do not consider plural forms in this paper.

Table 1. The distribution of Russian nouns by declension and gender[1]

Declension	Gender	NOM.SG inflection	% in RNC	Examples
I	F	*-(j)a*	29% nouns	*mama* 'mom', *zemlja* 'earth'
I	M (animate)	*-(j)a*	1% nouns	*papa* 'dad', *djadja* 'uncle'
IIa	M	∅	46% nouns	*stol* 'table', *gel'* 'gel'
IIb	N	*-o/e*	18% nouns	*vino* 'wine', *pole* 'field'
III	F	∅	5% nouns	*myš'* 'mouse', *mel'* 'shallow'
indeclinable	different genders	—	1% nouns	*kino* 'cinema', *kofe* 'coffee'

[1] Percentages of nouns in the Russian National Corpus, or RNC (www.ruscorpora.ru), are taken from Slioussar and Samoilova (2015). Their counts were based on the grammatically disambiguated subcorpus and did not take substantivized adjectives into account (e.g., *moroženoe* 'ice cream'). Among the nouns with a zero inflection in the nominative singular, feminine nouns in the III declension end in a palatalized or

Table 1 shows that the connection between gender and declension is pervasive but not absolute and does not work at all for indeclinable nouns (i.e., indeclinability is not associated with a particular gender). This table also does not include several minor noun groups that complicate the picture even further. There are different theoretical approaches to the problem of gender and declension (e.g., Caha 2021; Corbett and Fraser 2000; Kramer 2015; and Rice 2005), but we will not discuss them in this paper because our data do not let us tease them apart. In any case, because of this problem and because the affixes in different classes coincide in some forms, the gender of the noun can be unambiguously determined only from agreeing adjectives, participles, and verb forms, or from the choice of pronouns.[2]

All declinable Russian nouns have stems ending in a consonant. Most derivational affixes are associated with a particular declension, but if a noun stem does not have any, it is impossible to decide based on its final consonant to which declension and gender it belongs. There are certain tendencies and restrictions (e.g., see the comments to Table 1), but it is still true that stems in classes I and IIa may end in any consonant. If a noun is inanimate, it is also absolutely impossible to predict its declension and gender based on its semantics.[3] As we show below, these fundamental generalizations do not hold for indeclinable nouns.

Indeclinables are mostly loanwords ending in *-o*, *-e*, *-(j)a*, *-i*, and *-(j)u* vowels. Several indeclinable nouns end in *-y* and *-è*, but they are infrequent—e.g., *Janczy* '(the river) Yangtze', *Xuanxè* 'Huang He, (the Yellow River)'—so we do not analyze them in this paper. If a loanword ends in a consonant, eventually it always becomes a class IIa masculine noun or a class III feminine noun.[4] Most loanwords ending in *-(j)a* become class I feminine nouns, i.e., the final vowel gets reanalyzed as a NOM.SG inflection, but there are exceptions. In some cases, the reason is clear; since all declinable Russian noun stems end in a consonant, loanwords ending in a hiatus, like *kinoa* 'quinoa', become indeclinable,

alveolo-palatal consonant, while masculine nouns in the IIa declension may end in any consonant. As for *(j)a,* the Russian alphabet has two separate letters (Я and Ю) that are transliterated in the Latin alphabet as *ja* and *ju*: they convey the same vowels as the letters *a* and *u* but change the articulation of the preceding consonant, introduce the consonant /j/, or are purely a matter of orthographic conventions.

[2] Only the latter option is available to differentiate between masculine and neuter nouns in any form except for the nominative and accusative singular: all affixes coincide not only in nominal, but also in adjectival paradigms.

[3] For animate nouns, especially those denoting people, the grammatical gender is closely connected with the biological sex and/or the social gender of the referent.

[4] The former option is less restricted phonologically and in general much more frequent. In addition to that, many female first and last names that end in a consonant are indeclinable.

unless *j* is appended to the stem. For example, *Ikea*, the brand name, is usually pronounced as *Ikeja* in colloquial Russian and declines as a class I feminine noun. The other cases are more mysterious, like *kannada* 'Kannada (language)'. There is no hiatus, but the word does not decline.

Nouns ending in *-(j)u, -è, -y* and *-i* (like *tabu* 'taboo', *mjusli* 'muesli') do not fit into any existing declension and are doomed to be indeclinable (although *-y/i* can be reanalyzed as a plural affix yielding a pluralia tantum noun). Nouns ending in *-o* and *-e* (like *pal'to* 'coat', *žele* 'jelly') are the most noteworthy: they could be reanalyzed as class IIb neuter nouns, but this almost never happens.[5] Moreover, some native proper names and toponyms ending in *-o,* like *Ivanovo* (an old Russian city), show variation in declinability. In general, although the group of indeclinables is very small in Russian, it is still much larger than in other Slavic languages (Mučnik 1971: 256, Thomas 1983; Swan 2002; and Shigemori Bučar 2011)—an observation that still awaits its explanation and more precise calculations.

Another reason why indeclinables are interesting is that they prefer the neuter gender, but all three genders are represented in this set. According to a dictionary study by Murphy (2000), 67% of indeclinable nouns are listed in various dictionaries as neuter, 17% as masculine, 8% as feminine, and the remaining nouns demonstrate gender variation. Our corpus study shows much more extensive variation in real language use, but neuter is still the most frequent gender. This is remarkable, given that in declinable nouns, masculine is the most frequent and neuter the least frequent, as Table 1 shows. Magomedova and Slioussar (2023) try to explain this, developing a novel approach to the problem of gender markedness in Russian. At the same time, the fact that neuter is not the only choice for such nouns is hard to explain in any model studying the connection between genders and declensions.

Finally, indeclinable nouns are remarkable because their gender assignment was shown to be affected by their semantics and by their stem-final segment. Both factors were noted in the previous studies based on individual examples and some experimental data, while we conducted a corpus study to estimate their role based on a large dataset of naturally occurring examples. Speaking of the semantic factor, there are cases when the gender of a semantically related declinable noun influences gender assignment of an indeclinable (Rozental' et al. 1998; Murphy 2000; Galbreath 2010; Savchuk 2011, etc.). The nature of these semantic relations is difficult to formalize; they may involve synonymy, hypernymy, or even sporadic conceptual matches. Follow-

[5] For example, the word *èxo* 'echo', clearly borrowed a long time ago, is declinable, although some examples of indeclinable usage are reported by Corbett (2023). At the same time, for many nouns that are indeclinable in modern standard Russian, some declinable forms have been registered in non-standard variants—for example, see Henry (2020) for the history of the word *pal'to* 'coat'.

ing Corbett (1991), we call these relations *semantic analogy*. For example, *avenju* 'avenue' is feminine in the dictionaries and is often used with feminine agreement in our dataset because the synonymous noun *ulica* 'street' is feminine. *Mango* 'mango' and *kivi* 'kiwi' are listed as neuter in the dictionaries, but our dataset contains numerous examples of masculine agreement because their hypernym *frukt* 'fruit' is masculine.

The semantic analogy effect was noted to influence the gender assignment of loanwords cross-linguistically. Besides the case of Russian indeclinables, Corbett (1991: 75–82) provides examples from Archi (Nakh-Dagestanian), Hausa (Afro-Asiatic), and Polish (Slavic). However, as far as we can judge, Russian indeclinables are different in the following way. In all these languages, semantic analogy affects gender assignment when the loanword is adapted to the new language (which affixes and agreement patterns it will be associated with, how its phonological form will be changed), but then its role is the same as for any other noun in this language—in particular, it plays no role for inanimate nouns in Hausa and Polish. Russian also has similar examples—they are briefly discussed at the beginning of §4. However, the situation is different for Russian inanimate indeclinable nouns: semantic analogy affects gender assignment even in the nouns that were borrowed a long time ago and are listed in dictionaries, competing with other factors and resulting in massive gender variation that is not attested for any declinable Russian nouns.[6]

The second major factor is morphophonological analogy with declinable nouns. Some stem-final vowels of indeclinables resemble NOM.SG (*-(j)a*, *-o*, *-e*) and NOM.PL (*-i*) affixes. In the dictionary data overview, Murphy (2000: 109–14) discussed how indeclinables ending in *-o* and *-e* are more likely to be neuter, while the ones ending in *-(j)a* show a higher share of feminine gender than other indeclinables. This is confirmed in an experimental study by Wang (2014) and by some observations by Mjakilja (2000). Savchuk (2011) discusses the role of morphophonological analogy for nouns like *spagetti* 'spaghetti' and *mjusli* 'muesli' that are mostly used as pluralia tantum. What is especially interesting here is that morphophonological analogy works despite the fact that the nouns remain indeclinable, i.e., their final vowels do not get reanalyzed as inflectional affixes.

The fact that indeclinable nouns consist of a root alone and yet have highly variable gender that appears to depend on the semantic and phonological properties of that root poses several theoretical questions. The main two questions we address in this paper are the following: which theoretical approach can handle these data and how to formalize the analysis? Namely,

[6] Some declinable inanimate nouns show gender variation, but it is limited (e.g., Savchuk 2011). Animate nouns that can be used with masculine and feminine agreement are analyzed in terms of common gender rather than gender variation (the choice of grammatical gender depends on the biological or social gender of the referent).

which formal analysis methods of the selected theoretical approach predict our data distribution the most accurately? In the next section, we present the data in more detail and then turn to these questions.

3. Corpus Study

3.1. Data Collection and Preliminary Analysis

We collected data from the LiveJournal subcorpus (8.72 billion words) of the General Internet Corpus of Russian (GICR, http://www.webcorpora.ru/); it contains blogs that did not undergo editing or proofreading. Our dataset is based on blog posts published between 2010 and 2013. We took the list of inanimate indeclinable nouns from the *Grammatical Dictionary of Russian* (Zaliznjak 1987). We added a few other nouns that we noticed to be in use (both based on our native speaker intuition and on the dictionary study by Murphy 2000) and excluded several very infrequent nouns that would not be familiar to many Russian speakers (for instance, *kavallo*, a statue of a horseman in classical Italian art). We also did not include several highly frequent nouns: *kofe* 'coffee', *kafe* 'café', *metro* 'subway', *kino* 'movie (theatre)', *pal'to* 'coat', and *taksi* 'taxi'.[7]

We searched for combinations of an indeclinable noun and an attributive adjective preceding it as an agreement target. One instance of gender agreement with one noun produced by one speaker was taken as one observation (the corpus has information allowing us to identify different speakers). If one speaker produced numerous instances of the same agreement for the same noun, they were not included in the dataset. However, in 496 cases, one and the same noun was used in different genders by one and the same speaker (in different blog posts or even within one post). These cases were included as separate observations. They show that the gender of indeclinables can vary even in the grammar of a single speaker.

The resulting dataset contained 66,939 datapoints. After manually cleaning the data, we excluded all nouns with less than two examples of gender agreement and annotated the dataset for gender, number, and case. Many instances were annotated automatically based on the inflections of adjectives; other instances were manually disambiguated. Plural forms were not annotated for gender (there is no gender agreement in plural). In the singular, masculine and neuter adjective forms coincide in all cases except for nominative

[7] All these nouns are much more frequent than the majority of other nouns in our list, so we had a feeling that they could skew the results. They show almost no gender variation in contemporary Russian, being neuter (although see Comrie, Stone, and Polinsky 1996 for some interesting diachronic data), except for the word *kofe* 'coffee'. *Kofe* is masculine according to the conservative literary norm, but many speakers use it with neuter agreement, which provoked one of the biggest orthoepic debates in modern Russian. Gender variation in other indeclinable nouns is not widely discussed.

and accusative, so these instances were labeled as NM. For the analysis in this study, we selected only examples in nominative and accusative singular where gender can be determined unambiguously.[8] We also did not include monosyllabic nouns.[9] As a result, we had 131 nouns (types) and 32,792 observations (tokens) in the final dataset.

We annotated all nouns in this dataset for the root-final vowel, lexical stress (on the final, penultimate, or other syllables), and semantic analogy. The last task was less straightforward. Studies discussing semantic analogy in gender assignment do not report any problems with the procedure (Poplack, Pousada, and Sankoff 1982; Fuller and Lehnert 2000; and Violin-Wigent 2006). But, according to our intuitions, for some Russian indeclinables, one semantic analogy is found easily, while for the others, several variants come to mind, or there is no obvious variant at all. Therefore, we conducted a survey with 25 native speakers of Russian using the IbexFarm platform (https://korpling.german.hu-berlin.de/ibex/). The participants were asked which word they would use to briefly describe an indeclinable noun presented on the screen. Participants could also answer that they did not know the noun or did not know how to explain its meaning briefly. Several indeclinables were frequently described with nouns of different genders, e.g., *bungalo* 'bungalow' as *dom* 'house.M' and *žilišče* 'housing.N'. However, only the nouns in the M/F group were numerous.[10] All other mixed groups contained fewer than five nouns. Therefore, the following labels for semantic analogy were adopted: *F, M, N, M/F,* and *?*. The last label was used when no single noun was present in more than 75% of answers (because our participants struggled to give a short definition or because the definitions they gave were not uniform).

3.2. Statistical Analysis, Results, and Discussion

In this section, we discuss the main results of data analysis, while the following sections focus on fitting them into different morphological frameworks and modeling them. These results are also interesting for the discussion of the gender markedness problem; there is a debate as to whether the masculine or

[8] It would be interesting to find out how often different indeclinables are used as pluralia tantum nouns, but this is difficult to do based on naturally occurring examples (most indeclinables can be used in plural, so we would have to show that some nouns are used in plural more often than expected).

[9] Monosyllabic words with a final vowel are very unusual for the Russian nominal system; almost all of them are names of Latin and Cyrillic letters and musical notes.

[10] The M/F group primarily includes car brands. The feminine semantic analogy (*mašina* 'car.F') is very strong for them; the other analogy (*avtomobil'* 'automobile.M, car.M') is less prominent but nevertheless appears regularly and can be shown to affect the behavior of these nouns.

the neuter is the unmarked gender in Russian. We address this question in a separate paper (Chuprinko et al. 2023), which also presents more details on gender variation in indeclinable nouns, providing numerous examples and comparisons.

The first general observation is that gender variation in indeclinable nouns is enormous. Almost every noun in the sample varies in gender to some extent. Individual speakers often select different genders for one and the same noun, even within one text. Secondly, neuter is the most frequent gender in indeclinables (43% of the instances in our dataset), but masculine follows it closely (37% of instances). Feminine examples are clearly a minority (20% of instances). Thus, the picture we get from the corpus differs from the one we get from the dictionaries, although the hierarchy of genders is the same. Notably, many instances of masculine gender assignment are due to semantic analogy. Out of the words that have one salient semantic analogy in our dataset, 52 can be associated with masculine, 21 with feminine, and 15 with neuter (presumably because masculine is the most frequent and neuter the least frequent among declinable nouns). However, the neuter is often assigned in the absence of any clear cues. This pattern is discussed in more detail in Chuprinko et al. (2023).

Before turning to the statistical analysis, a caveat should be mentioned. Semantic and morphophonological features are distributed non-uniformly in our dataset—for example, there are only 12 indeclinable nouns ending in *-(j)a*, and each of them is relatively infrequent. Still, the dataset was large enough to analyze the factors of interest.

The statistical analysis was done in the R programming environment (R Core Team 2021). We tried two approaches to data modeling: mixed-effects logistic regressions with a random intercept by item (*lme4* package; Bates et al. 2015) and logistic regressions with fixed effects only using the built-in *glm* function. We explain the theoretical implications of both approaches and compare the models in §5. The mixed-effects approach turned out to be superior, so below we report the results of mixed-effects models, focusing on the main significant findings (the full description and outputs of all models can be found in the Appendix, available at https://osf.io/wsuaz/). For post hoc analyses, we ran Tukey's tests with the Holm-Bonferroni correction, using the *glht* function from the *multcomp* package (Bretz et al. 2010, and Hothorn et al. 2008). So, all *p*-values reported in tables 2 and 3 are corrected.

We made three mixed-effects models testing the probability of a word to be masculine, feminine, and neuter.[11] We used the same independent vari-

[11] Thus, the dependent variable in our three models is binary (e.g., the word is masculine vs. not masculine depending on the factors included in the model). It is also possible to make a multinominal model with a categorically distributed dependent variable (masculine vs. feminine vs. neuter) using the *mclogit* package (Elff 2022).

ables for each model: the root-final vowel, the gender of the semantic analogy noun, lexical stress (all non-final stresses were grouped), and case (nominative or accusative). We had no specific predictions for the case factor and only wanted to check that it does not influence our results in any unexpected way. We chose neuter as a reference level for the semantic analogy factor and *-i* as a reference for the final vowel factor (the case and stress factors were binary, so no reference level was needed).

The root-final vowel and semantic analogy factors were significant for all genders. Since these factors are not binary, we discuss pairwise comparisons between different values below. Lexical stress was significant only for feminine and neuter models. Nouns with the final stress are significantly less likely to be feminine ($\beta = -1.75$, $SE = 0.59$, $p = 0.003$) than masculine or neuter and significantly more likely to be neuter ($\beta = 1.41$, $SE = 0.43$, $p = 0.001$) than masculine or feminine. This may be due to the fact that the stress makes root-final vowels more salient because Russian has unstressed vowel reduction. The case factor reached significance in all three models, and we specifically checked that it improves their performance using the methods described in §5, but these results are difficult to explain. For instance, it may be that some nouns that tend to be assigned feminine have a relatively higher incidence of accusative forms than those that tend to be assigned neuter, etc. We tried including interactions between different factors in the models, but the resulting models were inferior.

Now let us discuss two main factors in more detail. Figures 1 and 2 on the following page show the role of the semantic analogy. Figure 1 groups the data by the gender of the semantic analogy, while Figure 2 groups them by the assigned gender, which corresponds to our statistical models. This gives us different perspectives. In Figure 1, we can see the general preference for the neuter gender, but masculine and M/F semantic analogies override it, while feminine semantic analogies visibly reduce the share of neuter.

In Figure 2, we can see that the masculine gender group includes mostly nouns with masculine and M/F semantic analogies. The feminine group almost exclusively consists of nouns with feminine and M/F semantic analogies. The neuter group, in contrast, includes a variety of nouns.

However, one value of the dependent variable must be taken as a reference level in such models; for example, if we take the neuter, the model will give us two sets of pairwise comparisons: N vs. M and N vs. F. For the third comparison, another model with a different reference level is needed. We believe that this is not optimal for interpreting the results and for comparing different models, which is one of the goals of this paper. Moreover, as we will show below, it is clear from the data that the factors of interest affect the choice of different genders in different ways. Therefore, we wanted to analyze each gender separately, not comparing it to one of the others but to both of them at once.

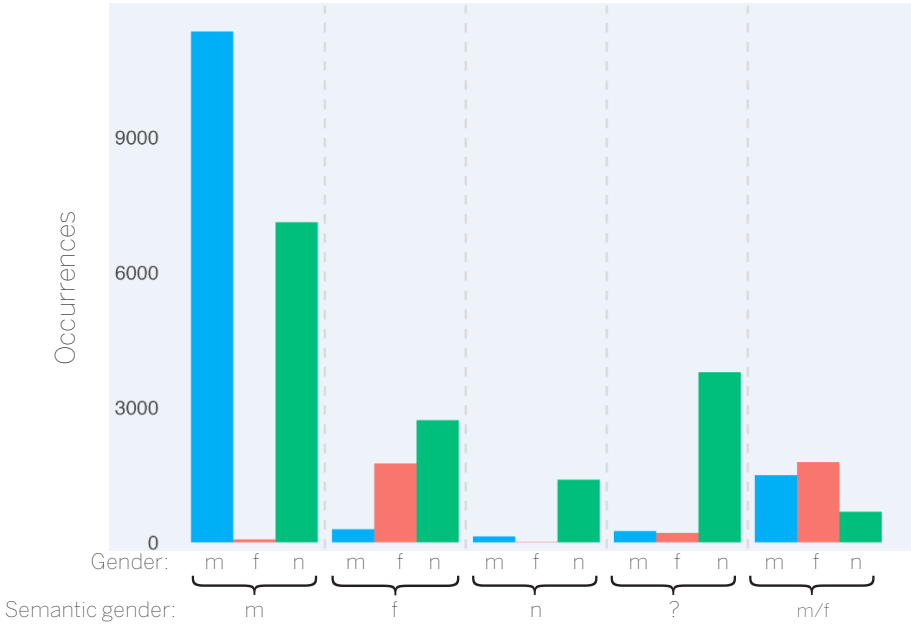


Figure 1. The distribution of genders depending on the semantic analogy[12]

The results of pairwise comparisons can be found in Table 2 (only significant results are reported). Table 2 shows the estimated effect (beta) in log odds, the standard deviation, and the significance for each pair of semantic analogy values for each gender. For example, *gender = masculine, m - n == 0, β = 3.42, SE = 0.53, p < 0.001**** means that when compared to nouns with masculine semantic analogy, nouns with neuter semantic analogy are less likely to end up masculine by approximately 3.42 in log odds with a standard deviation of 0.53, and this difference is significant at the level of $p < 0.001$.

[12] "?" indicates that there is no obvious semantic analogy noun, so its gender cannot be determined. "M/F" indicates that there are two salient semantic analogy nouns of masculine and feminine gender.

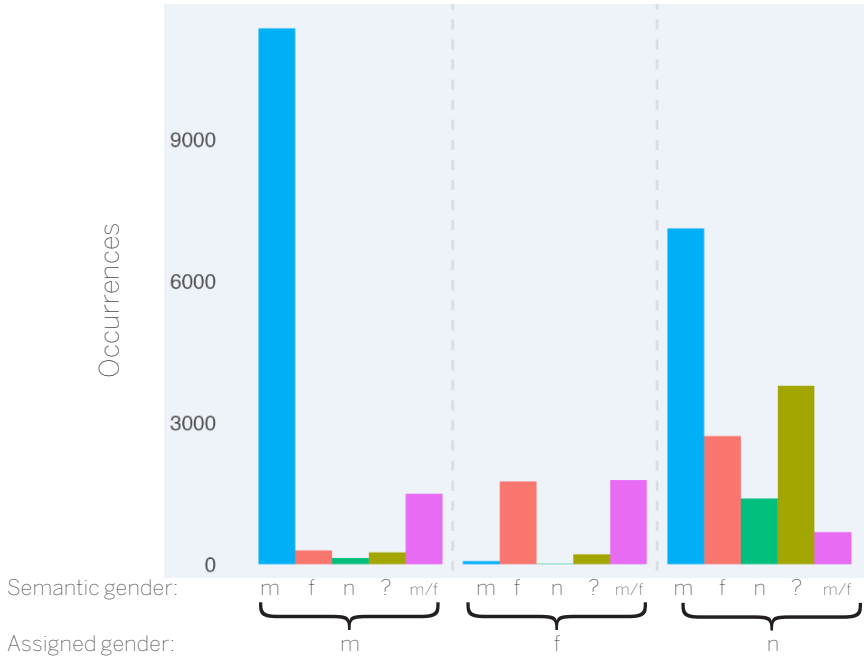


Figure 2. Semantic analogy in each gender group

Table 2. Pairwise comparisons for the semantic analogy factor

Assigned gender	Pair	Beta	SE	*p*	Significance code
masculine	m - n == 0	3.420	0.526	< 0.001	***
masculine	m/f - n == 0	2.107	0.636	0.005	**
masculine	m - ? == 0	3.409	0.450	< 0.001	***
masculine	m/f - ? == 0	2.100	0.579	0.002	**
masculine	m - f == 0	4.140	0.516	< 0.001	***
masculine	m/f - f == 0	2.829	0.622	< 0.001	***
masculine	m/f - m == 0	−1.311	0.508	0.040	*
feminine	f - n == 0	6.444	1.038	< 0.001	***
feminine	m/f - n == 0	6.693	1.046	< 0.001	***
feminine	f - ? == 0	5.016	0.726	< 0.001	***
feminine	m/f - ? == 0	5.265	0.737	< 0.001	***
feminine	m - f == 0	−5.783	0.630	< 0.001	***
feminine	m/f - m == 0	6.032	0.640	< 0.001	***

–continued on next page–

—continued—

Assigned gender	Pair	Beta	SE	*p*	Significance code
neuter	f - n == 0	−2.250	0.643	< 0.001	***
neuter	m - n == 0	−3.230	0.553	< 0.001	***
neuter	m/f - n == 0	−4.431	0.676	< 0.001	***
neuter	f - ? == 0	−2.161	0.571	< 0.001	***
neuter	m - ? == 0	−3.141	0.458	< 0.001	***
neuter	m/f - ? == 0	−4.343	0.598	< 0.001	***
neuter	m - f == 0	−0.980	0.498	0.098	.
neuter	m/f - f == 0	−2.181	0.622	0.002	**
neuter	m/f - m == 0	−1.201	0.530	0.070	.

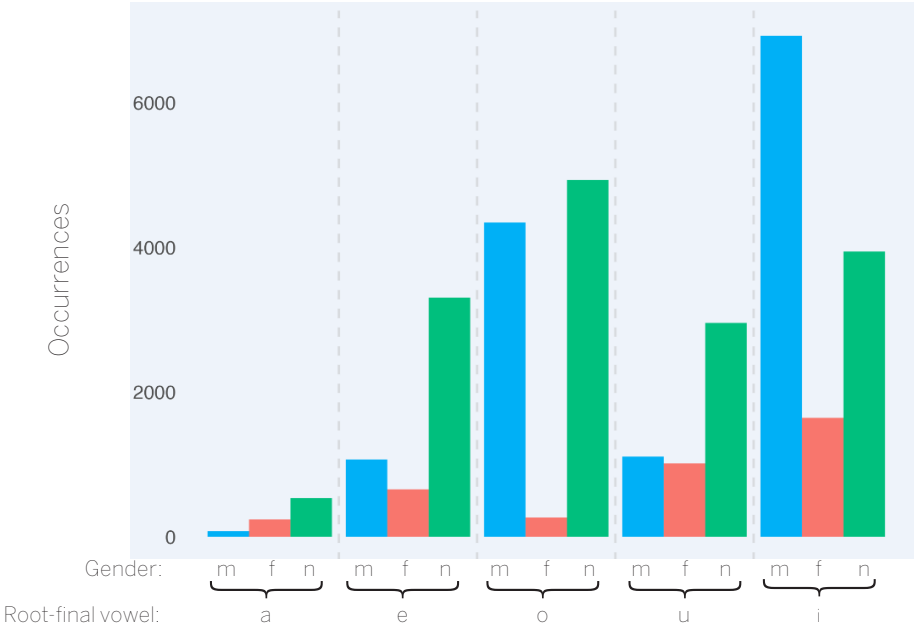


Figure 3. The distribution of genders depending on the root-final vowel

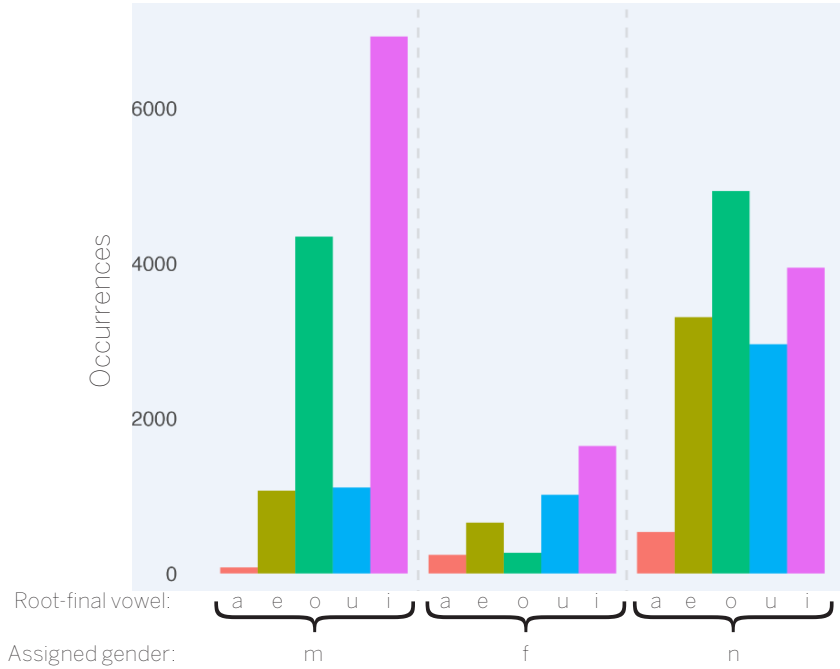


Figure 4. Root-final vowels in each gender group

Figures 3 and 4 show how the morphophonological analogy affects gender distribution. In Figure 3, we can see that the share of feminine agreement is larger for the nouns ending in *-(j)a*, while the share of neuter agreement is larger for the nouns ending in *-e*. For the nouns ending in *-i*, which do not have any morphophonological analogy in the singular, the share of masculine forms is the largest, due to the role of the semantic analogy factor. The semantic factor also explains a large share of masculine forms in the *o*-final group. Still, neuter is the most frequent gender in it, demonstrating the interplay of the two factors.

Figure 4 gives a different perspective on the same data. As we could see above, a word with a masculine semantic analogy has high chances to be masculine, while for a word with a feminine semantic analogy, this is less certain, and the morphophonological factor may affect the distribution. The neuter group is again the most diverse. The results of pairwise comparisons are reported in Table 3. They are significant only in the feminine and neuter models—presumably because the final vowel factor may change the share of these two genders—but not in the masculine, which is selected predominantly based on the semantic factor. In neuter models, the comparisons with *-u* and especially

-e are not significant because these groups are more diverse internally than *-o* (semantic analogy happens to play a stronger role for the nouns in these groups). In the next section, we will discuss these results in light of different theoretical frameworks.

Table 3. Pairwise comparisons for the root-final vowel factor

Assigned gender	Pair	Beta	SE	*p*	Significance code
feminine	a - i == 0	4.287	0.886	< 0.001	***
feminine	o - i == 0	−1.653	0.651	0.044	*
feminine	u - i == 0	1.903	0.750	0.044	*
feminine	e - a == 0	−4.885	0.890	< 0.001	***
feminine	o - a == 0	−9.940	0.924	< 0.001	***
feminine	u - a == 0	−2.384	0.807	0.016	*
feminine	u - e == 0	2.501	0.747	0.005	**
feminine	u - o == 0	3.556	0.802	< 0.001	***
neuter	o - i == 0	1.689	0.470	0.003	**
neuter	o - a == 0	2.300	0.703	0.010	**
neuter	u - o == 0	−1.745	0.566	0.017	*

4. Comparing Different Frameworks

In this section, we show why Russian indeclinable nouns are an interesting dataset to test different theoretical approaches to morphological gender. We compare three approaches representing opposing views on the grammatical architecture: Distributed Morphology (DM), Relational Morphology (RM), and Optimality Theory (OT). We compare both classic and probabilistic versions of these approaches. The crucial properties of the theories relevant for our data are (i) what kind of information is available when gender assignment takes place and (ii) whether the model predicts one outcome or several possible outcomes of different probability.

First, let us summarize some interesting properties of Russian indeclinable nouns. Crosslinguistically, relying on semantic analogy and morphophonology in assigning gender to loanwords is not uncommon (e.g., Corbett 1991: 75–82). The case of Russian indeclinables is interesting because most other examples discussed in the literature are either properly incorporated into the native inflectional system or come from languages with poor inflec-

tional morphology in which this problem does not arise. Russian also has numerous loanwords of the first type. For example, the English word *bug* (in the meaning 'a software error') was adopted into Russian in two versions: *bag* (class IIa masculine) and *baga* (class I feminine). The first version, which is more frequent, is closer to the original in the nominative singular, while the second version probably appeared due to semantic analogy with the word *ošibka* 'error.f'. The choice of the feminine gender predetermined the choice of the inflectional class. Examples like *baga* with an added affix are rare in Russian, while examples like *vendetta* 'vendetta.f' borrowed from Italian are frequent; in such loanwords, *-a* is reanalyzed as a class I inflectional affix, so they become feminine.

Indeclinables are different in three respects. Firstly, most lexemes demonstrate gender variation, which may be extensive: some nouns, like *xudi* 'hoodie', have comparable frequencies of all three genders in our dataset. Secondly, once an inanimate word is borrowed and becomes declinable, its semantics does not influence its gender, while in indeclinables, speakers continue to rely on this information. Some inanimate declinable nouns show gender variation, but it is never due to semantics—rather, class III loanwords like *šampun'* 'shampoo' migrate to the IIa class, and a gender change accompanies the change of declension.[13] Thirdly, as we noted above, loanwords like *vendetta* become feminine because *-a* is reanalyzed as a class I inflectional affix, and all inanimate nouns in this class are feminine. This pattern can be accounted for in most major morphological theories. In indeclinables, no reanalysis takes place, the last vowel belongs to the root, but the morphophonological analogy affects gender assignment nevertheless. In this section, we argue that its role is difficult to explain in several major frameworks.

4.1. Distributed Morphology

Structural theories mostly assume modularity and sequential processing, which restricts the availability of certain information at different derivation stages. Being a syntactic feature, gender must receive its value at an early derivation stage—when the syntactic tree is constructed. But phonological information is not available at this stage. This raises a problem for our results as we have shown that both phonological and semantic factors influence gender assignment and give rise to variation, which is especially difficult to account for. This problem is relevant for all sequential approaches; let us consider the most influential one, Distributed Morphology, as an example.

[13] Another small group of nouns that show gender variation are nouns with diminutive and augmentative affixes: their gender depends on the gender of the base noun as well as on their inflectional class (e.g., Steriopolo et al. 2021 and Magomedova and Slioussar 2023). Thus, this variation is also not due to semantics.

Kramer (2020) claims that gender is assigned by the Merge operation when a bare root or a stem is merged with a nominal head already containing a gender feature. Declension is also a syntactic feature situated lower in the structure. Therefore, it is easy to explain how declension can influence the choice of gender—like in Russian declinable loanwords and many other cases—but very difficult to explain the behavior of Russian indeclinables.

Several other authors working in the Distributed Morphology framework aim to "decompose" Russian declensions, reducing them to combinations of several features. For example, Privizentseva (2023) suggests the following system: class I is [+fem][−α] (the small group of animate masculine nouns in this class receives a separate treatment), class IIa is [−fem][+α], class IIb is [−fem] [−α], and class III is [+fem][+α], where *α* is a declension feature and *fem* a gender feature. Importantly, in all approaches, formal or functional, including Kramer (2020) and Privizentseva (2023), gender features are realized by inflections and not by roots.[14] Therefore, there is no way to explain why an indeclinable noun with a root-final *-a* is more likely to be feminine—unless the process can refer to the surface form.

4.2. Relational Morphology

Relational Morphology (e.g., Jackendoff and Audring 2020) postulates a continuum between the lexicon and the grammar: a unified space of schemas, where a schema is a multilayer representation of a real or possible word. Different layers contain semantic, syntactic, and phonological information. Different parts of this information are connected by indices. An example of a schema for the noun *reader* is given in (1).

(1) Semantics: $[\text{PERSON}_a; [\text{READ (a, Y)}]_1]_2$
Morphosyntax: $[\text{N } [\text{V}]_1 \text{ aff}_3]_2$
Phonology: $/\text{ri:d}_1 \text{ ər}_3/_2$

Schemas are connected to each other by the "same-same" relations: relations point to a part of a given schema that can also be found in other schemas. One of the basic principles of Relational Morphology is the availability of all information in a schema and its relations at any given moment. This allows us to account for the influence of semantic and phonological factors at once.

However, to be able to predict gender distribution, it is still necessary to look inside the root and to connect the root-final segment in indeclinables to

[14] In Caha's (2021) approach, different roots spell out different sets of features—he uses this to explain why they select different inflections (in order to get rid of arbitrary declension features). Unfortunately, this approach also does not predict that roots ending in particular vowels would be more likely to be assigned a particular gender.

inflectional affixes. We cannot do this (or properly formalize such operations) as each part of the schema must be substitutable with a variable—this guarantees the possibility to generate new words in this approach. It is possible to substitute a number of entities with a single variable, but it is not clear what would be the theoretical basis to divide a root into smaller structural pieces.

Notably, Doleshal (2000) developed a model of Russian gender that is similar in spirit to Relational Morphology. Her model even accounts for variation, although the distribution of variants cannot be predicted. However, Doleshal does not analyze indeclinables, so the problems outlined above are also relevant for her approach.

4.3. OT, Harmonic Grammars, and Maximum Entropy

In optimality-theoretic (OT) approaches, all kinds of information are available at the same time and thus can be used for gender assignment. It is also crucial for our data that these approaches do not require morphological reanalysis of a loanword to make morphophonological analogies possible. The derivation in this framework is a process of selecting the best candidate from a theoretically infinite set of possible realizations. All candidates are tested against a set of constraints. In the classic version of OT, these constraints are ranked, while in OT-based probabilistic theories, such as Harmonic Grammars and Maximum Entropy models, they are weighted. OT models of Russian gender not taking variation into account were developed by Rice (2005, 2006) and Galbreath (2010). In this study, we turn to probabilistic models because they allow for analysis of variation.

5. Modeling Our Data: Comparing Different OT Approaches

In the previous section, we demonstrated that our data can be analyzed only using OT-based approaches. To account for variation, we opted for probabilistic Maximum Entropy grammars over a strict OT grammar. Since it is possible to build statistical models based on these approaches, to choose the optimal one, we used them to model our data and determined which models have the largest predictive power.

Maximum Entropy grammars are essentially logistic regressions (Hayes and Wilson 2008), where betas (or estimates, see the comments for Table 2) are the weights of constraints and binary independent variables are the constraints themselves. These models were shown to capture many wider grammatical generalizations but were not ideal for predicting the idiosyncratic behavior of different lexemes: to do so, we would need to introduce different weights of the same constraint for specific words. Zymet (2019) used

mixed-effects regressions to capture this idiosyncratic behavior, which he introduces to the theory as Hierarchical MaxEnt.

The predictive power of the models is generally the reason why we trust statistics, but choosing between different statistical methods can also be seen as a part of the theoretical analysis in the case of probabilistic approaches. In this section, we compare the predictive power of mixed-effects and fixed-effects regressions and conclude that Zymet's Hierarchical MaxEnt approach accounts for our data more precisely. We also discuss the levels of the Hierarchical MaxEnt: what kind of behavior should be accounted for using constraints and what should be treated as idiosyncratic?

Our mixed-effects models treated each indeclinable noun as a random effect because each lexeme may have a different predisposition for gender assignment. For example, semantic analogy is more salient for some nouns than for others, etc. Speakers were not included as random effects, as we mostly had only one instance per lexeme from a speaker in our dataset. To separate language-wide factors (fixed effects) from idiosyncratic behavior (random effects) we excluded feature values that could be found only in a very few nouns (five or less). Namely, we did not introduce *m*/*n* and *f*/*n* values for the semantic analogy factor—they were included in the *?* group. Our motivation was simple: if we only have a couple of words with a certain value, it is impossible to tell whether they are different because of this feature value or because of some idiosyncrasy.[15]

We also tried including several minor factors as fixed effects. For example, we discussed in footnote 10 that nouns denoting car brands are more likely to be used with feminine agreement than any other group of nouns, so we included this as a separate semantic factor. Other minor factors we considered were word length and word-final hiatus. However, these factors did not improve the models, which indicates that such properties of smaller noun sets are better treated as idiosyncratic.[16]

Now let us focus on the comparison between two sets of models: three mixed-effects models and three models with fixed effects alone. In both sets, only four major factors were included as fixed effects: the semantic analogy gender, root-final vowel, lexical stress, and case. As we already mentioned in §3, we tried including interactions between different factors, but the resulting models were inferior according to the tests. All information on these six models, as well as pairwise comparisons for non-binary factors, can be found in

[15] To test this intuition, we built the models including all possible feature values, but they had smaller predictive power. We did not include them in the comparisons reported below for reasons of space.

[16] We tested this using the methods discussed in this section, but we cannot report all the results for reasons of space.

the Appendix (https://osf.io/wsuaz/), while below we discuss different estimates of their performance.

Indeclinable nouns are a limited set of words, so we tested the models on the same data that they were trained on. In particular, we could not afford to divide the dataset into test and training sets because mixed effects were calculated by word, and one and the same word could not appear in both sets. Unfortunately, this approach has its pitfalls: we cannot detect overfitting. As we show below, to control for this, we relied on the Akaike and Bayesian Information Criteria.

First of all, we used the pROC package (Robin et al. 2011) for ROC-AUC metrics. The ROC (Receiver Operating Characteristic) curve and the AUC (Area Under the Curve) metrics are widely used for evaluating the performance of binary classification models. The ROC curve is a graphical representation of the true positive rate (sensitivity) against the false positive rate (specificity) for a binary classifier system as its discrimination threshold is varied. The AUC refers to the area under the ROC curve. It provides a scalar value that characterizes the overall performance of the model across all thresholds. Figure 5 shows the ROC-AUC metrics for our models. Mixed-effects models are better than fixed-effects models for neuter (89% vs. 79%), slightly better for feminine (97% vs. 94%), and the same for masculine (82% vs. 82%).

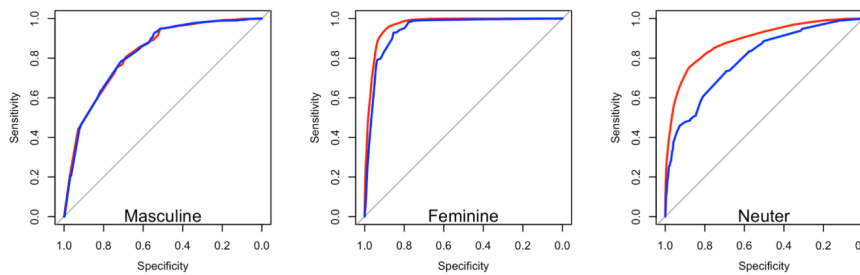


Figure 5. The performance of fixed- and mixed-effects models (shown in blue and in red, respectively)

To estimate the performance of our models in terms of the maximum likelihood, errors, and complexity, we used the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), the adjusted, marginal and conditional r-squared (R^2), and the root mean squared error (RMSE), calculated with the help of the performance package (Lüdecke et al. 2021). The information-theoretic indicators are based on maximal likelihood (ML) models because restricted maximal likelihood (REML) models are not available for binomial logistic regressions in the *lme4* package. As Table 4 shows, mixed-effects models are superior for all three genders according to all these criteria.

Table 4. Performance estimates for mixed- and fixed-effects models

Model	AIC	BIC	RMSE	marginal R^2	conditional R^2	adjusted R^2
Masculine: mixed effects	26202	26337	0.353	0.360	0.666	
Masculine: fixed effects	33159	33251	0.412			0.299
Feminine: mixed effects	9598	9732	0.212	0.571	0.796	
Feminine: fixed effects	12178	12271	0.243			0.414
Neuter: mixed effects	27969	28103	0.365	0.301	0.644	
Neuter: fixed effects	36019	36111	0.430			0.252

The RMSE criterion shows the square root from the squared sum of residuals divided by the number of observations, where a residual is a distance from a datapoint to the regression curve. The smaller this number is, the better the model. As we can see in Table 4, mixed-effects models are more precise by a large margin. R^2 is a more sophisticated criterion based on the same principle; it is calculated as the sum of squared differences between each actual datapoint and the model prediction divided by the sum of squared differences between each actual datapoint and the mean of all datapoints, all subtracted from one. Therefore, the bigger this number is, the better the model. Marginal R^2 shows how fixed effects alone account for variance in a mixed-effects model, while conditional R^2 shows an estimate for both fixed and mixed effects. As we can see in Table 4, conditional R^2 is larger than marginal R^2 for every model. In addition to that, adjusted R^2 calculated for fixed-effects models are always smaller than both marginal and conditional R^2 calculated for mixed-effects models.

However, both RMSE and R^2 criteria are known to prefer more complex models and do not penalize for overfitting. To take this problem into account, we used the Akaike and Bayesian Information Criteria. The AIC and BIC are based on the likelihood function and are almost the same, but BIC penalizes complex models more. The absolute value of both criteria is of no importance for the model selection; these numbers are used relatively—the smallest number among all the models trained on the same data indicates the best model. As we can see in Table 4, both criteria strongly prefer mixed-effects models

for all three genders. There may be some further complications in comparing mixed- and fixed-effects models (e.g., see McNeish and Kelley 2019), which makes us interpret the results in Table 4 with a certain caution. However, this does not apply to the ROC-AUC metrics in Figure 5, which show the superiority of mixed-effects models very clearly.

We can conclude that our results support the Hierarchical MaxEnt approach (Zymet 2019) over the ordinary MaxEnt one. The feminine gender is predicted more successfully than masculine and neuter, which is probably because it is assigned to indeclinables only in the presence of very strong cues (a stem-final *-(j)a* or an obvious semantic analogy word of feminine gender). This makes the feminine gender much less frequent in this group of nouns than masculine and neuter, and at the same time, easier to predict. Interestingly, if we look at pairwise comparisons for the semantic analogy and final vowel factors in the Appendix (https://osf.io/wsuaz/), we can see that fixed-effects models give a larger number of significant results than mixed-effects ones. Since the performance of fixed-effects models was inferior according to all used criteria, we can conclude that this significance is in fact misleading. This shows once again how important it is to compare different approaches to choose an optimal one.

6. Conclusions

We analyzed grammatical gender assignment in a set of 32,792 instances of 131 indeclinable inanimate nouns from the General Internet Corpus of the Russian Language. We found a substantial variation in gender for the majority of nouns and identified several contributing factors: the gender of the semantic analogy noun and the root-final vowel, as well as the lexical stress and case. The first two factors were noted in earlier studies, but their role has never been estimated on corpus data or discussed in the context of different morphological frameworks.

We believe that Russian indeclinables are noteworthy because in declinable inanimate Russian nouns, the gender does not depend on semantics or on the properties of the stem—rather, it strongly correlates with the inflectional class. When a new loanword becomes declinable, its semantics can sometimes affect the choice of declension and gender, and its phonology very often does (for example, the final *-a* is usually reanalyzed as a class I nominative singular affix). But once the loanword has been incorporated in the Russian nominal system, these factors stop playing a role, unlike in indeclinable nouns. It is especially remarkable that the final vowel in indeclinables influences gender assignment although it does not get reanalyzed as an affix and remains part of the root.

We argued that indeclinable nouns are an interesting dataset for comparing different morphological frameworks. We selected three approaches

representing opposing views on the grammatical architecture: Distributed Morphology, Relational Morphology, and Optimality Theory, for which classic and probabilistic versions were compared. We identified two properties of these approaches as crucial for our dataset: (i) what kind of information is available when gender assignment takes place and (ii) whether the model predicts one outcome or several possible outcomes of different probability. As for (i), in DM, no phonological information is available when gender assignment takes place. In RM, phonological information may be accessed, but not inside the root. In OT, constraints can take any information into account. As for (ii), only probabilistic OT approaches can deal with multiple outcomes of different probability.

Finally, within the family of probabilistic OT approaches, we aimed to select the one that would be optimal for our data, accounting both for some general tendencies and for some idiosyncratic properties of small noun groups or individual words. Since it is possible to build statistical models based on these approaches, we used them to model our data and determined which models have the largest predictive power. As a result, we demonstrated that the hierarchical Maximum Entropy (MaxEnt) models (Zymet 2019) were superior to the classical MaxEnt models.

References

Aronoff, Mark. (1994) *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT press.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. (2015) "Fitting linear mixed-effects models using lme4". *Journal of statistical software* 67(1): 1–48.

Bretz, Frank, Torsten Hothorn, and Peter Westfall. (2010) *Multiple comparisons using R*. Boca Raton, FL: CRC Press.

Caha, Pavel. (2021) "Modeling declensions without declension features: The case of Russian". *Acta linguistica academica* 68(4): 385–425.

Chuprinko, Kirill, Varvara Magomedova, and Natalia Slioussar. (2023) "Gender variation in indeclinable inanimate nouns and gender markedness in modern Russian". *Acta linguistica academica* 70(3): 317–38.

Comrie, Bernard, Gerald Stone, and Maria Polinsky. (1996) *The Russian language in the twentieth century*. Oxford, UK: Clarendon Press.

Corbett, Greville. (1991) *Gender*. Cambridge, UK: Cambridge University Press.

————. (2023) "The dog didn't bark, the noun didn't inflect: A typology of significant absences". Paper presented at DGfS (45th annual conference of the German Linguistic Society), Cologne, Germany, March 2023.

Corbett, Greville, and Norman M. Fraser. (2000) "Gender assignment: A typology and a model". Gunter Senft, ed. *Systems of nominal classification*. Cambridge, UK: Cambridge University Press, 293–325.

Doleschal, Ursula. (2000) "Gender assignment revisited". Barbara Unterbeck, Matti Rissanen, Terttu Nevalainen, and Mirja Saari, eds. *Gender in grammar and cognition. Part I: Approaches to gender*. Berlin: Mouton de Gruyter, 117–66.

Elff, Martin. (2022) "mclogit: Multinomial logit models, with or without random effects or overdispersion". R package version 0.9.6. Available at: https://CRAN.R-project.org/package=mclogit. Last accessed 10 December 2022.

Fuller, Janet M., and Heike Lehnert. (2000) "Noun phrase structure in German-English codeswitching: Variation in gender assignment and article use". *International journal of bilingualism* 4(3): 399–420.

Galbreath, Blake Lee Everett. (2010) *Gender assignment in contemporary standard Russian: A comprehensive analysis in optimality theory*. PhD dissertation, University of Virginia, Charlottesville, Virginia.

Halle, Morris. (1994) "The Russian declension: An illustration of the theory of Distributed Morphology". Jennifer Cole and Charles Kisseberth, eds. *Perspectives in phonology*. Stanford, CA: CSLI Publications, 29–60.

Hayes, Bruce, and Colin Wilson. (2008) "A maximum entropy model of phonotactics and phonotactic learning". *Linguistic inquiry* 39(3): 379–440.

Henry, Louise. (2020) "Morfologičeskij status slova *pal'to* v russkom jazyke: Korpusnoe issledovanie". *Movna osobistist': Lingvistika i lingvodidaktika* 5: 350–56.

Hothorn, Torsten, Frank Bretz, and Peter Westfall. (2008) "Simultaneous inference in general parametric models". *Biometrical journal* 50(3): 346–63.

Jackendoff, Ray, and Jenny Audring. (2020) *The texture of the lexicon: Relational Morphology and the Parallel Architecture*. Oxford: Oxford University Press.

Kramer, Ruth. (2015) *The morphosyntax of gender*. Oxford: Oxford University Press.

————. (2020) "Grammatical gender: A close look at gender assignment across languages". *Annual review of linguistics* 6: 45–66.

Lüdecke, Daniel, Mattan S. Ben-Shachar, Indrajeet Patil, Philip Waggoner, and Dominique Makowski. (2021) "performance: An R package for assessment, comparison and testing of statistical models". *Journal of open source software* 6(60): 3139. DOI 10.21105/joss.03139.

Magomedova, Varvara, and Natalia Slioussar. (2023) "Gender variation and gender markedness in Russian nouns". *Voprosy jazykoznanija* 2: 7–28.

McNeish, Daniel, and Ken Kelley. (2019) "Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations". *Psychological methods* 24(1): 20–35.

Mjakilja, Kari. (2000) "K probleme roda nesklonjaemyx zaimstvovannyx imen naricatel'nyx v sovremennom russkom jazyke". *Scando-Slavica* 46: 93–103.

Mučnik, Iosif. (1971) *Grammatičeskie kategorii glagola i imeni v sovremennom russkom literaturnom jazyke*. Moscow: Nauka.

Murphy, Dianna L. (2000) *The gender of inanimate indeclinable common nouns in modern Russian*. PhD dissertation, Ohio University, Athens, Ohio.

Poplack, Shana, Alicia Pousada, and David Sankoff. (1982) "Competing influences on gender assignment: Variable process, stable outcome". *Lingua* 57(1): 1–28.

Privizentseva, Mariia. (2024) "Semantic agreement in Russian: Gender, declension, and morphological ineffability." *Natural language and linguistic theory* 42: 767–814.

R Core Team. (2021) "R: A language and environment for statistical computing". Vienna, Austria: R Foundation for Statistical Computing. Available at https://www.R-project.org/.

Rice, Curt. (2005) "Optimizing Russian gender: A preliminary analysis". Steven Franks, Frank Gladney, and Mila Tasseva-Kurktchieva, eds. *Formal Approaches to Slavic Linguistics #13: The South Carolina Meeting.* Ann Arbor, MI: Michigan Slavic Publications, 265–75.

————. (2006) "Optimizing gender". *Lingua* 116(9): 1394–417.

Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, and Markus Müller. (2011) "pROC: An open-source package for R and S+ to analyze and compare ROC curves". *BMC bioinformatics* 12, article 77. DOI 10.1186/1471-2105-12-77

Rozental', Ditmar, Evgenija Džandžakova, and Natal'ja Kabanova. (1998) *Spravočnik po pravopisaniju, proiznošeniju, literaturnomu redaktirovaniju*. Moscow: ČeRo.

Savchuk, Svetlana. (2011) "Korpusnoe issledovanie variantov rodovoj prinadležnosti imen suščestvitel'nyx v russkom jazyke". *Computer linguistics and intellectual technologies* 10: 562–79.

Shigemori Bučar, Chikako. (2011) "Creative competence in borrowings: Words of Japanese origin in Slovene". *Linguistica* 51(1): 245–62.

Shvedova, Natalia, ed. (1980) *Russkaja grammatika.* Moscow: Nauka.

Slioussar, Natalia, and Maria Samoilova. (2015) "Častotnosti različnyx grammatičeskix xarakteristik i okončanij u suščestvitel'nyx russkogo jazyka". In *Proceedings of the conference "Dialogue"*. Available at: www.dialog-21.ru/digests/dialog2015/materials/pdf/SlioussarNASamoilovaMV.pdf.

Steriopolo, Olga, Giorgos Markopoulos, and Vassilis Spyropoulos. (2021) "A morphosyntactic analysis of nominal expressive suffixes in Russian and Greek". *The linguistic review* 38(4): 645–86.

Swan, Oscar E. (2002) *A grammar of contemporary Polish*. Bloomington, IN: Slavica Publishers.

Thomas, George. (1983) "A comparison of the morphological adaptation of loanwords ending in a vowel in contemporary Czech, Russian, and Serbo-Croatian". *Canadian Slavonic papers* 25(1): 180–205.

Violin-Wigent, Anne. (2006) "Gender assignment to nouns codeswitched into French: Observations and explanations". *International journal of bilingualism* 10(3): 253–76.

Wang, Qiang. (2014) "Gender assignment of Russian indeclinable nouns". Master's thesis, University of Oregon, Eugene, OR.

Zaliznjak, Andrej A. (1987) *Grammatičeskij slovar' russkogo jazyka: Slovoizmenenie*. 2nd ed. Moscow: Russkij Jazyk.

Zymet, Jesse. (2019) "Learning a frequency-matching grammar together with lexical idiosyncrasy: MaxEnt versus Hierarchical Regression". Katherine Hout, Anna Mai, Adam McCollum, Sharon Rose, and Matthew Zaslansky, eds. *Proceedings of the 2018 Annual Meeting on Phonology*. Available at: https://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/4495/4198.

Varvara Magomedova
Independent researcher
Warsaw, Poland
varya.magomedova@gmail.com

Kirill Chuprinko
Center for Cognitive Science of Language
University of Nova Gorica
Nova Gorica, Slovenia
chuprinko_kirill@mail.ru

Natalia Slioussar
School of Linguistics
HSE University
Moscow, Russia
slioussar@gmail.com

OV/VO Word Order in Heritage Russian: Is Transfer at Play?

Maria Martynova, Yulia Zuban, Luka Szucsich, and Natalia Gagarina

Abstract: The present study investigates the choice of OV/VO word order in heritage and monolingual Russian. In monolingual Russian, OV/VO order is claimed to be sensitive to the object realization (noun vs. pronoun) and clause type (main vs. embedded). In heritage Russian, OV/VO order is claimed to be prone to changes under language contact. Analyzing spoken and written narratives produced by heritage speakers (HSs) of Russian residing in the US and Germany, we scrutinize HSs' choice of OV/VO orders in comparison to the monolingual speakers from Russia. According to the results of the binomial generalized linear mixed-effects model, the OV/VO choice in heritage Russian was best predicted by the clause type and object realization. Specifically, the likelihood of producing the OV order was lower in the embedded clauses than in the main clauses among all speaker groups. Furthermore, all three speaker groups preferred the OV order with the pronominal object, while the preference shifted towards the VO order when the object was realized by a noun. Finally, both HS groups behaved similarly to the monolingual speakers in their choice of OV/VO orders. The results of the study do not provide any clear evidence for cross-linguistic influence from the majority languages and suggest that the word order choice of heritage and monolingual speakers depends on multiple factors, such as clause type and object realization.

1. Introduction*

The number of migrants worldwide increases continuously, establishing robust language contact situations in receiving countries. Some countries, such as the US and Germany, have a substantial proportion of immigrants, around 15.5% and 16% of the population, respectively (United Nations 2019).

* This research was funded by the German Research Foundation (DFG) as part of the research unit "Emerging grammars in language contact situations: A comparative approach" (FOR 2537), project no. 313607803, GZ: ZE 940/4-1, SZ 263/6-1, GA 1424/10-1.

In the US, approximately 880,000 speakers report that they speak Russian at home, making it the ninth most spoken family language (US Census Bureau 2015). Russian-speaking communities in the US are typically not located in particular neighborhoods but rather are scattered across the country (Dubinina and Polinsky 2013; Laleko 2013). Nevertheless, there are some places with a higher density of Russian speakers. The largest Russian-speaking communities are situated in New York (226,290 speakers), California (151,685), and Washington (56,537) (US Census Bureau 2015).

In Germany, approximately 3.487 million people originate from the countries of the former Soviet Union (Statistisches Bundesamt 2019a, 2019b). However, the real number of Russian speakers in Germany may reach up to 6 million people (cf. Anstatt 2011; Aref'ev 2012; Brehmer 2007; Brehmer and Mehlhorn 2015; Gagarina 2014, 2017; Gagarina and Klassert 2018; Hamann et al. 2019; Topaj 2018).

The present study focuses on heritage Russian that is spoken in the US and Germany. We adopt the following definition of HSs and heritage languages (HLs): "Heritage languages are languages spoken by the children of immigrants or by those who immigrated to a country when young" (Cho et al. 2004: 23). HSs are therefore bi- or multilingual speakers, using their HL(s) alongside the majority language(s) (ML) of the surrounding community (cf. among many others, Benmamoun et al. 2013; Guijarro-Fuentes and Schmitz 2015; Montrul 2015; Polinsky 2015; Rothman 2009; Valdés 2005).

The language performance of HSs of Russian was often reported to differ from the monolingual speakers of Russian (Laleko 2019; Romanova 2008). Compared to the monolingual speakers, heritage Russian was found to show salient changes in lexicon (Isurin 2011; Polinsky 2006), nominal morphosyntax such as case and gender agreement (Laleko 2018; Polinsky 2006, 2008a), verbal morphology (Romanova 2008), verbal aspect (Gagarina et al. 2020; Laleko 2010, 2011, 2015; Polinsky 2006, 2008b), pro-drop (Dubinina and Polinsky 2013; Isurin 2011; Laleko and Polinsky 2017), and word order (Brehmer and Usanova 2015; Dubinina and Polinsky 2013; Isurin and Ivanova-Sullivan 2008; Kisselev 2019; Laleko and Dubinina 2018; Polinsky 2006, 2011; Zuban et al. 2021).

Although word order in heritage Russian has been investigated in a number of studies, none of them, to our knowledge, explicitly focused on the OV/VO orders in the narrow sense, i.e., the ordering of objects/internal arguments relative to the main verb. The present study aims to fill in this gap. Grouping the data into OV/VO orders allows us to include the dataset by setting minimal exclusion criteria, as specified in §4.3. Limiting ourselves to the OV/VO distinction is also grounded in its importance for the theoretical discussion on head directionality in syntactic structures in general (cf. for a discussion, among many others, Baker 2001; Broekhuis 2006; Haider 2015; Haider and Szucsich 2022; Kayne 1994).

OV/VO orders represent a promising field for investigation in language contact for a number of reasons. The linearization of objects (internal arguments) relative to the main verb—especially in languages with so-called free word order—is determined not only by the directionality (licensing of verbal complements by initial or final heads), but also by information-structural properties of the utterance (see Bailyn 1995; Junghanns and Zybatow 1995; Kallestinova 2007; Slioussar 2011, among many others for Russian). Head-directionality is associated with core syntax (Aboh 2015; Uriagereka 2007)[1] and is considered to be less susceptible to contact-induced and diachronic change (Kroch and Taylor 2000; Pintzuk and Taylor 2006). Linearization of constituents due to information structure, on the other hand, concerns interfaces of core grammar with discourse. Furthermore, in monolingual Russian, OV/VO choice is highly influenced by object realization (noun vs. pronoun) (Kallestinova 2007).

In light of the Interface Hypothesis (see Sorace 2011 for a detailed discussion), one can come up with two competing predictions for languages with free word order: (i) If one assumes that head-directionality is the determining factor for the distribution of OV/VO orders, one would expect it to be less prone to contact-induced change resulting in a higher degree of optionality; and (ii) if one assumes that information structure is the decisive factor, one would expect the distribution of OV and VO in HSs to deviate from monolingual production by showing a higher degree of variability under constant conditions (object type: nominal vs. pronominal, and clause type: main vs. embedded).

Since the Interface Hypothesis is difficult to test, the aim of the current study is more modest, viz., to investigate the possible influence of the MLs, English and German, on the choice of OV/VO orders in heritage Russian by comparing two HS groups to monolingual speakers of Russian. The data of the participants are analyzed with the generalized mixed-effects modeling that accounts for the individual speaker variation and allows inclusion of other factors that influence the choice of OV/VO orders (they will be specified in detail later).

The paper is organized as follows. Section 2 provides theoretical background on verb placement in monolingual English, German, and Russian. Section 3 specifies the research questions alongside the hypotheses, derived from the previous studies on word order in heritage and monolingual Russian. Section 4 describes the experimental design, participants, and data analysis. Section 5 presents the results of the study, which are discussed in §6.

[1] Of course, this is true only if parametrized directionality is accepted within syntactic theory; for an alternative, see Kayne's (1994) proposal of the Linear Correspondence Axiom, which universally allows only for head-initial derivations.

2. Theoretical Background

2.1. Verb Placement in Majority Languages English and German

English is an SVO language with residual V2, e.g., quotative inversion, and very limited reordering of constituents, e.g., topicalization of objects (Birner and Ward 1998; Eppler 1999; Kempen and Harbusch 2019). This holds irrespective of what we call "clause type" in this article, i.e., it is valid for both main and embedded clauses, as illustrated in the example below:

(1) My friends know that I like this book.

In contrast to English, German is an SOV language with V2 in main clauses and reordering options for non-verbal constituents, i.e., German is well-known for exhibiting scrambling (Eppler 1999; Gärtner 2000; Haider 2012; Kempen and Harbusch 2019; Wegener 1993). In embedded clauses, introduced by complementizers and relativizers/wh-elements, verbal categories marked for finiteness—finite main verb or finite auxiliaries—canonically occupy the clause-final position (with the exception of particular, so-called non-integrated embedded clauses introduced by *weil* 'because', *obwohl* 'although', etc.; see Gärtner 2000 and Wegener 1993):

(2) Ich mag das Buch, das ich zum Geburtstag bekam.
I like the book which I for birthday got
'I like the book that I got for my birthday.'

Thus, German makes a clear distinction between main and embedded clauses when it comes to surface order, with robust OV in embedded and predominant VO order in main clauses with synthetic verb forms and OV with analytic verb forms. In the generative tradition, the V2 effect is derived by movement of the finite verbal element to the C-head, i.e., it is a grammaticalized process (Gärtner 2000 and Haider 2012, among many others).

2.2. Verb Placement in Monolingual and Heritage Russian

Monolingual Russian is considered to have an SVO order with full (non-pronominal) object NPs in neutral contexts, i.e., with sentential ("all new") and VP focus in both main and embedded clauses. At the same time, Russian is a language which allows massive re-orderings/permutations of constituents (Brehmer and Usanova 2015; Junghanns and Zybatow 1995; Kallestinova 2007; Slioussar 2007, 2011; Švedova 1980, 2005). In the generative literature on Russian, it is widely assumed that objects are "base-generated" to the

right of the verbal head (Bailyn 1995; Jasinskaja and Šimík, forthcoming; Junghanns and Zybatow 1995; Slioussar 2007, 2011).[2] One of the arguments for this assumption is that the VO order is by far the most frequent relative order in clauses with a verb selecting an internal argument (object). Bailyn (2012: 7) states that Russian has a head-initial VP. The vast majority of authors assume that the above-mentioned positional alternations are governed by information structure, but crucially they are also taken to be derived from the basic/neutral structure by additional operations (movement), although the exact motivation and landing sites of these movements might differ (Bailyn 1995, 2004; Junghanns and Zybatow 1995; and Slioussar 2007, among others). The following example contains both basic and derived word orders:

(3) Context: [The very beginning of a text] On November 29, 2018, an accident occurred at the address…

Mužčina	uronil	mjač	na	dorogu	za	nim	pobežala	sobaka.
man	dropped	ball	on	road	after	him	ran	dog

'A man dropped a ball on the road and a dog ran after it.'

(RUmo65FR_fwR,[3] RUEG Corpus)

In example (3) from a narration of a monolingual speaker, the SVO word order is used in an "all-new" context at the beginning of the story, while the following OVS order is used to introduce a new subject, a dog, into the discourse (see, e.g., Junghanns and Zybatow 1995; Kallestinova 2007; and Slioussar 2007 for detailed discussion). Since we will investigate VO vs. OV ordering in Russian quantitatively, a qualitative assessment of the data from an information-structural perspective must be left for further research. Although word order in monolingual Russian (under appropriate information structural conditions) is claimed to vary independently of the clause type, the variation is expected to be higher in the main clause than in the embedded one (Bailyn 2012). Zuban et al. (2021) showed that monolingual speakers of Russian produced fewer word order combinations in embedded clauses (namely, five) compared to the main clauses (namely, six). Moreover, the frequency of the SVO order produced by monolingual speakers was higher in embedded clauses (around 62%) compared to the main clauses (roughly 56%) (Zuban et al. 2021).

In addition to the information structure, word order in monolingual Russian is influenced by the object realization. Specifically, there is a strong

[2] But for an alternative view, see Haider and Szucsich (2022), in which the authors argue that Slavic languages are ambivalent between SVO and SOV (so-called T3-languages).

[3] The meaning of the participants' code is provided in Table 3.

tendency for the pronominal objects to be placed preverbally (Kallestinova 2007), as shown in example (4):

(4) Context: The first driver braked and that was it.

Vtoroj v nego vrezalsja.
second in him bumped

'The second one bumped into him.' (RUmo24FR_iwR)

Additionally, word order in monolingual Russian is influenced by other factors that will not be investigated in our study (e.g., intonation, grammatical weight of constituents, etc.; for more detailed information, see Bailyn 2012; King 1995; Laleko and Dubinina 2018; Lobanova 2011; Luchkina and Cole 2016).

Word order in HLs is claimed to be prone to change possibly induced by transfer. As a result, in languages with multiple word order options, the word order variation is reduced (Polinsky 2018). Such reduction of word order options is shown in several studies on different HLs such as heritage Korean (O'Grady et al. 2011), heritage Norwegian (Johannessen and Laake 2015), heritage Spanish (Cuza 2012; Cuza and Frank 2011), and heritage Hungarian (Fenyvesi 2005). However, Polinsky (2018: 273) points out that the word order reduction might not be a general outcome of language contact. Since the majority of studies focused on the populations whose ML is English, the behavior of HSs with a different ML is less predictable.

Several studies on word order in heritage Russian in the US found that HSs reduced their word order flexibility and increased the use of the SVO orders (e.g., Isurin 2005; Kisselev 2019; Laleko and Dubinina 2018; Polinsky 2006; Zuban et al. 2021).

Polinsky (2006: 237) states that HSs of Russian in the US clearly limit their word order repertoire by mainly producing SVO order with both nominal and pronominal arguments. Although Polinsky (2006) reports on the "frozen" word orders of HSs of Russian in the US, it has to be noted that these observations are not based on any numerical analysis of the word order patterns.

Isurin (2005) conducted a longitudinal study with one Russian girl who came to the US at the age of nine. The girl took part in different tasks that were repeated after a 9-month break. The results revealed that during the second elicitation session the girl produced more SVO and fewer inverted word orders compared to the first session. The results of the study were explained with reference to transfer from the ML and language internal factors such as frequency of inverted and basic word orders in the HL (for more details, see Isurin 2005: 1122).

Laleko and Dubinina (2018) investigated word order patterns in the "frog stories" of 21 HSs of Russian in the US and 19 monolingual speakers of Russian. The results of the study showed that HSs produced significantly more SVO

orders and significantly fewer non-SVO orders than the monolingual speakers (Laleko and Dubinina 2018: 202). Laleko and Dubinina (2018: 202) conclude that the relative flexibility of word order typical for monolingual Russian is reduced in HSs of Russian in the US.

Another study by Kisselev (2019) examined word order patterns of HSs of Russian in the US in written essays. It was found that HSs produced more SV and fewer VS orders than monolingual speakers. However, SV orders not only included the SVO order, but also all other word orders in which the verb follows the subject (i.e., SOV and OSV). Furthermore, although HSs were found to produce various word order combinations, they did not produce any verb-initial word orders, contrary to the monolingual speakers who produced all six grammatically possible word order combinations (Kisselev 2019: 161).

A more recent study by Zuban et al. (2021) examined the word order patterns of three speaker groups, namely HSs of Russian residing in the US and Germany, and monolingual speakers of Russian residing in Russia. Word order was investigated in a corpus of semi-spontaneous spoken and written narratives collected according to the "Language Situations" method (Wiese 2020). First, only the results related to the HSs in the US and monolingual speakers will be summarized, while the HSs in Germany will be discussed later. The results of the study showed that HSs residing in the US were found to produce significantly more SVO and significantly fewer OVS orders than monolingual speakers in the overall dataset. Furthermore, the split of the data into main and embedded clauses revealed further interesting findings. In the main clauses, both the HSs and monolingual speakers behaved similarly regarding the overall distribution of different word orders. In the embedded clauses, however, the HSs significantly differed from the monolingual speakers by predominantly producing the SVO order. Although monolingual speakers were found to increase the rate of SVO order in embedded clauses compared to main ones, that increase was not as substantial (from around 56% to 62%) as in the case of the HSs (from around 65% to 91%). Also, the HSs in the US produced fewer word order combinations in the embedded clauses (four combinations) than in the main ones (six combinations) (Zuban et al. 2021: 270).

Although word order in heritage Russian was often investigated for HSs with the ML English (e.g., Kisselev 2019; Laleko and Dubinina 2018; Polinsky 2006; Zuban et al. 2021), much less is known about HSs with the ML German. In one of a few studies on heritage Russian word order in bilingual speakers in Germany, Brehmer and Usanova (2015) examined word order patterns in written productions of HSs of Russian residing in Germany. Their study aimed at investigating the potential influence of German on the verb placement in declarative sentences in heritage Russian. The data were elicited from 20 HSs of Russian in Germany and 20 monolingual speakers from Russia. All participants were asked to write two texts representing

two different situational settings: a magazine article about how to create a boomerang (task 1: "Fast Catch Boomerang" (more formal academic language was expected)) and a story about how a boomerang can be used (task 2: "Boomerang in the Park" (elements of narrative speech were expected)).

The results of the study showed that HSs did not significantly differ from the monolinguals regarding the placement of V2 in main clauses (Brehmer and Usanova 2015: 174). As for the V-final position in embedded clauses, HSs produced significantly more V-final linearizations than monolingual speakers in narrative texts, but more formal texts did not differ across the two speaker groups (Brehmer and Usanova 2015: 179). Besides, in main clauses HSs also produced V-final orders significantly more frequently than monolingual speakers (for main clauses, Brehmer and Usanova (2015: 180) did not distinguish between different tasks in their analysis of main clauses). The results of the V-final position in embedded clauses were interpreted as language transfer from German to Russian, while the results of the V-final placement in main clauses were explained with reference to pragmatic unmarking (for details on pragmatic unmarking, see Brehmer and Usanova 2015: 182). Since Brehmer and Usanova (2015) do not explicitly focus on OV/VO orders, one can only derive the frequency of OV/VO from all word order patterns (OV: OV, SOV, OSV, OVS vs. VO: VO, SVO, VOS, VSO) (Brehmer and Usanova 2015: 179). However, Brehmer and Usanova (2015) do not report the frequencies of different word order patterns in embedded clauses. Thus, the distribution of OV and VO given in Table 1 is restricted to main clauses. It can be seen that HSs produced OV orders more frequently than monolingual speakers.

Table 1. Aggregated absolute and relative distributions of OV/VO patterns in Brehmer and Usanova (2015) in main clauses

Group	OV vs. VO, task 1	OV vs. VO, task 2
Bilinguals	31 (28.2%) vs. 79 (71.8%)	9 (18.4%) vs. 40 (81.6%)
Monolinguals	7 (4.6%) vs. 146 (95.4%)	2 (4.1%) vs. 47 (95.9%)

The study by Zuban et al. (2021), which was mentioned above, also investigated the word order patterns produced by HSs of Russian in Germany. The results of the study, contrary to the results of Brehmer and Usanova (2015), revealed that HSs in Germany showed no significant difference from monolingual speakers regarding the word order patterns produced in both main and embedded clauses. Furthermore, the study showed that clause type had an effect on the word order repertoire and on the frequency of SVO word order. Specifically, both heritage and monolingual speakers produced six word or-

der combinations in the main clauses and five word order combinations in the embedded clauses. Besides that, both speaker groups increased the frequency of SVO order in the embedded clauses compared to the main ones.

To sum up, the results regarding word order in heritage Russian in Germany are conflicting. On one hand, HSs were found to differ from monolingual speakers by producing more V-final linearizations in both main and embedded clauses (Brehmer and Usanova 2015). On the other hand, HSs were found to pattern with the monolingual speakers regarding the production of different word order patterns (Zuban et al. 2021). The studies on word order in heritage Russian in the US typically report that the HSs increase the proportion of SVO orders and decrease word order variation (Kisselev 2019; Laleko and Dubinina 2018). Interestingly, clause type was found to be a major factor that can influence the word order choice of the HSs (Zuban et al. 2021). However, the aforedescribed studies differ from each other in terms of methodology (e.g., tasks, elicitation techniques, participant sample, data annotation, and analysis), and this factor might have contributed to the divergent results of the studies.

3. Research Questions and Predictions

Taking the previous studies on word order in heritage Russian into consideration, we aim at contributing to our understanding of whether OV and VO orders are subject to transfer from the ML of English or German. The current study seeks to answer the following research questions:

RQ1 Do we find a crosslinguistic influence from the ML English on the choice of OV/VO orders in heritage Russian?

H1 HSs are expected to produce VO orders more frequently than monolingual speakers regardless of the object realization in both main and embedded clauses (Isurin 2005; Kisselev 2019; Laleko and Dubinina 2018; Polinsky 2006; Zuban et al. 2021).

RQ2 Do we find a crosslinguistic influence from the ML German on the choice of OV/VO orders in heritage Russian?

H2 HSs are expected to produce OV orders more frequently than monolingual speakers in embedded clauses with nominal objects (Brehmer and Usanova 2015).

4. Methodology

The present study analyzes 96 narrations drawn from the RUEG corpus (Wiese et al. 2021). The RUEG corpus was created within the Research Unit "Emerging Grammars in Language Contact Situations: A Comparative Approach"

(RUEG). RUEG investigates the linguistic systems and linguistic resources of HSs of different MLs and HLs (German, English, Greek, Russian, and Turkish). This article focuses on the narrations of the HSs of Russian. In §4.1 and §4.2, we will describe in detail the design of the RUEG subcorpus for Russian (RUEG-RU), including the data elicitation method, an overview of parameters and factors controlled, participants' coding and metadata, and available annotation.

4.1. Experimental Design

The data were elicited according to the Language Situations Method (Wiese 2020). The participants were shown a video of a fictional car accident. Multiple characters were involved in the accident: a young woman with a stroller accompanied by her son with a ball, another woman with a small dog who was putting her groceries in the trunk of her car, and two drivers (of a white car and of a blue car). The action unfolds as follows: the boy suddenly drops the ball, the dog runs after it, the cars stop abruptly, and the second car slightly bumps into the first one. Additionally, the woman drops a paper bag with her groceries, and they roll away. At the end of the story, everybody is safe, the boy helps the woman to collect her groceries, and the two car drivers call the police. After the participants watched the video, they were asked to narrate what they saw in four different communicative situations that differed with respect to their formality and mode:

1. **Informal spoken**: a voice message to a friend via WhatsApp
2. **Informal written**: a text message to a friend via WhatsApp
3. **Formal spoken**: a witness call to the police
4. **Formal written**: a written witness report to the police

HSs had two elicitation sessions, one in their HL and one in their ML. These sessions were separated from each other by at least three days. Each elicitation was performed by two different elicitors, one for the formal and one for the informal situation. The elicitations took place in two different rooms: one room looked more formal (similar to a typical office or a school room), while another room looked more informal (with some tablecloths, snacks, and juice). The participants had a break between the formal and the informal elicitations. Besides this, the order of four different narrations was randomized. Finally, the participants filled out a background questionnaire about their language use and input, family, and socio-economic status.

4.2. Participants

Table 2 gives detailed information about the participants that formed part of the current study: country of elicitation, number of participants, speakers' gender, number of narrations, speakers' age per group, and age of onset (AoO) of the ML.[4]

Table 2. Participants

Country	Speakers	Females	Narrations	Mean age	AoO ML
Germany	8	4	32	17.0, SD = 0.84	3.1, SD = 1.3
Russia	8	4	32	16.6, SD = 0.49	n/a
US	8	4	32	15.7, SD = 1.51	4.2, SD = 2.26

The participants in the US were recruited in the greater Washington area mainly with the help of the Metaphor Academic Center for Russian Language & Culture in Virginia. The participants in Germany were recruited in the Berlin and Brandenburg urban area via calls in mailing lists, social media, educational institutions (e.g., schools, universities, and language courses), and public organizations (libraries, youth and sport clubs, shopping centers, and consulting offices). The HSs of Russian in the US and Germany were invited to participate if they grew up in one of these two countries speaking Russian with at least one Russian-speaking caregiver at home. Only candidates with one HL were admitted to the experiment.

All HSs were either born in the US or Germany or moved there before the age of six. Furthermore, all HSs in our sample still attended school or just finished school, but more importantly, still lived with their Russian-speaking caregivers at the moment of testing. In addition, none of the HSs attended bilingual schools, but some of the participants learned Russian in a language school (one HS in the US) or in school (five HSs in Germany) as a foreign language. Almost all participants could write in the Russian Cyrillic

[4] The participants' sample used in the present study is the same as the one in the study by Zuban et al. (2021). However, in contrast to Zuban et al. (2021), the present study investigates another syntactic phenomenon involving a different set of data. Specifically, Zuban et al. (2021) investigated six trivalent word order patterns (SVO, SOV, VOS, VSO, OSV, OVS), whereas the present study analyzes OV and VO word orders irrespective of the subject position if applicable. Also, a more advanced statistical modeling allowing us to control for several factors and a different annotation scheme were applied in the present study.

script.[5] Also, four HSs in the US and two HSs in Germany participated in other Russian-speaking activities such as music lessons, dance classes, etc. As can be seen in Table 2, the participants in both HSs groups are quite homogeneous regarding the AoO in their ML, which is also true for their language background in Russian, although HSs in Germany may have slightly more exposure to formal education in Russian than those in the US.

As for the ML of the HSs, some of the participants (a total of three HSs in the US and Germany) started acquiring their ML from birth, while other HSs started acquiring their ML later than their HL. As can be seen in Table 2, on average HSs in the US started to acquire their ML slightly later than HSs in Germany, namely, at the age of 4.2 and at the age of 3.1, respectively.

In the background questionnaire, the bilingual participants were asked about their use of different types of media in their HL and ML, including video, audio, writing, and reading. The HSs of both groups reported usage of the media more frequently in their ML than in their HL, suggesting that majority English and German are the main languages in their daily life.

Next to the self-reports about the use of different media, the HSs assessed their knowledge of Russian in four language domains: understanding, speaking, reading, and writing. The HSs of both groups reported that it was fairly easy for them to understand, speak, and read in Russian. As for the writing, while the HSs in the US found it easy to write in Russian, the HSs in Germany found it more difficult.

Finally, the HSs were asked about the use of their HL and ML with their core family, i.e., which language/languages is/are typically used to communicate with their parents. According to their self-reports, the HSs in the US were exposed to Russian at home to a lesser degree than the HSs in Germany.

Monolingual participants were recruited in St. Petersburg, Russia, primarily through calls in social media and educational institutions. Monolingual speakers were considered to be those whose mother tongue was the only one spoken in their household but who might possess knowledge of foreign languages acquired through language instruction. All monolingual speakers were either born in St. Petersburg or arrived there as young children from other regions of Russia or the countries of the former Soviet Union.

Since the participants' proficiency in Russian was not explicitly tested, we calculated an individual innovation ratio score as a language performance measure for each monolingual and HS in the sample. More specifically, we manually annotated the innovations in the domain of morphology and lexicon in the oral and written productions according to the classification of the Russian Learner Corpus (Rakhilina et al. 2016). Then, we calculated the individual innovation ratio score by dividing the overall number of innovations

[5] Those HSs who were not able to write in Cyrillic were allowed to write in Latin script.

produced by one speaker by the overall number of words per narration produced by the same speaker. The individual innovation ratio was taken as an indicator of the language performance.

The individual innovation ratio scores were compared among the speakers of the three groups. First, the HSs in the US and in Germany had comparable innovation ratio scores (W = 41, p = 0.382).[6] Next, the HSs in the US showed innovation scores that were significantly higher than the scores of the monolingual speakers (W = 61, p = 0.001). Finally, the HSs in Germany also showed innovation scores that were significantly higher than the scores of the monolingual speakers (t = 2.513, df = 14, p = 0.02). To sum up, both HS groups demonstrated similar innovation ratio scores, while monolingual speakers showed significantly fewer innovations compared to the HSs.

4.3. Data Analysis

This section provides information about the basic annotation layers of the RUEG corpus and additional manual syntactic annotations[7] of the data. In the RUEG corpus, all the collected data were anonymized, and each participant was assigned a code, as shown in Table 3 on the following page. The data were annotated on the so-called diplomatic layer, dipl. This layer provides word-level segmentation without taking the orthographic rules of standard Russian into account. The spoken data were annotated in Praat (Boersma 2001), while the written texts were automatically transferred to the dipl layer without any modifications. Some texts were written in the Latin alphabet. Each document was transcribed by two student assistants on the basis of the dual control principle in order to minimize typing errors and increase the accuracy. The sample used in the present study was drawn from the dipl layer of the RUEG-RU corpus.

[6] The Unpaired Wilcoxon Rank Sum Test was chosen for the comparisons between HS groups and between the HSs in the US and monolingual speakers since the data of the HSs in the US are not normally distributed. A two-tailed unpaired t-test was chosen for the comparison between the HSs in Germany and monolingual speakers since the data of these two speaker groups are normally distributed.

[7] Semi-automatic syntax annotation of the RUEG-RU corpus in SUD framework (Gerdes et al. 2018, 2019; Gerdes and Kahane 2016) is still in progress.

Table 3. Participants' codes

RU, DE, US	country of elicitation: Russia, Germany, US
mo, bi	mode of language acquisition: monolingual, bilingual
01–99	participants' number
F, M	gender: female, male
R	heritage language or, for monolinguals, their only language: Russian
fs	mode of elicitation: formal spoken
is	mode of elicitation: informal spoken
fw	mode of elicitation: formal written
iw	mode of elicitation: informal written
R	language of elicitation: Russian

After the data had been exported from the corpus, they were enriched with manual annotations for syntactic functions. The annotations were carried out by two PhD students of linguistics who are L1 speakers of Russian. The annotations were crosschecked by the two aforedescribed students. First, the utterances were tagged for verb type, word order, and clause type (main vs. embedded clause),[8] as shown in the following examples:

(5) Main clause

…sobaka:	(-) è:	pobežala	za	mjačom.
dog	uh	run	after	ball

'…the dog uh ran after the ball.' (USbi63MR_fsR)

(6) Embedded clause

Kogda	sobaka	uvidela	futbol'nyj	mjač…
when	dog	saw	soccer	ball

'When the dog saw the soccer ball…' (USbi68FR_fwR)

Verbless clauses and clauses with no object were excluded from further analysis. Second, for the declarative clauses with OV or VO orders, the object

[8] See for details Zuban et al. (2021).

realization (nominal or pronominal) was annotated. The final dataset consists of 96 documents and 1,010 data points with either a finite or non-finite verb and at least one pronominal or nominal object (direct and/or oblique) each. The examples below illustrate typical OV and VO clauses in the dataset:

(7) …i ženščina eë vzjala na ruki.
and woman her took on arms
O, pronominal **V**
'…and the woman took her (it) in her arms.' (USbi66FR_fwR)

(8) …on poterjal mjač.
he lost ball
V **O, nominal**
'…he lost the ball.' (DEbi58MR_iwR)

5. Results

As mentioned in §4.3, 1,010 OV and VO clauses were included in the dataset for further analysis. Table 4 provides absolute and relative frequencies of OV and VO orders aggregated by the country of elicitation. The numerical values provide an overall impression that monolingual speakers and the HSs in Germany behave similarly to each other, while the HSs in the US show a higher rate of VO orders compared to the two aforementioned groups.

Table 4. Distribution of OV/VO word orders

Country of elicitation	OV, abs. (rel.)	VO, abs. (rel.)	SUM
Germany	122 (30.5%)	278 (69.5%)	400
Russia	95 (33.2%)	191 (66.8%)	286
US	79 (24.4%)	245 (75.6%)	324

Tables 5 and 6 illustrate the distribution of OV and VO orders grouped according to the object realization, i.e., nominal and pronominal objects. It can be seen (Table 5) that all speaker groups prefer VO orders with nominal objects. As for the pronominal objects (Table 6), although all speaker groups prefer OV orders, this preference is less pronounced in the heritage group from the US than in the two other groups. This descriptive analysis gives an

impression that the first observed differences between some speaker groups might also be statistically significant.

Table 5. Distribution of OV/VO with nominal objects

Country of elicitation	OV, abs. (rel.)	VO, abs. (rel.)	SUM
Germany	43 (14%)	264 (86%)	307
Russia	51 (22.9%)	172 (77.1%)	223
US	41 (15.9%)	217 (84.1%)	258

Table 6. Distribution of OV/VO with pronominal objects

Country of elicitation	OV, abs. (rel.)	VO, abs. (rel.)	SUM
Germany	79 (84.9%)	14 (15.1%)	93
Russia	44 (69.8%)	19 (30.2%)	63
US	38 (57.6%)	28 (42.4%)	66

Finally, tables 7 and 8 show the distribution of OV and VO orders in main and embedded clauses across different speaker groups. It can be seen that all speaker groups prefer VO orders over the OV orders across different clause types. Besides, all speaker groups demonstrate an increase of the VO orders in the embedded clauses, and this increase is particularly substantial for the HSs in the US.

Table 7. Distribution of OV/VO in main clauses

Country of elicitation	OV, abs. (rel.)	VO, abs. (rel.)	SUM
Germany	86 (32.3%)	180 (67.7%)	266
Russia	81 (35.7%)	146 (64.3%)	227
US	73 (29.2%)	177 (70.8%)	250

Furthermore, we applied a generalized linear mixed-effects model which has the following advantages over the descriptive analysis. First, statistical

modeling can account for the speaker's random effect. Besides, such modeling allows investigation of multiple fixed effects. Furthermore, the generalized linear mixed-effect model is especially attractive when the data of each participant are not independent since the model can deal with clusters of dependent data by including the preferences of every individual speaker in the analysis (Winter 2019).

Table 8. Distribution of OV/VO in embedded clauses

Country of elicitation	OV, abs. (rel.)	VO, abs. (rel.)	SUM
Germany	36 (26.9%)	98 (73.1%)	134
Russia	14 (23.7%)	45 (76.3%)	59
US	6 (8.1%)	68 (91.9%)	74

The following predictors were included in the model: *country of elicitation* (Germany, US, and Russia), *clause type* (main vs. embedded clause), *object realization* (nominal vs. pronominal object), *formality* (formal vs. informal), and *mode* (spoken vs. written).[9] The model included the interactions of *country of elicitation* with the fixed effects of *clause type* and *object realization* since these interactions were directly relevant for our research questions and hypotheses.[10] The fixed effects were contrast-coded using sum contrast coding (either –0.5 or 0.5). For post-hoc tests of interactions (pairwise Tukey test), R's (R Core Team 2021) emmeans package (Lenth 2023) was used. *Participants* were treated as a random effect (SD = 0.5283). Furthermore, the main factor of interest, *country of elicitation,* had three levels (Germany, US, Russia), one of which was set as a reference level, namely Russia, since the present study compares the heritage varieties of Russian with the monolingual one.

The model revealed two main effects and one interaction (for the model summary and confidence intervals, see Table 9 and Table 10 in the Appendix). First, there was the main effect of *clause type*: the likelihood of producing an OV word order was lower in the embedded clauses than in the main clauses (see Figure 2 in the Appendix). In the embedded clauses, VO orders were almost always preferred by all speaker groups ($z = –2.57$, $p < 0.05$). Second, there

[9] The last two predictors, *formality* and *mode,* were included in the model due to the experimental setup that explicitly tested the role of formality and mode in the linguistic behavior of heritage and monolingual speakers.

[10] The predictors *formality* and *mode* could not be included as interactions with *country of elicitation* since the model failed to converge.

was the main effect of *object realization*: all speaker groups were more likely to produce an OV order with the pronominal object, while the speakers' preference shifted towards a VO order when the object was realized by a noun ($z = -6.33$, $p < 0.001$) (see Figure 3 in the Appendix).

Furthermore, there was a significant two-way interaction between *country of elicitation* and *object realization*. For the interaction of country and object realization, the Tukey multiple comparison test was run. The Tukey's adjustment revealed the following meaningful significant results. First, as reported earlier, all speaker groups preferred an OV order with the pronominal object, and they preferred a VO order with a nominal object (HSs in the US: $z = -6.13$, $p < 0.001$; HSs in Germany: $z = -10.38$, $p < 0.001$; monolingual speakers: $z = -6.33$, $p < 0.001$). Second, the HSs in Germany produced an OV order with the pronominal object significantly more frequently than the HSs in the US, and the former group produced a VO order with a pronominal object significantly less frequently than the latter group ($z = 3.81$, $p < 0.01$). Note that the HSs of both groups did not differ from the monolingual speakers regarding their preference for the pronominal object with the OV orders (HSs in the US/monolinguals: $z = 1.6$, $p = 0.6013$; HSs in Germany/monolinguals: $z = -2.27$, $p = 0.2037$) (see Figure 1).

To sum up, *clause type* and *object realization* were found to play an important role in the OV/VO choice by all speaker groups in a similar manner. Besides, the HSs in Germany differed from the HSs in the US by producing OV orders with the pronominal object more frequently. Finally, both HS groups were similar to the monolingual speakers regarding their preference for the OV/VO orders and the effect of *clause type* and *object realization*. The predictors *formality* and *mode* turned out to be insignificant.

6. Discussion

The present study investigated the OV/VO choice in heritage and monolingual Russian, using semi-spontaneous, manually annotated data and applying a generalized linear mixed-effects model. In the following, the research questions and hypotheses will be addressed in detail.

The first research question focused on the possible influence of the majority English on the choice of OV/VO orders in productions of the HSs in the US. Hypothesis 1 stated that HSs would produce VO orders more frequently than monolingual speakers regardless of the object realization in both main and embedded clauses.

The data refute Hypothesis 1 since there were no significant differences between the HSs in the US and the monolingual speakers. Apparently, the numerical differences that arose in the descriptive analysis of the data (e.g., the overall distribution of OV/VO orders between the HSs in the US and the monolingual speakers or the potential differences between these two speaker

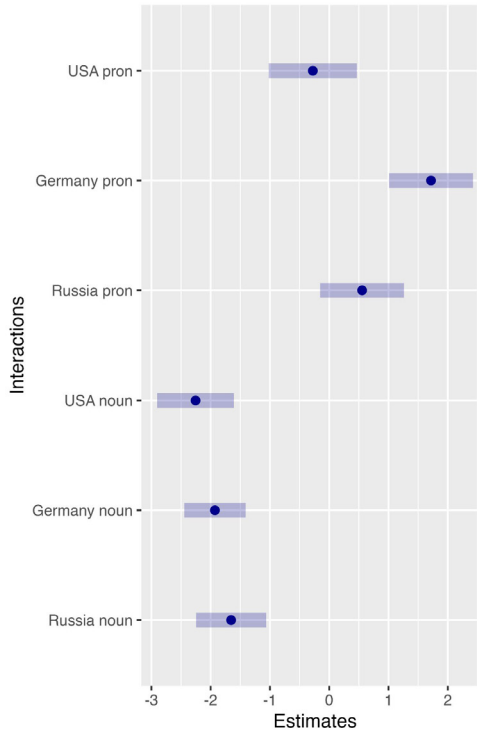


Figure 1. Likelihood of an OV word order:
Interaction of country of elicitation and object realization

groups in the embedded clauses or with the pronominal objects) turned out to be insignificant when speakers' variation was taken into account. This result contradicts those of the previous studies (e.g., Isurin 2005; Kisselev 2019; Laleko and Dubinina 2018; Polinsky 2006; and Zuban et al. 2021).

The differences between the abovementioned research and the results of this study can be explained with reference to the experimental and statistical methodology.

First, the annotation schemes in the abovementioned studies differ from the annotation scheme used in the present study. The study by Isurin (2005) includes the trivalent orders, namely SVO, SOV, SVX, SXV, VSX, XVS, and XSV, as well as two bivalent orders, SV and VS. The (X)OV(X) and (X)VO(X) orders are not mentioned, and it remains unclear whether they did not appear in the data or whether they were not annotated. Polinsky (2006: 238) does not provide any details about the annotation of word order but explicitly points out that the overuse of SVO needs to be considered cautiously due to the difficulties of annotating word order (the speech of HSs typically contains

many pauses). Laleko and Dubinina (2018) annotated clauses with a verbal predicate and at least one of the following constituents: subject, object, or a verbal modifier (adverbial). Clauses with the following word order were considered to be canonical: (S) (Modifier) V (DO) (IO) (Modifier). Clauses with at least one constituent aligned in a different way relative to the verbal predicate were considered to be non-canonical (Laleko and Dubinina 2018: 202). Thus, canonical clauses could contain, for instance, orders without any object (e.g., SV or only with a Modifier V). The annotation scheme in the study by Kisselev (2019) is based on six trivalent word order patterns, SVO, SOV, VSO, VOS, OSV, and OVS, and two bivalent word orders, SV and VS. Thus, OV and VO clauses without subjects were not included in the analysis. This is also true for the study by Zuban et al. (2021), where only trivalent word orders with realized subjects were accounted for.

All in all, the annotation schemes described above are only partially equivalent to the annotation scheme used in the present study (i.e., OV and VO orders with the overtly expressed object and other constituents being optional). Thus, the reoccurring result of the previous studies about the prevalence of the SVO order does not necessarily imply the prevalence of the VO order over OV.

Apart from the annotation scheme, the data analysis might have led to the discrepancies between the current study and the previous studies. Specifically, the present study, contrary to Kisselev (2019), Laleko and Dubinina (2018), and Zuban et al. (2021),[11] used statistical modeling that took into account the individual speaker variation. Corpus studies that look at the *inter-group* variation without the *within-group* variation have been criticized since, in this case, the independence assumption is violated, making the results vulnerable to the Type I error (false positive) (Baayen et al. 2008; Brezina and Meyerhoff 2014; and Winter 2011, 2019).

The second research question focused on the possible influence of the majority German on the choice of OV/VO orders in productions of the HSs in Germany. Hypothesis 2 stated that HSs would produce OV orders more frequently than monolingual speakers in embedded clauses with nominal objects.

Our data seem to provide evidence against Hypothesis 2 since there were generally no significant differences between the HSs in Germany and the monolingual speakers. Specifically, the results of the study showed that in embedded clauses HSs in Germany favor VO word orders with nominal objects, similar to monolingual speakers. Thus, no support for the possible influence of the ML German was found. These results contradict those of Brehmer and Usanova (2015). Although the current study and Brehmer and Usanova

[11] This does not apply to Isurin (2005) and Polinsky (2006) since the former study investigates one single participant while the latter study is a survey article.

(2015) used a similar approach to annotate the data, these two studies differed regarding the statistical method. Brehmer and Usanova (2015), contrary to this study, did not take the individual speaker variation into consideration.

Given these two aspects, namely different annotation schemes and statistical approaches, the results of the present study are not surprising, since neither of these two aspects appears to be fully comparable to those of the previous studies. Thus, the results of the present investigation are novel, inviting replications and further research using the advanced methodology.

In the following, the results will be discussed in terms of their general impact. First and foremost, our results are in line with previous research on monolingual Russian confirming a strong tendency for a postverbal linearization of nominal objects and a preverbal linearization of pronominal objects (Kallestinova 2007). Importantly, we found that this trend is not only true for the monolingual speakers but for the HSs as well. Apparently, this feature remains stable even under language contact. These results clearly speak against transfer effects from MLs.

Although we did not aim at testing the Interface Hypothesis (Sorace 2011), we will briefly discuss possible consequences for it. Our results can be interpreted in two ways. They may be taken to speak in favor of considering head-directionality within the VP as being the decisive factor for the distribution of OV/VO patterns in Russian, i.e., under this view, the overall OV/VO distribution is rather determined by core-grammatical (internal) principles. Alternatively, if one maintains the view that the overall OV/VO distribution is determined by information structural conditions, these results together with some previous studies may raise some questions about the predictive power of the Interface Hypothesis since external interfaces are claimed to be prone to variability under language contact (e.g., Slabakova et al. 2012 on L2 learners; Jin et al. 2022; Leal et al. 2014; and Méndez et al. 2015 on HSs). It might be the case that not all external interfaces are similarly challenging for HSs. It is also possible that the distinction between the internal and external interface is not a completely reliable predictor for the linguistic behavior of bilingual speakers (both L2 learners and HSs).

Furthermore, our results indicate that OV/VO choice is clause-type sensitive. Specifically, the likelihood of producing an OV order was lower in the embedded clauses than in the main clauses. This result suggests that both heritage and monolingual speakers tend to use (S)VO more frequently in embedded clauses than in main clauses. In fact, a closer look at embedded clauses with VO orders reveals that almost all the word orders (210 out of 211)[12] are of (S)VO(O) type. For monolingual Russian, it was already stated that word order in embedded clauses is less varied compared to main clauses, probably due to the differences of the two clause types in terms of their

[12] One clause uttered by a monolingual speaker has a VOS word order.

discourse structure (Bailyn 2012 and Zuban et al. 2021). Our findings provide support for this claim. However, one of the limitations of the present study is that the data were not controlled for information-structural appropriateness. Since word order patterns of HSs of Russian were not always found to be contextually appropriate (Brehmer and Usanova 2015; Laleko and Dubinina 2018), we assume that it might also be true for our data sample. Moreover, the authors are aware of the fact that the dataset used in the present study is not exhaustive in its size and homogeneity. Once the syntactic annotation in the RUEG-RU corpus is complete, further studies have to be conducted to reproduce the results reported in the present paper on a larger number of observations. Also, other relevant factors that can influence word order in Russian, such as intonation, animacy of the object, or grammatical weight of the constituents, were not considered in the present research. Future studies should look at these factors and their possible interplay since they may give new additional insights into the OV/VO choice.

The results of our study can be explained by the participants' high engagement with Russian. According to the self-reports, the HSs participating in our study actively use Russian in their daily life from birth (communication with caregivers, extra activities in Russian, media use in Russian) and almost all HSs can write in the Cyrillic script. This is especially remarkable if one compares the HSs' profiles from the present study with those in the studies carried out by Isurin (2005) and Polinsky (2006), where the participants were barely involved in any kind of activities using their HL. Finally, the HSs of both groups were found to differ from each other regarding their preference for OV/VO orders with the pronominal object. These results indicate that the HSs of Russian residing in different countries may behave differently from each other. It is difficult to pinpoint the exact reason behind these results. It is possible that the differences in the speaker communities or some other extra-linguistic factors contributed to the differences between the two HS groups.

To sum up, the HSs in the US and Germany were similar to the monolingual speakers of Russian regarding their choice of OV/VO orders. On one hand, these results call for a new evaluation of the previous studies and their implications for the possible cross-linguistic effects from the MLs. We believe that language transfer may arise in some particular populations of HSs (presumably those that are not extensively engaged in the practices involved in using Russian, e.g., speaking, listening, overhearing, responding, code-switching, etc.), while for other HS populations that are highly engaged with their HL, the transfer is not necessarily at play. On the other hand, we suggest that the factors that are relevant for word order choice in monolingual Russian (i.e., object realization and clause type) are also relevant for heritage Russian in the US and Germany to a similar extent.

Sources

Wiese, Heike, Artemis Alexiadou, Shanley Allen, et al. (2021) RUEG Corpus. Available at: https://zenodo.org/record/5808870.

References

Aboh, Enoch Oladé. (2015) *The emergence of hybrid grammars: Language contact and change*. Cambridge, UK: Cambridge University Press.

Anstatt, Tanja. (2011) "Russisch in der zweiten Generation. Zur Sprachsituation von Jugendlichen aus russischsprachigen Familien in Deutschland". Ludwig M. Eichinger, Albrecht Plewnia, and Melanie Steinle, eds. *Sprache und Integration: Über Mehrsprachigkeit und Migration*. Tübingen: A. Francke Verlag, 101–28.

Aref'ev, Aleksandr L. (2012) *Russkij jazyk na rubeže XX–XXI vekov*. Moscow: Centr social'nogo prognozirovanija i marketinga.

Baayen, R. Harald, Douglas J. Davidson, and Douglas M. Bates. (2008) "Mixed-effects modeling with crossed random effects for subjects and items". Special issue: Emerging data analysis, *Journal of memory and language* 59(4): 390–412. DOI 10.1016/j.jml.2007.12.005.

Bailyn, John. (1995) *A configurational approach to Russian free word order.* PhD dissertation, Cornell University.

————. (2004) "Generalized inversion". *Natural language & linguistic theory* 22(1): 1–50.

————. (2012) *The syntax of Russian*. Cambridge, UK: Cambridge University Press.

Baker, Mark. (2001) *The atoms of language: The mind's hidden rules of grammar.* New York: Basic Books.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. (2015) "Fitting linear mixed-effects models using lme4". *Journal of statistical software* 67(1): 1–48. DOI 10.18637/jss.v067.i01.

Benmamoun, Elabbas, Silvina Montrul, and Maria Polinsky. (2013) "Heritage languages and their speakers: Opportunities and challenges for linguistics". *Theoretical linguistics* 39(3–4): 129–81. DOI 10.1515/tl-2013-0009.

Birner, Betty, and Gregory Ward. (1998) *Information status and noncanonical word order in English*. Amsterdam: John Benjamins.

Boersma, Paul. (2001) "Praat, a system for doing phonetics by computer". *Glot international* 5(9–10): 341–45.

Brehmer, Bernhard. (2007) "Sprechen Sie Qwelja? Formen und Folgen russischdeutscher Zweisprachigkeit in Deutschland". Tanja Anstatt, ed. *Mehrsprachigkeit Bei Kindern und Erwachsenen. Erwerb, Formen, Förderung.* Tübingen: Narr, 163–86.

Brehmer, Bernhard, and Grit Mehlhorn. (2015) "Russisch als Herkunftssprache in Deutschland. Ein holistischer Ansatz zur Erforschung des Potenzials von Herkunftssprachen". *Zeitschrift für Fremdsprachenforschung* 26(1): 83–121.

Brehmer, Bernhard, and Irina Usanova. (2015) "Let's fix it? Cross-linguistic influence in word order patterns of Russian heritage speakers in Germany". Hagen Peukert, ed. *Transfer effects in multilingual language development.* Amsterdam: John Benjamins, 159–86.

Brezina, Vaclav, and Miriam Meyerhoff. (2014) "Significant or random. A critical review of sociolinguistic generalisations based on large corpora". *International journal of corpus linguistics* 19(1): 1–28.

Broekhuis, Hans. (2006) "The universal base hypothesis: VO or OV?" Jeroen van de Weijer and Bettelou Los, eds. *Linguistics in the Netherlands 2006.* Amsterdam: John Benjamins, 28–39.

Cho, Grace, Fay Shin, and Stephen Krashen. (2004) "What do we know about heritage languages? What do we need to know about them?" *Multicultural education* 11(4): 23–26.

Cuza, Alejandro. (2012) "Crosslinguistic influence at the syntax proper: Interrogative subject-verb inversion in heritage Spanish". *International journal of bilingualism* 17(1): 71–96. DOI 10.1177/1367006911432619.

Cuza, Alejandro, and Joshua Frank. (2011) "Transfer effects at the syntax-semantics interface: The case of double *que* questions in heritage Spanish". *Heritage language journal* 8(1): 66–89.

Dubinina, Irina, and Maria Polinsky. (2013) "Russian in the USA". Michael Moser and Maria Polinsky, eds. *Slavic languages in migration*. Münster: Lit Verlag.

Eppler, Eva. (1999) "Word order in German-English mixed discourse". *UCL working papers in linguistics* 11: 285–308.

Fenyvesi, Anna. (2005) "Hungarian in the United States". Anna Fenyvesi, ed. *Hungarian language contact outside Hungary: Studies in Hungarian as a minority language.* Amsterdam: John Benjamins, 265–318.

Gagarina, Natalia. (2014) "Die Erstsprache bei Mehrsprachigen im Migrationskontext". Solveig Chilla and Stefanie Haberzettl, eds. *Mehrsprachigkeit*. Vienna: Elsevier, 19–37.

————. (2017) "Monolingualer und bilingualer Erstspracherwerb des Russischen: Ein Überblick". Kai Witzlack-Makarevich and Nadja Wulff, eds. *Handbuch des Russischen in Deutschland: Migration – Mehrsprachigkeit – Spracherwerb.* Berlin: Frank & Timme, 393–410.

Gagarina, Natalia, and Annegret Klassert. (2018) "Input dominance and development of home language in Russian-German bilinguals". *Frontiers in communication* 3: 40. DOI 10.3389/fcomm.2018.00040.

Gagarina, Natalia, Maria Martynova, Irina Sekerina, and Luka Szucsich. (2020) "From first verbs to adulthood: Aspect in heritage Russian in the US and Germany". Evgenij Vasil'evič Golovko, Elena Viktorovna Gorbova, Petr Aleksandrovič Kočarov, Viktor Samuilovič Xrakovskij, and Oksana Jur'evna Čujkova, eds. *Vzaimodejstvie aspekta so smežnymi kategorijami. Materialy VII Meždunarodnoj konferencii Komissii po aspektologii Meždunarodnogo komiteta slavistov.* Petersburg: RGPU im. A. I. Gercena, 112–17.

Gärtner, Hans-Martin. (2000) "Are there v2 relative clauses in German?" *The journal of comparative Germanic linguistics* 3(2): 97–141.

Gerdes, Kim, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. (2018) "SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD". Universal Dependencies Workshop 2018, Brussels, Belgium. Available at: https://hal.inria.fr/hal-01930614. Last accessed 31 October 2024.

————. (2019) "Improving Surface-Syntactic Universal Dependencies (SUD): Surface-syntactic relations and deep syntactic features". TLT 2019—18th International Workshop on Treebanks and Linguistic Theories, Paris, France. Available at: https://hal.inria.fr/hal-02266003. Last accessed 31 October 2024.

Gerdes, Kim, and Sylvain Kahane. (2016) "Dependency annotation choices: Assessing theoretical and practical issues of universal dependencies". Annemarie Friedrich and Katrin Tomanek, eds. *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016).* Berlin: Association for Computational Linguistics, 131–40. DOI 10.18653/v1/W16-1715.

Guijarro-Fuentes, Pedro, and Katrin Schmitz. (2015) "The nature and nurture of heritage language acquisition". *Lingua* 164: 239–50. DOI 10.1016/j.lingua.2015.05.008.

Haider, Hubert. (2012) *The syntax of German*. Cambridge, UK: Cambridge University Press.

————. (2015) "Head directionality - in syntax and morphology". Antonio Fábregas, Jaume Mateu, and Michael Putnam, eds. *Contemporary linguistic parameters*. London: Bloomsbury Academic, 73–97.

Haider, Hubert, and Luka Szucsich. (2022) "Slavic languages–'SVO' languages without SVO qualities?" *Theoretical linguistics* 48(1–2): 1–39.

Hamann, Katharina, Kai Witzlack-Makarevich, and Nadja Wulff. (2019) "Russian in Germany". Arto Mustajoki, Ekaterina Protassova, and Maria Yelenevskaya, eds. *The soft power of the Russian language: Pluricentricity, politics and policies*. London: Routledge, 163–74. DOI 10.4324/9780429061110-14.

Hothorn, Torsten, Frank Bretz, and Peter Westfall. (2008) "Simultaneous inference in general parametric models". *Biometrical journal* 50(3): 346–63.

Isurin, Ludmila. (2005) "Cross linguistic transfer in word order: Evidence from L1 forgetting and L2 acquisition". James Cohen, Kara T. McAlister, Kellie Rolstad, and Jeff MacSwan, eds. *Proceedings of the 4th International Symposium on Bilingualism*. Somerville, MA: Cascadilla Press, 1115–30. Available at: http://www.cascadilla.com/isb4.html. Last accessed 31 October 2024.

————. (2011) *Russian diaspora: Culture, identity, and language change*. Berlin, New York: De Gruyter Mouton. DOI 10.1515/9781934078457.

Isurin, Ludmila, and Tanya Ivanova-Sullivan. (2008) "Lost in between: The case of Russian heritage speakers". *Heritage language journal* 6(1): 72–104.

Jasinskaja, Katja, and Radek Šimík. (forthcoming) "Slavonic free word order". Jan Fellerer and Neil Bermel, eds. *The Oxford guide to Slavonic languages*. Oxford, UK: Oxford University Press.

Jin, Jing, Sihui Echo Ke, and John Chi-Kin Lee. (2022) "Language interfaces in adult heritage language acquisition: A study on encoding of nominal reference in Mandarin Chinese as a heritage language". *Frontiers in psychology* 12. DOI 10.3389/fpsyg.2021.790102.

Johannessen, Janne Bondi, and Signe Laake. (2015) "On two myths of the Norwegian language in America". Janne Bondi Johannessen and Joseph C. Salmons, eds. *Germanic heritage languages in North America*. Amsterdam: John Benjamins, 299–322. Available at: https://www.jbe-platform.com/content/books/9789027268198-silv.18.14joh.

Junghanns, Uwe, and Gerhild Zybatow. (1995) "Syntax and information structure of Russian clauses". Wayles Browne, Ewa Dornisch, Natasha Kondrashova, and Draga Zec, eds. *Annual Workshop on Formal Approaches to Slavic Linguistics: The Cornell Meeting*. Ann Arbor: Michigan Slavic Publications, 289–319.

Kallestinova, Elena D. (2007) *Aspects of word order in Russian*. Iowa City: University of Iowa.

Kayne, Richard. (1994) *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Kempen, Gerard, and Karin Harbusch. (2019) "Mutual attraction between high-frequency verbs and clause types with finite verbs in early positions: Corpus evidence from spoken English, Dutch, and German". *Language, cognition and neuroscience* 34(9): 1140–51. DOI 10.1080/23273798.2019.1642498.

King, Tracy H. (1995) *Configuring topic and focus in Russian*. Stanford, CA: CLSI Publications.

Kisselev, Olesya. (2019) "Word order patterns in the writing of heritage and second language learners of Russian". *Heritage language journal* 69: 149–74.

Kroch, Anthony, and Ann Taylor. (2000) "Verb-object order in early middle English". *Diachronic syntax: Models and mechanisms* 132: 163.

Laleko, Oksana. (2010) *The syntax-pragmatics interface in language loss: Covert restructuring of aspect in heritage Russian*. PhD dissertation, University of Minnesota.

————. (2011) "Restructuring of verbal aspect in heritage Russian: Beyond lexicalization". *International journal of language studies* 5(3): 13–26.

————. (2013) "Assessing heritage language vitality: Russian in the United States". *Heritage language journal* 10(3): 89–102.

————. (2015) "From privative to equipollent: Incipient changes in the aspectual system of heritage Russian". Gerhild Zybatow, Petr Biskup, Marcel Guhl, Claudia Hurtig, Olav Mueller-Reichau, and Maria Yastrebova, eds. *Slavic grammar from a formal perspective*. Frankfurt am Main: Peter Lang, 273–86.

————. (2018) "What is difficult about grammatical gender? Evidence from heritage Russian". *Journal of language contact* 11(2): 233–67. DOI 10.1163/19552629-01102004.

————. (2019) "Resolving indeterminacy in gender agreement: Comparing heritage speakers and L2 learners of Russian". *Heritage language journal* 16(2): 151–82.

Laleko, Oksana, and Irina Dubinina. (2018) "Word order production in heritage Russian: Perspectives from linguistics and pedagogy". Susan Bauckus and Susan Kresin, eds. *Connecting across languages and cultures: A heritage language Festschrift in honor of Olga Kagan*. Bloomington, IN: Slavica Publishers, 191–215.

Laleko, Oksana, and Maria Polinsky. (2017) "Silence is difficult: On missing elements in bilingual grammars". *Zeitschrift für Sprachwissenschaft* 36(1): 135–63. DOI 10.1515/zfs-2017-0007.

Leal, Tania, Jason Rothman, and Roumyana Slabakova. (2014) "A rare structure at the syntax-discourse interface: Heritage and Spanish-dominant native speakers weigh in". *Language acquisition* 21(4): 411–29.

Lenth, Russell V. (2023) "Emmeans: Estimated marginal means, aka least-squares means" [R package version 1.8.6]. Available at: https://CRAN.R-project.org/package=emmeans. Last accessed 31 October 2024.

Lobanova, Anna. (2011) "The role of prominence scales for the disambiguation of grammatical functions in Russian". *Russian linguistics* 35(1): 125–42.

Luchkina, Tatiana, and Jennifer S. Cole. (2016) "Structural and referent-based effects on prosodic expression in Russian". *Phonetica* 73(3–4): 279–313.

Méndez, Tania L., Jason Rothman, and Roumyana Slabakova. (2015) "Discourse-sensitive clitic-doubled dislocations in heritage Spanish". *Lingua* 155: 85–97.

Montrul, Silvina. (2015) *The acquisition of heritage languages*. Cambridge, UK: Cambridge University Press. DOI 10.1017/CBO9781139030502.

O'Grady, William, Hye-Young Kwak, On-Soon Lee, and Miseon Lee. (2011) "An emergentist perspective on heritage language acquisition". *Studies in second language acquisition* 33(2): 223–45. Available at: http://www.jstor.org/stable/44486002.

Pintzuk, Susan, and Ann Taylor. (2006) "The loss of OV order in the history of English". Ans van Kemenade and Bettelou Los, eds. *The handbook of the history of English*. Carlton, Australia: Blackwell Publishing, 249–78.

Polinsky, Maria. (2006) "Incomplete acquisition: American Russian". *Journal of slavic linguistics* 14(2): 191–262.

————. (2008a) "Gender under incomplete acquisition: Heritage speakers' knowledge of noun categorization". *Heritage language journal* 6(1): 40–71.

————. (2008b) "Without aspect". G. G. Corbett and M. Noonan, eds. *Case and grammatical relations*. Amsterdam: John Benjamins, 263–82. Available at: https://www.benjamins.com/catalog/tsl.81.

————. (2011) "Reanalysis in adult heritage language: A case for attrition". *Studies in second language acquisition* 33: 305–28.

————. (2015) "Heritage languages and their speakers: State of the field, challenges, perspectives for future work, and methodologies". *Zeitschrift für Fremdsprachwissenschaft* 26: 7–27.

————. (2018) *Heritage languages and their speakers*. Cambridge, UK: Cambridge University Press. DOI 10.1017/9781107252349.

R Core Team. (2021) "R: A language and environment for statistical computing". Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/. Last accessed 31 October 2024.

Rakhilina, Ekaterina V., Anastasia Vyrenkova, Elmira Mustakimova, Alina Ladygina, and Ivan Smirnov. (2016) "Building a learner corpus for Russian". Elena Volodina, Gintarė Grigonytė, Ildikó Pilán, Kristina Nilsson Björkenstam, and Lars Borin, eds. *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition*. Umeå, Sweden: LiU Electronic Press, 66–75.

Romanova, Natalia. (2008) "Mechanisms of verbal morphology processing in heritage speakers of Russian". *Heritage language journal* 6(1): 105–26.

Rothman, Jason. (2009) "Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages". *International journal of bilingualism* 13(2): 155–63. DOI 10.1177/1367006909339814.

Slabakova, Roumyana, Paula Kempchinsky, and Jason Rothman. (2012) "Clitic-doubled left dislocation and focus fronting in L2 Spanish: A case of successful acquisition at the syntax-discourse interface". *Second language research* 28(3): 319–43.

Slioussar, Natalia. (2007) *Grammar and information structure. A study with reference to Russian*. Utrecht: LOT.

————. (2011) "Processing of a free word order language: The role of syntax and context". *Journal of psycholinguistic research* 40(4): 291–306.

Sorace, Antonella. (2011) "Pinning down the concept of 'interface' in bilingualism". *Linguistic approaches to bilingualism* 1(1): 1–33.

Statistisches Bundesamt. (2019a) "Bevölkerung in Privathaushalten nach Migrationshintergrund im weiteren Sinn nach Geburtsstaat in Staatengruppen. Ergebnisse des Mikrozensus". Available at: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Tabellen/migrationshintergrund-staatsangehoerigkeit-staatengruppen.html. Last accessed 23 April 2020.

————. (2019b) "Pressemitteilung Nr. 314 vom 21. August 2019: Jede vierte Person in Deutschland hatte 2018 einen Migrationshintergrund". Available at: https://www.destatis.de/DE/Presse/Pressemitteilungen/2019/08/PD19_314_12511.html. Last accessed 23 April 2020.

Švedova, Natal'ja Jul'evna, ed. (1980, 2005) *Russkaja grammatika*. Moscow: Institut russkogo jazyka im. V.V. Vinogradova RAN.

Topaj, Nathalie. (2018) *Grammatical and pragmatic use of referential expressions in picture-based narratives of bilingual and monolingual children in Russian and German*. PhD dissertation, Humboldt-Universität zu Berlin.

United Nations. (2019) "International migrant stock 2019: Department of Economic and Social Affairs, Population Division". Available at: www.unmigration.org. Last accessed 29 April 2020.

Uriagereka, Juan. (2007) "Clarifying the notion 'parameter'". *Biolinguistics* 1: 99–113.

US Census Bureau. (2015) "Detailed languages spoken at home and ability to speak English for the population 5 years and over: 2009–2013". Available at: https://www.census.gov/data/tables/2013/demo/2009-2013-lang-tables.html. Last accessed 1 July 2021.

Valdés, Guadalupe. (2005) "Bilingualism, heritage language learners, and SLA research: Opportunities lost or seized?" *The modern language journal* 89(3): 410–26. DOI 10.1111/j.1540-4781.2005.00314.x.

Wegener, Heide. (1993) "Weil – das hat schon seinen Grund. Zur Verbstellung in Kausalsätzen mit weil im gegenwärtigen Deutsch". *Deutsche sprache* 21: 289–305.

Wiese, Heike. (2020) "Language situations: A method for capturing variation within speakers' repertoires". Yoshiyuki Asahi, ed. *Methods in dialectology*. Volume 15. Bamberg: Bamberg Studies in English Linguistics.

Winter, Bodo. (2011) "Pseudoreplication in phonetic research". *ICPhS XVII*: 2137–40.

Zuban, Yulia, Maria Martynova, Sabine Zerbian, Luka Szucsich, and Natalia Gagarina. (2021) "Word order in heritage Russian: Clause type and majority language matter". *Russian linguistics* 45(3): 253–81.

Appendix

Table 9. Model summary

Predictors	Estimates	SE	z value	*p*
(Intercept)	–0.55024	0.28320	–1.943	5.202450e–02
Germany	0.44474	0.38519	1.155	2.482609e–01
USA	–0.71568	0.42193	–1.696	8.984753e–02
Nominal Objects	–2.21013	0.34942	–6.325	2.529034e–10
Emb. Clauses	–1.04441	0.40618	–2.571	1.013147e–02
Spoken Mode	0.29567	0.17171	1.722	8.507501e–02
Formal Setting	–0.03266	0.17127	–0.191	8.487528e–01
Germany : Nominal Objects	–1.43580	0.49541	–2.898	3.752886e–03
USA : Nominal Objects	0.23169	0.47586	0.487	6.263370e–01
Germany : Emb. Clauses	0.97298	0.51926	1.874	6.096140e–02
USA : Emb. Clauses	–0.32743	0.62513	–0.524	6.004306e–01
Observations	1010			
Marginal R^2 / Conditional R^2	0.317 / 0.371			

Table 10. Confidence intervals

Predictors	2.5%	97.5%
(Intercept)	–1.10531185	0.004823809
Germany	–0.31022657	1.199701204
USA	–1.54265041	0.111289172
Nominal Objects	–2.89497793	–1.525288231
Emb. Clauses	–1.84051030	–0.248316023
Spoken Mode	–0.04086513	0.632211324
Formal Setting	–0.36835491	0.303027532
Germany : Nominal Objects	–2.40677639	–0.464817764
USA : Nominal Objects	–0.70098285	1.164368365
Germany : Emb. Clauses	–0.04475457	1.990717871
USA : Emb. Clauses	–1.55267288	0.897806089

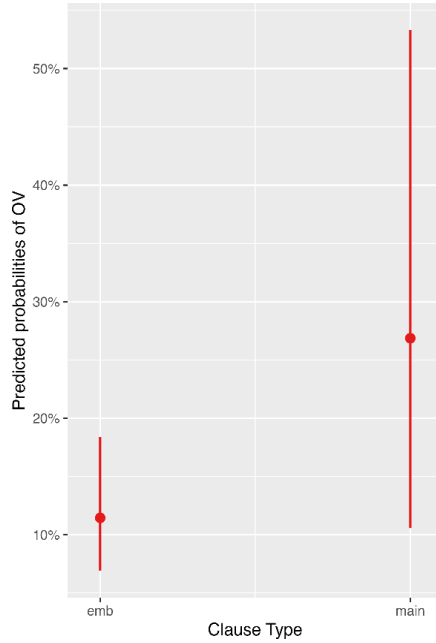


Figure 2. Likelihood of an OV word order in main and embedded clauses

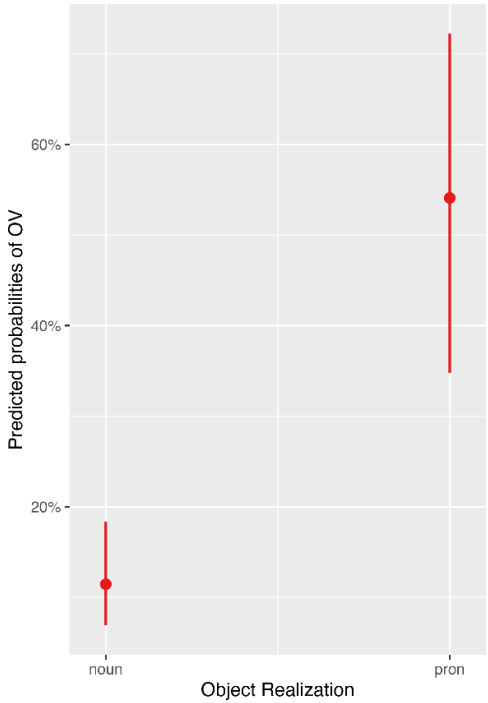


Figure 3. Likelihood of an OV word order with nominal and pronominal objects

Maria Martynova
Humboldt University Berlin
Institute of Slavic Studies
and Hungarology
Berlin, Germany
maria.martynova@gmx.com

Yulia Zuban
University of Stuttgart
Institute of Linguistics
Division of English Linguistics
Stuttgart, Germany
zubanyulia@gmail.com

Luka Szucsich
Humboldt University Berlin
Berlin, Germany
luka.szucsich@hu-berlin.de

Natalia Gagarina
Leibniz Center General Linguistics
Berlin, Germany
gagarina@leibniz-zas.de

The Clausative-Presentative Construction in Russian and Serbian

Igor Mel'čuk and Jasmina Milićević

Abstract: This paper describes the clausative-presentative construction, a particular type of the presentative construction found in Slavic languages, using data from Russian and Serbian. The discussion is carried out within the Meaning-Text linguistic approach.

The clausative-presentative construction is syntactically headed by a special presentative lexeme that is a clausative (i.e., can constitute a clause by itself or together with its obligatory actants): Rus. *èto*/Serb. *to* 'that situation is [Y]' and Rus. *vot*/Serb. *evo* 'I indicate here [Y]'. Such a lexeme takes as its actant—the presentee—a fully independent clause (i.e., a clause without a complementizer), which is, typologically, a rare occurrence.

The semantic and syntactic properties of the clausative-presentative construction follow from the lexical meaning of the clausative-presentative lexeme that heads it. Therefore, this construction is a syntactic construction only in a very general sense (≈ 'a configuration of syntactically linked items'); in fact, it is a clausative lexeme taken with its actant. This construction is compared with the cleft construction, which is a genuine syntactic construction: it expresses communicative information by a particular syntactic configuration.

Lexical entries of Russian and Serbian clausative-presentative lexemes are supplied, together with an overview of their co-polysemous or homophonous lexical "partners".

1. Introduction: The Goal of the Paper and Underlying Notions*

This paper is dedicated to the so-called clausative-presentative construction in Russian and Serbian, illustrated with the following sentences: Rus. *Èto*/*Vot Ivan činit našu mašinu*. / Serb. *To*/*Evo Ivan popravlja naša kola*. ≈ 'This is Ivan repairing our car.'

Since this construction is present in all other Slavic languages, what we say about Russian and Serbian can be extended to the Slavic domain in general. We do not, however, provide data from those other languages, as doing so would make the paper unduly long (and overcomplicated).

* We offer our most heartfelt thanks to Wayles Browne, Lidija Iordanskaja, Artëmij Keidan, François Louis, Sébastien Marengo, and Alain Polguère for their comments upon a prefinal version of this paper. Thanks are also due to two anonymous reviewers and the editors for their constructive suggestions. Any remaining blunders are, of course, our own.

The discussion is couched in the framework of the Meaning-Text linguistic approach (Mel'čuk 1988, 2012–15, vol. 2: 259–376, 2021; Mel'čuk and Milićević 2020) and presupposes a familiarity of the reader with its basic notions, such as dependency, the semantic representation of utterances, dependency trees, deep and surface syntax, and semantic and syntactic actants. The corresponding formalisms are taken for granted. Nevertheless, a characterization of the crucial notions specific to the paper will be provided as we go along.

The goal of our paper is threefold:

1. Propose a rigorous definition of the Slavic clausative-presentative construction [Claus-PresC].
2. Formally describe this construction in Russian and Serbian.
3. Highlight the crucial differences between the Slavic Claus-PresC and the cleft construction, as the two construction types are sometimes equated in the linguistic literature.

A Claus-PresC is a particular case of presentative construction, the latter being a syntactic construction of the following form:

presentative + (copula/copula-like verb +) presentee.

The term "copula/copula-like verb" is used here to cover all BE-verb lexemes meaning 'be identical', 'be an element of a class', etc., including the genuine copula (as in *John* ***is*** *young*)—in order to distinguish them from the locative BE-verb (as in *John* ***is*** *in London*).

The notions of presentative and presentee require special discussion.

1.1. Presentatives

From a semantic viewpoint, a presentative is an expression used by the Speaker to draw the Addressee's attention to a presentee—an entity/a set of entities or a situation to be talked about in the subsequent exchange. A presentative is a kind of linguistic demonstrating or indicative "gesture," an ostensive introduction of a discourse participant: the Speaker is saying "You see here/there…" or "I indicate here/there…". Presentatives thus form a small, semantically characterized lexical class.

From a syntactic viewpoint, presentatives come in two major types: standard presentatives and special presentatives.

A standard presentative is a demonstrative (= deictic) pronoun, which can be nominal or adverbial (like Eng. *it, this, that* or *here, there*). It behaves syntactically as a normal pronoun of this type should, following the usual syntactic patterns of the language.

A special presentative is a demonstrative clausative [DemClaus].[1] DemClaus lexemes exist in all Slavic languages but are absent from most other languages of the Standard Average European type, notable exceptions being, as will be seen, French and Italian.[2]

Special Russian and Serbian presentative lexemes are listed in Table 1; for the linguistic details of Rus. *èto*/*vot* lexemes and Serb. *to*/*evo* lexemes, see §5, pp. 83–96.

Table 1. Clausative-presentative lexemes in Russian and Serbian

	Clausative-presentative lexemes					
	Descriptive lexemes: pronominal clausatives			Signalative lexemes: particle-like clausatives		
Russian	proximal	non-proximal		proximal	non-proximal	
	èto lit. 'this'	arch/poet *to* lit. 'that'		***vot*** lit. 'I indicate here'	*von* lit. 'I indicate there'	
Serbian	proximal	non-proximal	distal	proximal	non-proximal	distal
	ovo lit. 'this'	***to*** lit. 'that'	*ono* lit. 'that there'	***evo*** lit. 'I indicate here'	*eto* lit. 'I indicate there'	*eno* lit. 'I indicate over there'
	≈ 'this here (there, over there) is'			≈ 'I indicate here (there, over there)'		

Table 1 calls for the following three comments.

(i) All Russian and Serbian clausative-presentative [Claus-Pres] lexemes are invariable; pronominal clausatives *èto*/*to* should not be

[1] A clausative (Mel'čuk 2006b: 41), a.k.a. *word-sentence* or Fr. *mot-phrase* (*phrasillon*, in Tesnière 1959: 94–99), is an element of a distinct part of speech: (i) it forms an utterance either on its own (*Yes. / Wow! / Not at all.*) or together with its actant, which can be a complete clause (*Sorry that I cannot be with you*, where the clause *that I cannot be with you* is an actant of *sorry*); (ii) it is not a verb: it does not have properties most typical of verbs in a given language (for instance, in Russian and Serbian, it does not conjugate). A demonstrative is an expression whose meaning contains a semantic component of indication, or pointing out, or else of reference to another expression in the same text; in the present article, it is the semanteme 'this' = 'which is under consideration'. Such a demonstrative is anaphoric or cataphoric.

[2] As F. Louis indicated to us, Breton has a presentative clausative similar to Fr. *voici*: ***Setu*** *mand eo daet Yann en-dro.* lit. 'Look.here that is come Yann in.return.' = 'Here has come back Yann.' Such clausatives probably exist in other Celtic languages.

confounded with the neuter singular forms of the demonstrative adjectival pronouns Rus. *ètot ~ tot* and Serb. *ovaj ~ taj ~ onaj*. Compare the following examples:

(1) a. ['What a strange sound!'] (Russian)
Èto$_{(CLAUSATIVE)}$ plačet kakoj-to rebënok.
'This is some child crying.'
vs.
Èto$_{(ADJ,\ pronominal)NEU.SG.ACC}$ okno$_{(N,\ neu)SG.ACC}$ pora myt'.
'It is time to wash this window.'

b. ['What a strange sound!'] (Serbian)
To$_{(CLAUSATIVE)}$ plače neko dete.
'This is some child crying.'
vs.
To$_{(ADJ,\ pronominal)NEU.SG.NOM}$ dete$_{(N,\ neu)SG.NOM}$ stalno plače.
'This child is always crying.'

(ii) A descriptive lexical item communicates information about the extralinguistic world: *This is **disgusting**! / I find it **curious** that John is absent*.

A non-descriptive lexical item that is a signalative only signals some internal state or an opinion of the Speaker and does this in a specific linguistic form: ***Yuck!*** / ***Curiously***, *John is absent*. A clause that contains a signalative cannot be negated/questioned: **It is not true that yuck! / *Is it true that yuck?*; it cannot be subordinated as a complement—for instance, to a verb of speaking or thinking: **Mary says/knows that yuck* ⟨*that, curiously, John is absent*⟩.[3] For non-descriptive lexical units, see Mel'čuk (2001: 351ff).

(iii) In what follows, reference is only made to the Claus-Pres lexemes that are the unmarked members of their respective paradigms (these lexemes are boldfaced in Table 1). In Russian, *èto* is unmarked with respect to *to*; the former is used across registers, the latter is appropriate only in archaic/poetic ones. In Serbian, the converse is true; it is *to* that is unmarked because it fits more contexts than *ovo/ono*. Only *to* is used to refer to a situation that is not directly perceivable, regardless of the degree of proximity, while the

[3] The angular brackets in this paper indicate alternative variants.

reference to a directly perceived situation is made using any of the three DemClaus lexemes. Therefore, the proximal Rus. *èto* and the non-proximal Serb. *to* are taken as translational equivalents. Their difference in degree of proximity, seen in the glosses, is not relevant for the purposes of the present paper. (The clausatives *vot*/*evo* in the two languages have fewer discrepancies in the degree of proximity.)

1.2. Presentees

From a semantic viewpoint, a presentee can denote either an entity/a set of entities or a situation. From a syntactic viewpoint, presentees are, like presentatives, of two major types.

A standard presentee is a noun phrase (in a broad sense); exactly which subclass of noun phrases or its syntactic equivalents is allowed for a given presentative construction (e.g., can it be a THAT-clause?, a pseudo-relative clause?, etc.) is determined by the particular presentative lexeme and can vary from one language to the next. A special presentee is a full independent clause.

1.3. Standard and Special Presentative Constructions

In accordance with the characteristics of the presentative and the presentee, two types of presentative constructions are distinguished: standard and special.

In a standard presentative construction, both the presentative and the presentee are standard, and the copula is present, for instance:

(2) a. This/That is John's new car.
b. These/Those are my friends.
c. This/That/It is John repairing our car.

The standard presentative construction is probably language universal.

A detailed study of the standard presentative construction in English is found in Wood and Zannuttini (2023). Although the authors touch upon the special presentative construction only tangentially (the French and Italian subtypes, cf. our examples (6) and (7)), some properties of the standard presentative construction they point out and characterize (such as the impossibility of embedding and particular co-occurrence with the negation) are quite similar to those we have established for special presentative constructions (see §2 below).

In a special presentative construction, the presentative and/or the presentee are special or the copula is absent. One particular type of special presentative construction appears in Slavic languages; it does not include a

copula, the presentative is special, and the presentee is special or standard syntactically (depending on the presentative). In these languages, the presentative construction is syntactically headed by a DemClaus lexeme, and therefore it can be naturally called clausative-presentative construction [Claus-PresC]. Here are examples of the Russian and Serbian Claus-PresC.

(3) a. ['What is this noise over there?'] (Russian)

Èto Ivan idët
this.situation.is Ivan.NOM come.PRES.3SG[4]

⟨činit našu mašinu⟩.
repair.PRES.3SG our car.SG.ACC

'This is Ivan coming ⟨repairing our car⟩.'

b. (i) **Vot** Ivan ⟨on⟩.
I.indicate.here Ivan.NOM he.NOM

'Here is Ivan ⟨he⟩.'

(ii) **Vot** Ivan idët
I.indicate.here Ivan.NOM come.PRES.3SG

⟨činit našu mašinu⟩.
repair.PRES.3SG our car.SG.ACC

'Here is Ivan coming ⟨repairing our car⟩.

(4) a. ['What is this noise over there?'] (Serbian)

To Ivan dolazi
that.situation.is Ivan.NOM come.PRES.3SG

⟨popravlja naša kola⟩.
repair.PRES.3SG our car.PL.ACC

'This is Ivan coming ⟨repairing our car⟩.'

b. (i) **Evo** Ivana ⟨ga⟩.
I.indicate.here Ivan.GEN he.CLIT.GEN

'Here is Ivan ⟨he⟩.'

(ii) **Evo** Ivan dolazi
I.indicate.here Ivan.NOM come.PRES.3SG

⟨popravlja naša kola⟩.
repair.PRES.3SG our car.PL.ACC

'Here is Ivan coming ⟨repairing our car⟩.'

[4] A period between two semantic elements in glosses indicates that these elements are physically expressed together—by one lexical unit.

These examples illustrate a typical speech act in which a Claus-PresC is used, i.e., an affirmative-declarative sentence that is an appropriate answer to a question about a given situation, namely, a question such as *What is happening? / What is this? / What is this noise? / Why all this mess?*, etc. In the answer—'This.situation.is [Y]'—the presentative expresses the semantic Theme (= Topic) of the sentence and the presentee expresses its semantic Rheme (= Comment). This is always the case with *èto/to*; but *vot/evo* can also express the semantic Rheme (see §3 below).

Sentences in (3) and (4) contrast with sentences such as Rus. *Èto Ivan* 'This is Ivan', which features a standard presentative construction, just like the English one illustrated in (2). Here, we see a different lexeme *èto*—a nominal demonstrative pronoun, and the sentence contains the copula *byt´* 'be', which has a zero form in the present tense; cf. Serb. *To* ***je*** *Ivan* 'That is Ivan', where the copula is non-zero, and Rus. *Èto* ***byl*** ⟨***budet***⟩ *Ivan* 'This was ⟨will be⟩ Ivan'. However, in the Claus-PresC, no copula intervenes because one of the senses of the verb *byt´/biti* is included within the meaning of the Claus-Pres lexeme (i.e., *èto/to* and *vot/evo*): the meaning 'be.identical'.

To avoid misunderstandings, remember that there is an important difference between a semanteme being present in the meaning of an expression and a lexeme being physically present in that expression. When in a gloss we write, for short, 'this.situation.is' (instead of 'this situation is identical to or a result of', see the lexicographic definition of *èto*^{1}**III**/*to*^{1}**III** on p. 84), we indicate that all the semantemes connected by periods are lexicalized together; none receives a separate lexical expression.

Special presentative constructions include, along with the Slavic Claus-PresC, other types, for instance, the one found in Mandarin Chinese (Li and Thompson 1989: 153); the sentence in example (5) is an appropriate answer to a question like *What's happening?* or *What is this noise?*

(5) Shì Wāng zài nàr dǎ gǔ.
be Wang at there hit drum
'This is Wang playing drums there.'

It is the presentee that makes this Chinese presentative construction special; it is a full-fledged clause without complementizer. A fundamental difference between the special presentative construction in Chinese and that in Slavic is that, in the former, the syntactic head of the construction is the copula verb *shì* 'be', while in the latter, the syntactic head is not verbal—it is a clausative.[5]

[5] With Chinese being a strongly Pro-Drop language, example (5) does not overtly include a subject similar to Eng. *it/this*; however, the subject *zhè* 'this' is quite normal in such sentences, if the context warrants it.

Works that treat Claus-Pres lexemes and the (special) Claus-PresC in Russian and/or Serbian (although under different terms) include Browne (1976); Padučeva (1982); Junghanns (1997); Progovac (1998) and (2005: 54–56, 197–212); Kordić (2002: 93–128); Peti (2005); Reeve (2008) and (2012); Kimmelman (2009); Krylova (2014); Raković (2017); and Burukina and den Dikken (2020). For the Claus-PresC in French, see Morin (1985); Lambrecht (1988) and (2000); and Porhiel (2012), among many others. Grenoble and Riley (1996) compare discourse properties of Fr. *voici/voilà* and Rus. *vot/von*. For the corresponding construction in Italian, see Zanuttini (2017) and Wood and Zannuttini (2023: 596).

However, none of these studies explores the aspects of Claus-PresC we are interested in here, which are the formal definition of the notion, the syntactic structure of the construction in terms of syntactic dependency, and its comparison with the cleft construction. Furthermore, our approach is different from most others in the following three respects.

First, it adopts a synthesis orientation (text production, as opposed to the analysis, or text comprehension approach)—that is, a linguistic description is carried out strictly from Meaning to Text. Second, it is aimed at modeling the corresponding linguistic phenomena by means of sets of rules, understood as formal expressions that can be applied mechanically. Third, our stance is resolutely lexicalist, as we take the properties of the construction under discussion to stem directly from the lexical meaning of the DemClaus lexemes that control it. (A formal lexicographic description of these lexemes is provided in §5.) Under this view, the Claus-PresC turns out to be a syntactic construction only in a very general and rather vague sense of the term ('a configuration of syntactically linked items'), applicable to any syntactic phrase. More on this will be said below.

For all these reasons, it is difficult to compare our analysis with previous ones; we will, however, make reference to the specifics of some anterior proposals, where appropriate.

From now on, we will be using distinctive lexicographic numbers with our special Claus-Pres lexemes: Rus. *èto*^{1}**III**/Serb. *to*^{1}**III** and Rus. *vot*^{1}/Serb. *evo***1**.

Most of the examples are our own; we also have used some found on the Internet and in the linguistic literature on the subject. Symbols and abbreviations are presented in a table on p. 98.

2. The Clausative-Presentative Construction

Unlike a standard presentative construction illustrated in (2), whose syntactic head, i.e., the top node, is a verb (more precisely, the verb 'be'), a special presentative construction is headed by a clausative. (For the considerations used to determine the orientation of the syntactic dependency within the Claus-PresC, see below.)

definition 1: the clausative-presentative construction

A Claus-PresC is a deep-syntactic [DSynt-] subtree of the following form:

$$L_{1(\text{CLAUSATIVE, presentative})} \text{–}\textbf{II}\text{→} L_2,$$

where:

L_1 is a Claus-Pres lexeme;

L_2 is either a noun/a syntactic equivalent of a noun or a finite verb that is the syntactic head of a clause $C_{(\text{presentee})}$.

In the surface-syntactic structure, the head of $C_{(\text{presentee})}$ is either its main verb (i.e., L_2) or a semantically empty subordinating conjunction, i.e., a complementizer.

In prose, this means that a clausative-presentative construction is in fact simply a clausative-presentative lexeme taken with its deep-syntactic actant **II**.

The presentee component in the Claus-PresC is the second actant of the Claus-Pres lexeme—both on the semantic and the deep-syntactic levels of linguistic representation. This is explained by the meaning of such a lexeme: its first actant should be the situation in question (for *èto/to* [Y], which means 'this situation is identical to Y or a result of Y', §5.1) or the Speaker (for *vot/evo* [Y]: 'I indicate here something that is Y', §5.3), but this semantic component is included ("incorporated," as it were) in the Claus-Pres lexeme's meaning.[6]

Outside the Slavic family, Claus-PresCs are found among Standard Average European languages, for example as mentioned above, in French and Italian. The lexemes Fr. *voici/voilà* 'one.sees.here/there' and It. *ecco* 'look.here' are Claus-Pres lexemes heading a presentative construction similar, in some respects, to the Slavic one.[7]

[6] Clausatives lacking DSynt-actant **I**, while having DSynt-actant **II**, are widespread in natural languages; here are some English examples: *Down*–**II**→*with the mullahs!*, *Kudos*–**II**→*to everyone!*, *Off/Out*–**II**→*with you!* But not only clausatives can have such an actantial frame: there are meteorological verbs like (to) *rain* and (to) *dawn*, verbs of physiological state of the Rus. *tošnit´* 'nauseate' type (*Menja tošnit*. lit. '[It] nauseates me.'), various modal expressions, etc. For more, see Mel'čuk (2015: 67–68).

[7] The French and Italian presentative lexemes are verb-like in that they accept a regular direct object; It. *ecco*, in particular, is similar to an imperative verb form because a clitic object follows it. (This explains different glosses we adopted for it and its French counterparts, with which the clitics precede, as is normal for non-imperative sentences in Romance.) However, these lexemes are not verbs (not even defective ones), and this is so because they lack one indispensable verbhood attribute—the capacity to control a syntactic subject. For a different viewpoint, see Morin (1985); see also Porhiel (2012).

(6) a. **Voici** Jean. ~ Le **voici**. (French)
one.sees.here Jean he.ACC one.sees.here
'Here is Jean.' 'Here he is.'

b. **Voici** Jean ~ Le **voici**
one.sees.here Jean he.ACC one.sees.here
qui arrive. qui arrive.
who come.PRES.3SG who come.PRES.3SG
'Here is Jean who is coming.' 'Here is he who is coming.'

c. **Voici** que Jean arrive.
one.sees.here that Jean come.PRES.3SG
'Here comes Jean.'

d. **Voici** arriver Jean.[8]
one.sees.here come.INF Jean
'Here comes Jean.'

(7) a. **Ecco** Giovanni. ~ **Ecco**lo. (Italian)
look.here Giovanni look.here=he.ACC
'Here is Giovanni.' 'Here he is.'

b. **Ecco** Giovanni/lo che arriva.
look.here Giovanni he.ACC who come.PRES.3SG
'Here is Giovanni ⟨*he*⟩ who is coming.'

c. **Ecco** che arriva Giovanni
look.here that come.PRES.3SG Giovanni
'Here comes Giovanni.'

d. **Ecco** arrivare Giovanni.
look.here come.INF Giovanni
'Here comes Giovanni.'

e. **Ecco** arriva Giovanni.
look.here come.PRES.3SG Giovanni
'Here comes Giovanni.'

The Claus-PresC in French does not admit a finite verb as the syntactic head of the $C_{(\text{presentee})}$ clause; in Italian it accepts in this role only a few "ap-

[8] This construction is stylistically marked.

pearance" verbs ('appear', 'arrive', 'come', etc.), as shown in (7e). However, the corresponding Slavic construction admits any verb—including a transitive one (together with its object)—as the head of $C_{(presentee)}$. To put it differently, in the Slavic Claus-PresC, the presentative lexeme governs as its $C_{(presentee)}$ any full clause without a complementizer. Examples such as (6c) and (7c), in which $C_{(presentee)}$ contains a complementizer, are ungrammatical in Russian and Serbian.

An important question naturally arises: How does one establish the orientation of Sem- and Synt-dependencies between a Claus-Pres lexeme and the head of the $C_{(presentee)}$ clause? That is, how can the researcher decide whether the head of $C_{(presentee)}$ depends on the presentative lexeme or, conversely, the presentative lexeme depends on the head of $C_{(presentee)}$? Let us first consider the situation with *èto*^{1}**III** (and its Serbian counterpart *to*^{1}**III**).

The semantic dependency between the corresponding semantemes, i.e., between 'èto^{1}**III**' and the meaning underlying the head of $C_{(presentee)}$, is given by the decomposition of the former, which, as already mentioned, is 'this situation is identical to Y or a result of Y'. We therefore conclude that 'èto^{1}**III**' is the semantic governor and the meaning underlying the head of $C_{(presentee)}$ is written as follows: 'èto^{1}**III**$_{-2}\rightarrow C_{(presentee)}$'.

As far as the syntactic dependency between *èto*^{1}**III** and the head of $C_{(presentee)}$ is concerned, things are more complicated. The established criteria for determining the orientation of the syntactic dependency between two lexemes L and L′ are formulated based on the syntactic and morphological behavior of the pair L—L′ **within** a sentence (for instance, Mel'čuk 2021: 41ff.). But in the case of Claus-PresC, these intra-sentential criteria do not allow for determining the syntactic status of *èto*^{1}**III** with respect to the $C_{(presentee)}$ clause because the latter is a full-fledged sentence in its own right. W. Browne aptly called the Serbian *to*^{1}**III** a "sentence introducer".[9] Thus, in our case, the only considerations available are semantic ones.

First, the meaning 'èto^{1}**III**' is a predicate with one Sem-actantial slot—'Y', and it is reasonable to consider the clause whose meaning fills in this slot to be not only a semantic but also a syntactic actant of *èto*^{1}**III**. Second, *èto*^{1}**III** imposes specific semantic constraints on the $C_{(presentee)}$ clause, to wit, $C_{(presentee)}$ must be declarative and its main verb must denote a dynamic fact (Progovac 1998 explicitly identifies Serb. *to*^{1}**III** as "event pronominal"); and imposing constraints on syntactically linked expressions is more typical of syntactic governors.

Therefore, at the deep-syntactic level, *èto*^{1}**III** governs the head of the $C_{(presentee)}$ by the DSyntRel **II** (cf. Definition 1, p. 69). Since the meaning 'èto^{1}**III**' includes one of the copular meanings—'be.identical'—the syntactic

[9] "The word *to* in this special use exhibits the behavior of a sentence introducer" (Browne 1976: 199).

dependency between *èto*^{1}**III** and $C_{(\text{presentee})}$ is encoded at the surface-syntactic level by the **copular-completive** SSyntRel, foreseen for the copula:

èto^{1}**III**–**copular-completive**→$C_{(\text{presentee})}$.

The situation with syntactic dependency is more obvious in the case of *vot*^{1} and its Serbian counterpart, *evo***1**. Here we have the following facts that argue for *vot*^{1}/*evo***1** as the governor.

First, *vot*^{1}/*evo***1** determines the syntactic behavior of its partner. A noun without *vot*^{1}/*evo***1** does not form a genuine independent clause (that is, a phrase that can, for instance, be coordinated with another clause), but it does so with *vot*^{1}/*evo***1**. Here is an example featuring *vot*^{1} (*evo***1** behaves identically).

(8) a. *Ivan, i teper′ my možem idti.
Intended: '[Here's] Ivan, and now we can go.'
vs.
Vot Ivan, i teper′ my možem idti.
'Here's Ivan, and now we can go.'

b. *My dolgo ždali, i, nakonec, Ivan.
Intended: 'We have been waiting for a long time and finally Ivan [is here].'
vs.
My dolgo ždali, i **vot**, nakonec, Ivan.
'We have been waiting for a long time and here, finally, is Ivan.'

Second, another weighty argument in favor of the governor status of Claus-Pres lexemes within our construction is provided by Serb. *evo***1**. It controls the form of the nominal presentee, which takes the genitive case (cf. Criterion B2 for the direction of a syntactic dependency in Mel'čuk (2021: 42), that is, the *control of inflection* or *morphological contact point* criterion); this is shown in (4b(i)) and in the following example (for more on the government of *evo***1**, see §5.3).

(9) A evo nam i **Marije** ⟨**nje**⟩.
and I.indicate.here we.DAT.ETH also Marija.GEN she.FULL.SG.GEN
'And here is our Marija ⟨she⟩ as well.'

Third, there are typological considerations. Thus, the direction of syntactic dependency is fairly obvious in the case of Fr. *voilà*/*voici* and It. *ecco*, as,

generally speaking, is the case with all clausatives of this kind, which are the syntactic governors in the corresponding constructions.

Thus, clausative-presentative lexemes turn out to be syntactic governors.[10] The resulting construction is monoclausal: it is a clausative with this presentee clause as its actant; there is only **one** subject—main verb pair. (This apparently paradoxical situation is brought about by the special nature of clausatives, which can be functionally equivalent to clauses but, in point of fact, are not clauses.)

3. The Clausative-Presentative Construction in Russian and Serbian

As illustrated in (3a) and (4a) as well as in (3b) and (4b), there are two subtypes of the presentative construction in both languages, with *èto*^{1}**III**/*to*^{1}**III** and with *vot*^{1}/ *evo***1**. We will take them in turn.

3.1. The Clausative-Presentative Construction Headed by Rus. *èto*^{1}III/ Serb. *to*^{1}III

The clausative Rus. *èto*^{1}**III**/Serb. *to*^{1}**III** takes the $C_{(\text{presentee})}$ clause as its actant **II** at the DSynt-level and as a copular complement at the SSynt-level:

(10) [What is this mess?]

a. Èto Ivan moet ⟨myl,
this.situation.is Ivan.NOM wash.PRES.3SG wash.PAST.MASC.3SG

budet myt′⟩ posudu.
will.be.AUX.FUT.3SG wash.INF tableware.SG.ACC

'This is/was/will be Ivan washing dishes.'

b. (i) To Ivan pere sudove.
that.situation.is Ivan.NOM wash.PRES.3SG dish.PL.ACC

'This is Ivan washing dishes.'

(ii) To je Ivan prao
this.situation.is was.AUX.PAST.3SG Ivan.NOM wash.P.PART

sudove.
dish.PL.ACC

'This was Ivan washing dishes.'

[10] The SSynt-dependency status of *èto*^{1}**III** had previously been considered as being inverse, that is, *èto*^{1}**III** was taken to be a SSynt-dependent of $C_{(\text{presentee})}$; see Iomdin (2010) and Mel'čuk (2021: 75, no. 44).

(10) b. (iii) To ću Ivan prati
this.situation.is will.be.AUX.FUT.3SG Ivan.NOM wash.INF
sudove.
dish.PL.ACC
'This will be Ivan washing dishes.'

The main verb of the presentee clause can be in any tense, but due to the semantics of the Claus-Pres lexemes, it is most natural in the present tense.

Moreover, in Serbian, where the past and the future tense forms use a clitic auxiliary, the clitic's adjacency to *to*[1]**III** may cause processing difficulties for the Addressee, leading to an erroneous interpretation of *to*[1]**III** as the "ordinary" demonstrative pronoun; therefore, the past and the future tense of the main verb of the Serbian presentee clause tend to be avoided.[11]

Since *èto*[1]**III**/*to*[1]**III** are descriptive lexemes, the construction they are heading, i.e., the Claus-PresC, can be used not only in affirmative sentences, as illustrated in (3a–4a) and in (10), but also in negative sentences (see (11) with the underlying question 'Is this the rain falling?'), and in general questions (see (12)). A sentence including this presentative can be subordinated as an object clause—for instance, to a verb of speaking or thinking (see (13)).

(11) a. Èto ne idët dožd' ⟨ ≡ Èto ne dožd' idët⟩ (, èto prosto Ivan prinimaet duš).[12]
'This is not the rain falling(, this is simply Ivan taking a shower).'

b. To ne pada kiša ⟨ ≡ Ne pada to kiša⟩ (, to se samo Ivan tuširа).
'This is not the rain falling (, this is simply Ivan taking a shower).'

In (11), it is not the existence of the situation Y that is negated but rather Y's identity; to put it differently, what is negated is the link 'is' between *èto*[1]**III**/*to*[1]**III** and the presentee. Therefore, it is natural, and in Serbian highly desirable, for the Speaker to offer an explanation about the context of situation Y (provided in parentheses in our examples). The presentative lexeme is often

[11] The form *to* in the sequence *To je Ivan …* can be interpreted in two different ways, depending on the continuation: either as the "ordinary" demonstrative pronoun *to* in the role of a fronted direct object, cf. (i), or as the clausative-demonstrative *to*[1]**III** heading the ClausPresC, cf. (ii).

(i) To$_{(Pron.Dem),\ DirO}$←**dir-objectival**–[*je Ivan*]–*uradio*.
lit. 'That Ivan did.' = 'That was done by Ivan.'

(ii) To$_{(Claus)}$–**copular-attibutive**→*je Ivan došao*.
lit. 'This.situation.is Ivan came.'

[12] The symbol "≡" indicates semantic equivalence.

repeated in the explanation (*Èto/To ne…*, ***èto/to***…). On the particularities of the negation used in sentences with *èto*[1]**III**/*to*[1]**III**, see §5.1, p. 85.

(12) a. Èto idët Ivan ⟨ ≡ Ivan idët⟩?
'Is this Ivan coming?'

b. Jel' to Ivan dolazi ⟨ ≡ dolazi Ivan⟩?
'Is this Ivan coming?'

(13) a. Maša govorit ⟨znaet⟩,→čto→èto→idët→Ivan ⟨ ≡ Ivan←idët⟩.
'Masha says ⟨knows⟩ that this is Ivan coming.'

b. ?Marija kaže ⟨zna⟩ da→to→dolazi→Ivan ⟨ ≡ Ivan←dolazi⟩.
'Marija says ⟨knows⟩ that this is Ivan coming.'

The subordination of the Claus-PresC is less felicitous in Serbian than in Russian.

As for the communicative role of *èto*[1]**III**/*to*[1]**III**, we have already indicated that it necessarily expresses the semantic [Sem-] Theme of the sentence. Its actant, the $C_{(\text{presentee})}$ clause, expresses the Sem-Rheme and can feature further communicative subdivisions, i.e., a secondary Rheme and a secondary Theme. Thus, from the semantic representation [SemR] in Figure 1, it is possible to produce both sentences in (14)—propositionally synonymous but with different secondary thematizations (not shown in Figure 1).

(14) [What is happening?]

a. [Èto]$_{\text{Theme-1}}$ [[Ivan]$_{\text{Theme-2}}$ [idët]$_{\text{Rheme-2}}$]$_{\text{Rheme-1}}$ ≡

b. [Èto]$_{\text{Theme-1}}$ [[idët]$_{\text{Theme-2}}$ [Ivan]$_{\text{Rheme-2}}$]$_{\text{Rheme-1}}$
'This is Ivan coming.'

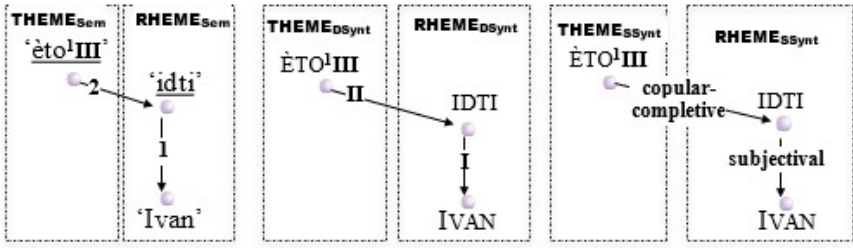


Figure 1. Partial representations of example (14)
at the semantic, deep-syntactic, and surface-syntactic levels

A word of caution is in order here. Both sentences in (14), illustrating the Claus-PresC, are uttered with a specific prosody—namely, with a possible phrasal stress on *èto*, a possible pause after it, and without phrasal stress on a wordform in the $C_{(presentee)}$ clause. If *èto* is unstressed and there can be no pause between it and the following wordform that carries phrasal stress, this is a different *èto*—a focalizing particle *èto*[3], and the corresponding construction is completely different (see example (40), §5.2).

Quite naturally, the $C_{(presentee)}$ clause can feature focalization of the secondary Rheme by means of prosody and/or word order, as in (15); **BOLDFACE SMALL CAPS** and the acute symbol indicate prosodic emphasis—rising intonation and phrasal stress:[13]

(15) a. [Èto]$_{Theme\text{-}1}$ [[moet posudu]$_{Theme\text{-}2}$ [**IVÁN**]$_{Rheme\text{-}2,\ Focalized}$]$_{Rheme\text{-}1}$
this.situation.is washes dishes Ivan

b. [Èto]$_{Theme\text{-}1}$ [[Ivan moet]$_{Theme\text{-}2}$ [**POSÚDU**]$_{Rheme\text{-}2,\ Focalized}$]$_{Rheme\text{-}1}$
this.situation.is Ivan washes dishes

Secondary Theme ~ Rheme division within the $C_{(presentee)}$ clause is possible in sentences with Serb. *to*[1]**III** as well, even though the resulting word order permutations are less free than in Russian and the SVO order is strongly preferred.

In addition, in Serbian, unlike in Russian, if the $C_{(presentee)}$ clause's subject that expresses the secondary semantic Theme is Focalized, it can be extracted from the $C_{(presentee)}$ clause and linearly placed at the very beginning of the sentence, thus depriving *to*[1]**III** of its first position (in this connection, see Progovac 2005: 55, which attributes the dislocation of the subject of the $C_{(presentee)}$ clause to its being communicatively given and thus outside of the "scope" of *to*[1]**III**).

(16) [What is this noise?]

a. [**To**]$_{Theme\text{-}1}$ [[Ivan]$_{Theme\text{-}2}$ [pere sudove]$_{Rheme\text{-}2}$]$_{Rheme\text{-}1}$
that.situation.is Ivan washing dishes
'That is Ivan washing dishes.'

b. [Ivan]$_{Theme\text{-}2,\ Focalized}$ [**to**]$_{Theme\text{-}1}$ [[pere sudove]$_{Rheme\text{-}2}$]$_{Rheme\text{-}1}$
Ivan that.situation.is washes dishes
'Ivan, that is him washing dishes.'

[13] Incidentally, the same state of affairs obtains in the Chinese special presentative construction, i.e., the focalization of the secondary Rheme is marked by a phrasal stress on one of its wordforms (boldfaced in the example below), very much like what happens in Russian and Serbian, as can be seen in (i), based on example (5) above:

(i) a. Shì **Wāng** zài nàr dǎ gǔ. 'This is **WÁNG** is playing drums there.'
b. Shì Wāng **zài nàr** dǎ gǔ. 'This is **THÉRE** Wang is playing drums.'
c. Shì Wāng zài nàr dǎ **gǔ.** 'This is **DRÚMS** Wang is playing there.'

Rus. *èto*[1]**III** can occupy a non-initial position in the sentence in highly colloquial style.

(17) Ivan tam èto činit mašinu!
Ivan there this.situation.is repairs car
'It is Ivan who is repairing the car over there.'

In neutral style, *èto*[1]**III** can be preceded only by one of the modal particles mentioned in §5.1 (see example (24), p. 84).

3.2. The Clausative-Presentative Construction Headed by Rus. *vot*[1]/ Serb. *evo***1**

In this construction, the dependent element—the actant of *vot*[1]/*evo***1**—can be either a noun, as in (3b(i)–4b(i)), or an independent clause, as in (3b(ii)– 4b(ii)).

Since *vot*[1]/*evo***1** are non-descriptive lexemes, the sentence they are heading does not admit negation or interrogation and cannot be syntactically subordinated.

(18) a. *Vot ne idët Ivan. / *Evo ne dolazi Ivan.
I.indicate.here not comes Ivan
Intended: 'I indicate here that it is not the case that Ivan is coming.'

b. *Vot idët Ivan? / *Evo dolazi Ivan?
I.indicate.here comes Ivan
Intended: 'Do I indicate here that Ivan is coming?'

c. ??Maša govorit ⟨znaet⟩, čto vot idët Ivan. /
Masha says ⟨knows⟩ that I.indicate.here comes Ivan
*Marija kaže ⟨zna⟩ da evo dolazi Ivan.
Intended: 'Masha says/knows that I indicate here that Ivan is coming.'

Because *vot*[1]/*evo***1** are, so to speak, linguistic indicative gestures used in "here and now" situations, the main verb of the $C_{(presentee)}$ clause can only be in the present tense.

Unlike *èto*[1]**III**/*to*[1]**III**, which express exclusively the Sem-Theme of the corresponding sentence, the clausatives *vot*[1]/*evo***1** appear in both communicative roles—as the Sem-Theme and as the Sem-Rheme, the latter being somewhat less natural for Serbian *evo***1** than for its Russian counterpart. Let us first illustrate the case of *vot*[1].

Lexeme *vot*[1] can express the Sem-Theme of the sentence, while its actant, the presentee (a noun or a clause), expresses the Sem-Rheme—with further communicative divisions being possible in case the presentee is a clause (see (19a) and (19b)).

Lexeme *vot*[1] can also function as the Sem-Rheme—provided the presentee is expressed by a nominal, as in (19c) and (19d). In (19d), the nominal presentee is communicatively Given (i.e., its referent is present in the mind of interlocutors). This communicative marking triggers the "split" implementation of the Sem-Theme: by the noun IVAN—in the SSynt-role of postlepsis or prolepsis—and the substitute pronoun *on* 'he', co-referential with the noun. (The arrows above the examples indicate, quite roughly, the appropriate intonation. Example (19b) does not carry any prosodic indication because the corresponding sentence is uttered with neutral declarative prosody.)

(19) a. [Vot]$_{\text{THEME}}$ [Ivan.]$_{\text{RHEME}}$
I.indicate.here Ivan.NOM
'Here is Ivan.'

b. [Vot]$_{\text{THEME}}$ [Ivan činit našu mašinu.]$_{\text{RHEME}}$
I.indicate.here Ivan.NOM repair.PRES.3SG our car.SG.ACC
'Here is Ivan repairing our car.'

c. [Ivan]$_{\text{THEME}}$ [– vot.]$_{\text{RHEME}}$ ~ [**VÓT**]$_{\text{RHEME}}$ Ivan.]$_{\text{THEME}}$
Ivan.NOM I.indicate.here I.indicate.here Ivan.NOM
'Ivan is here.' 'Here is Ivan.'

d. [Vot on]$_{\text{RHEME}}$ [Ivan.]$_{\text{THEME, Given}}$
I.indicate.here he.NOM Ivan.NOM
'Here he is, Ivan.'

~ [Ivan]$_{\text{THEME, Given}}$ [**VÓT** on.]$_{\text{RHEME}}$
Ivan.NOM I.indicate.here he.NOM
'Ivan, here he is.'

As for *evo***1**, it is naturally used in the role of the Sem-Theme, similarly to *vot*[1] in (19a); it can also be used like *vot*[1] in (19b), but only if the word order in the presentee clause is unmarked (Subject-Verb-Object).

(20) a. (i) [Evo]$_{\text{THEME}}$ [Ivana.]$_{\text{RHEME}}$
I.indicate.here Ivan.GEN
'Here is Ivan.'

(ii) [Evo ga]$_{\text{THEME}}$ [Ivan.]$_{\text{RHEME, Given}}$
I.indicate.here he.CLIT.GEN Ivan.NOM
'Here is Ivan.'

b. [Evo]$_{\text{THEME}}$ [Ivan popravlja naša kola.]$_{\text{RHEME}}$
I.indicate.here Ivan.NOM repair.PRES.3SG our car.PL.ACC
'Here is Ivan repairing our car.'

However, Serbian examples (21a–b), analogous to Russian (19c–d), where *evo***1** expresses the semantic Rheme, are highly marked and found "bizarre" by most speakers.

(21) a. (i) ?[Ivan]$_{\text{THEME}}$ [– evo ga ⟨njega⟩.]$_{\text{RHEME}}$
Ivan I.indicate.here he.CLIT.GEN he.FULL.GEN
'Ivan, here he is.'

(ii) ?[Ivan]$_{\text{THEME}}$ [– evo ga on.]$_{\text{RHEME}}$
Ivan I.indicate.here he.CLIT.GEN he.NOM
'Ivan, here he is.'

b. ?[Evo ga on,]$_{\text{RHEME}}$ [Ivan.]$_{\text{THEME}}$
I.indicate.here he.CLIT.GEN he.NOM Ivan.NOM
'Here he is, Ivan.'

The representations of Russian sentences in (19) at the three relevant levels—semantic, deep-syntactic, and surface-syntactic—are featured in Figure 2 through Figure 5. The representations of the Serbian counterparts of (19), cited in example (20), can be seen in Figure 6 through Figure 8. (These representations are incomplete because irrelevant details—e.g., grammemes—are omitted.)

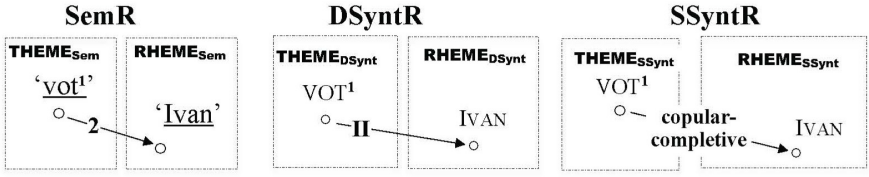


Figure 2. Representations of example (19a): *Vot Ivan*. 'Here is Ivan.'

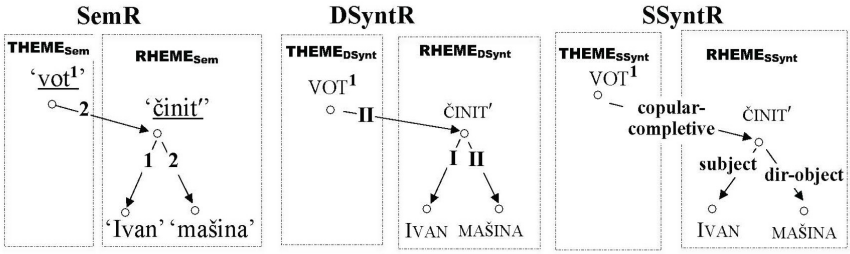


Figure 3. Representations of example (19b):
Vot Ivan činit našu mašinu. 'Here is Ivan repairing our car.'

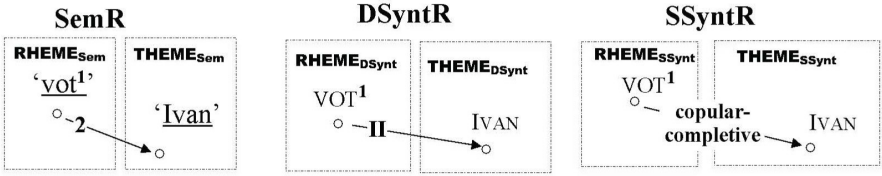


Figure 4. Representations of example (19c):
Ivan – vot. 'Ivan is here.' ~ ***vót*** *Ivan.* 'Here is Ivan.'

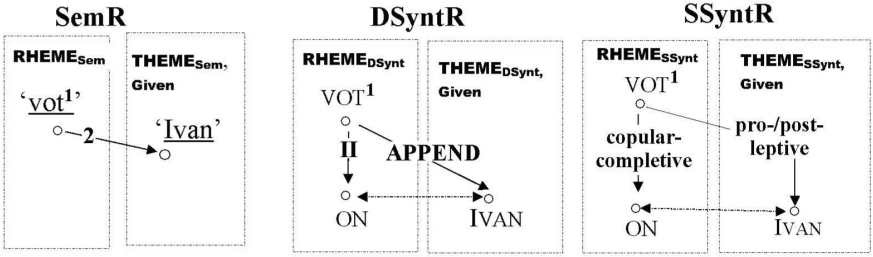


Figure 5. Representations of example (19d):[14]
Vot on, Ivan. 'Here he is, Ivan.' ~ *Ivan,* ***vót*** *on.* 'Ivan, here he is.'

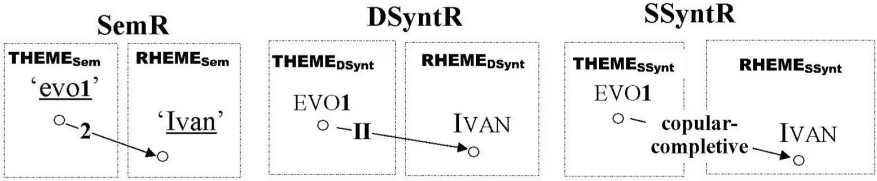


Figure 6. Representations of example (20a(i)):
Evo Ivana. 'Here is Ivan.'

[14] The double-headed dashed arrow shows a coreference link.

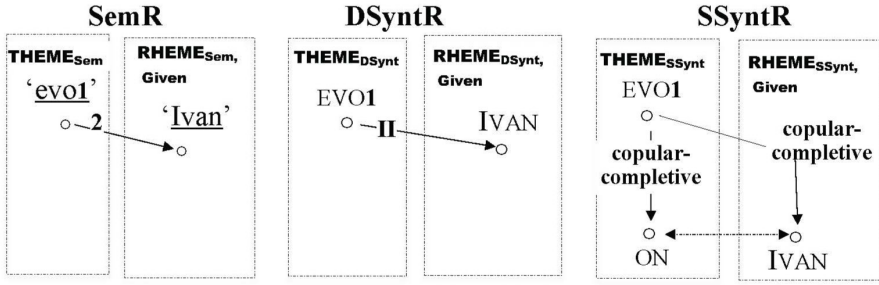


Figure 7. Representations of example (20a(ii)):
Evo ga Ivan. 'Here is Ivan.'

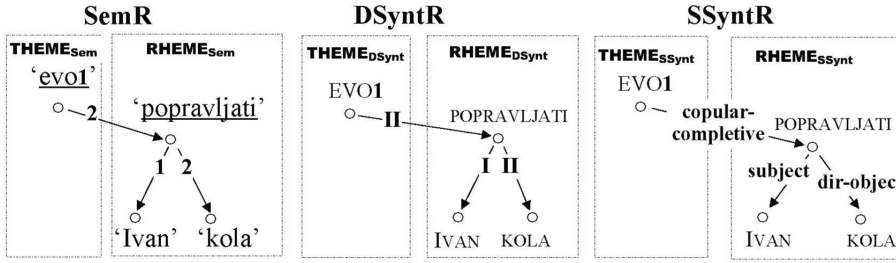


Figure 8. Representations of example (20b):
Evo Ivan popravlja naša kola. 'Here is Ivan repairing our car.'

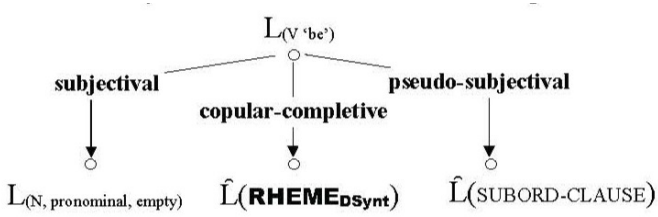
Sentence (20a(ii)) contains the resumptive (= semantically empty) clitic *ga*, the genitive form of the pronoun *on* 'he', coreferential with the proper noun *Ivan*. The use of the clitic indicates that 'Ivan' is communicatively Given. Sentence (20a(ii)) is only approximately synonymous with (20a(i)), which, however, does not prevent the interchangeability of these two sentence types.

4. Clausative-Presentative Constructions vs. Cleft Construction

Slavic Claus-PresCs are often compared and even identified with what is called the *cleft construction* in Germanic and Romance languages. The reason for this is obvious: both types of construction play the same discursive role—they serve to attract the Addressee's attention to a particular component of the Speaker's utterance. However, semantically and syntactically the Slavic Claus-PresCs and the Germanic/Romance cleft constructions are quite different; the similarity of their discourse roles—both are, in a sense, answers to a question—is not sufficient grounds for their identification.

DEFINITION 2: THE CLEFT CONSTRUCTION

A cleft construction is a surface-syntactic subtree of the following form:



where:

1. $L_{(V\ 'be')}$ is a copular verb 'be'.
2. $L_{(N,\ pronominal,\ empty)}$ is a semantically empty pronoun—a dummy subject (such as Eng. *it*[1]**5**, Fr. *il*, Ger. *es*; it can be zero—e.g., in Spanish and Italian).
3. $\hat{L}$(**RHEME$_{DSynt}$**) is the lexeme that expresses the DSynt-Rheme or is the SSynt-head of the phrase that expresses the DSynt-Rheme; in other words, it is the focalized element.
4. $\hat{L}$(SUBORD-CLAUSE) stands for 'SSynt-head of a subordinate clause'; this head is:
   a) either an empty subordinating conjunction, i.e., a complementizer (such as Eng. *that*[2]**1**, Fr. *que*, Ger. *dass*);
   b) or a finite verb that is the syntactic head of this subordinate clause and that has as a dependent a relative pronoun (such as Eng. *that*[1]**2** or *who*).

As a particular type of linguistic sign, the cleft construction is a surface-syntactic idiom, formed by three "grammatical" words (such as IT, BE, THAT); however, it is impossible to discuss the cleft construction here (see Mel'čuk, forthcoming).

Here are stock examples of the cleft construction from English:

(22) a. It was in May that we came to Boston.

b. It was John who came to Boston.

c. It is with John that we came to Boston.

d. It was John with whom we came to Boston.

The cleft construction is not characteristic of Slavic languages. It exists, as far as we know, only in Czech and Ukrainian and has rather limited usage there.

(23) a. Je to manželka, kdo rozhoduje. (Czech)
is that wife who decides
'It is the wife who decides.' (Reeve 2012: 167)

b. To je spadok, ščo xvyljuje joho. (Ukrainian)
that is inheritance that worries he.ACC
'It is the inheritance that worries him.' (Duma 2022: 1)

As is immediately seen, there are fundamental differences between the Claus-PresC and the cleft construction.

- The Claus-PresC (a) is defined on the **deep** Synt-level—its motivation is propositional-semantic; (b) it contains **no** grammatical lexemes; and (c) it is **mono**clausal.
- The cleft construction (a) is defined on the **surface** Synt-level—its motivation is communicative; (b) it contains **three** grammatical lexemes: a dummy pronoun, a copular verb, and either a complementizer or a relative pronoun; and (c) it is **bi**clausal.

Of course, a researcher is free to call Claus-PresCs "Slavic clefts", as do Reeve (2008) and Kimmelman (2009), for instance. This is, however, like calling English prepositions "English cases" because English prepositions often play the same syntactic role as cases in case languages. Therefore, we propose to avoid this terminological use: the Slavic Claus-PresC is not a cleft construction.

5. Lexicographic Descriptions of Russian and Serbian Clausative-Presentative Lexemes

As promised, we offer here the lexical entries for the Claus-Pres lexemes considered in this paper: Rus. *èto*[1]**III**/Serb. *to*[1]**III** and Rus. *vot*[1]/Serb. *evo***1**, along with a summary and non-exhaustive characterization of other related lexemes (co-polysemous or homophonous with ours), especially those with which our presentative lexemes may be confounded. Since the equivalent Russian and Serbian presentatives are similar enough in their linguistic properties, it makes sense to describe them in parallel, which we will do. Our presentation follows the standard format of a lexicographic entry in an explanatory combinatorial dictionary (see, for instance, Mel'čuk 2012–15, vol. 2: 276–314).

A detailed description of Russian lexeme *èto* is found in Padučeva (1982); the results of this paper are used in what follows. Lexemes *vot* and *von* were described in Krylova (2014). As for Serbian counterparts of these items, they

were described, rather cursorily, in two dictionaries: Stevanović et al. (1967), and its abridged version, Vujanić et al. (2011).

5.1. The Lexicographic Entry for Rus. *èto*¹III/Serb. *to*¹III

Rus. ***èto*¹III**/Serb. ***to*¹III**: clausative, pronominal, demonstrative (anaphoric), presentative.

Definition

'èto¹III/to¹III Y' = 'this/that situation is identical to Y or a result of Y'.

1. 'This/that [Ξ]' = '[Ξ] which is now in the active zone of your conscience' (i.e., Ξ is directly perceived or thought of by the Addressee).
2. The above definition is disjunctive (on such definitions and criteria for them, see Mel'čuk 2012–15, vol. 2: 324ff.). To illustrate, the first disjunct is implemented, for instance, in examples (28a) and (29a), and the second one in examples (28b–c) and (29b–c).

Government pattern

'Y' ⇔ II
1. **–copular-completive**→CLAUSE$_{(declarative)}$ **obligatory**

Syntactic properties

The only syntactic dependents (along with the $C_{(presentee)}$ clause) accepted by Rus. *èto*¹III are modal particles, such as *da* 'but, well', *nu* 'well', *prosto* 'simply', *razve* 'Is it the case that…?' and *ved´* 'but'/'after all' (see (24)). As for Serb. *to*¹III, it accepts as syntactic dependents the modal particles *a* 'well', *ma* 'but' and *samo*/*prosto* 'simply' (see (25)). A modal particle depends on *èto*/*to* via the restrictive surface-syntactic relation.

(24) a. Ved´ èto Ivan idët/prišël.
'But this is Ivan coming/having come.'

b. Razve èto Ivan idët/prišël?
'Is it the case that Ivan is coming/has come?'

(25) A/Ma to se Ivan nešto ljuti.
'Well, that ⟨But that⟩ is Ivan getting angry for some reason.'

Both *èto*^{1}III and *to*^{1}III can only express the Sem-Theme of the sentence. In neutral style, Rus. *èto*^{1}III is always linearly placed sentence-initially; Serb. *to*^{1}III can occupy the same position, but it can also be linearly preceded by the extracted element of the presentee clause which expresses the secondary Theme, provided the latter is **Focalized** (see example (16b), §3.1, p. 76).

The negation of the copular meaning 'be' inside the meaning 'èto^{1}**III**'/'to^{1}**III**'—that is, the meaning 'this situation←$_{1}$–**is**←$_{1}$–**not** (a result of) Y'—is expressed on the surface in the following way:

- In Russian, the particle *ne* 'not', subordinated to the main verb of the $C_{(presentee)}$ clause, is linearly placed clause-initially in $C_{(presentee)}$; the corresponding positioning indication must be added (during the DSyntR ⇔ SSyntR transition) to the syntactics of *ne* (see examples (11a) and (26)).[15]

(26) Èto, po vsej vidimosti, ne$_{(clause\text{-}initial)}$←[Ivan]–idët, a prosto polovicy skripjat. ~

Èto, po vsej vidimosti, ne$_{(clause\text{-}initial)}$←idët Ivan, a prosto polovicy skripjat.

'This is, apparently, not Ivan coming, but simply the floorboards creaking.'

- In Serbian, the particle *ne* 'not', subordinated to the main verb of the $C_{(presentee)}$ clause, is also linearly placed clause-initially, but in addition, it must be adjacent to the main verb (see examples (11b) and (27)).

(27) To, po svemu sudeći, ne$_{(clause\text{-}initial,\ verb\text{-}adjacent)}$←dolazi Ivan, to samo/prosto pod škripi. ~

*To, po svemu sudeći, ne $_{(clause\text{-}initial,\ verb\text{-}adjacent)}$ ←[Ivan]–dolazi, to samo/prosto pod škripi.

'This is, apparently, not Ivan coming, but simply the floorboards creaking.'

[15] Syntactics is the third component of the linguistic sign **s**, along with the Saussurean signified and signifier, which stores the idiosyncratic combinatorial properties of **s**; see Mel'čuk (2006a: 384ff).

Examples

(28) a. [Čto slučilos′? 'What happened?'] (Russian)

Da èto Ivan uronil knigi.

lit. 'But this.situation.is Ivan dropped books.'

'This situation, i.e., something happened, is Y, Y being "Ivan dropped books."'

b. [A kuda Ivan bežit? 'And where is Ivan running?']

Èto on spešit na avtobus.

lit. 'This.situation.is.a.result.of he hurries on bus.'

'This situation, i.e., Ivan is running, is a result of Y, Y being "Ivan is hurrying to catch the bus."'

c. [Ivan vinovato ulybnulsja. 'Ivan smiled guiltily.']

Èto on xotel pokazat′ nam, čto on vsë ponjal.

lit. 'This.situation.is.a.result.of he wanted to.show us that he everything understood.'

'This situation, i.e., his guilty smile, is a result of Y, Y being "Ivan wanted to show that he understood everything."'

(29) a. [Šta se dešava? 'What is going on?'] (Serbian)

To neko viče na ulici.

lit. 'That.situation.is some.people shout on street.'

'That situation, i.e., something that is going on, is Y, Y being "people are shouting in the street."'

b. [Kaže da će platiti, ali me izbegava. 'He.says he will pay, but he is avoiding me.']

To on hoće da vrati samo pola para.

lit. 'That.situation.is.a result.of he wants to return only half of.money.'

'That situation, i.e., he is avoiding me, is a result of Y, Y being "he wants to give back only half of the money."'

(29) c. [Upon seeing shopping bags scattered around]
To je Tanja opet bila u kupovini.
lit. 'That.situation.is.a.result.of is Tanja again been in shopping.'[16]
'That situation, i.e., the presence of shopping bags, is a result of Y, Y being "Tanja went shopping again."'

Two comments seem useful here.

Since 'Y', the Sem-actant of 'èto^{1}**III**'/'to^{1}**III**', denotes a situation, the $C_{(presentee)}$ clause must be declarative, i.e., it cannot be interrogative or exclamative. Therefore, in such sentences as (30), we find not *èto*^{1}**III**/*to*^{1}**III**, but *èto*^{2}/*to*^{2} (see also examples (38) and (39)).

(30) a. Èto kto ⟨Kto èto⟩ zdesʹ šumit? (Russian)
lit. 'This who ⟨Who this⟩ here makes.noises?'
'Who is this who is making noise here?'

b. Ko to ⟨*To ko⟩ ovde pravi buku? (Serbian)
lit. 'Who this here makes noise?'
'Who is this who is making noise here?'

In both Russian and Serbian ClausPresCs, the $C_{(presentee)}$ clause often features ellipsis of the main verb.

(31) a. Ja podumal, čto èto Ivan (Russian)
I thought that this.situation.is Ivan
prosto tak ⟨iz zavisti⟩.
just so out.of envy
'I thought that Ivan [**did/said**] this just so ⟨out of envy⟩.'

b. Niko mu ne plaća: to (Serbian)
nobody to.him not pays that.situation.is
on onako, za džabe.
he like.that for free
'Nobody pays him: he [**does**] that just so, free of charge.'

[16] The *je* after *to* is not the copula but rather the auxiliary of the past tense.

5.2. Other Rus. *èto* and Serb. *to* Lexemes

Here are lexemes that are co-polysemous/homophonous partners of *èto*^{1}**III**/ *to*^{1}**III**.

***èto*^{1}/*to*^{1}**

Nominal demonstrative pronouns whose meanings include the semanteme 'this/that [Ξ]' = '[Ξ] which is now in the active zone of your conscience'; they are anaphoric.

èto^{1}**I** (entity)

(32) a. Èto←**subj**–byl dom Ivana.
'This was Ivan's house.'

b. Dotronut´sja do–**prepos**→ ètogo ja ne sposoben.
'To touch this I am unable.'

to^{1}**I** (entity)

(33) a. To←**subj**–su naša deca.lit.
'That are our children.'

b. Ko ⟨Gde, Šta⟩ je–**subj**→ to?
'Who ⟨Where, What⟩ is that?'

èto^{1}**II.1** (entity or fact)

(34) a. Aspirin – èto←**subj**–Ø$^{BYT'}$ sredstvo ot vsego.
'Aspirin, this is a means against everything.'

b. Priznat´ ošibku – èto←**subj**– bylo nelegko.
lit. 'To admit the mistake this was not easy.'

c. Ko–[vsemu]→ètomu my privyknem.
lit. 'To all this we will get accustomed.'

to^{1}**II.1** (entity or fact)

(35) a. Aspirin, to←**subj**–je lek za sve.
'Aspirin, that is a remedy for everything.'[17]

b. Ti to←**dir.objectival**– [meni]– (kažeš)?
'You (are saying) that to me?'

c. Ko bi se toga setio?
'Who would remember that?'

[17] The noun *aspirin* in (34a) and (35a), as well as the infinitive *priznat´* in (34b), are prolepses. (Prolepsis is a fronted clause element, prosodically detached from the clause and morphologically independent of it, that syntactically depends on the head of the clause and fulfills a communicative role.)

èto^{1}**II.2** (fact) *to*^{1}**II.2** (fact)

This wordsense of *to* was described in Progovac (1998: 12–16).

(36) a. Marina xodit bez šapki, i èto pri takom moroze!
lit. 'Marina is.walking.around without a.hat, and that with such a.cold!'

b. On sil´no p´ët, no èto tol´ko raz v godu.
lit. 'He heavily drinks, but that only once a.year.'

(37) a. Marina uvek peva, i to lepo.
lit. 'Marina always sings, and that beautifully.'

b. Uskoro odlazim, i to sama.
lit. 'Soon I.am.leaving, and that alone.'

Interestingly, while *èto*^{1}**II.2** and *to*^{1}**II.2** do not exhibit semantic differences, their syntactic behavior is quite different; thus, a literal translation of (36b) into Serbian is ungrammatical, and the same is true of the translations of (37a–b) into Russian. We cannot, however, pursue this matter further here.

***èto*^{2}/*to*^{2}**

A modal invariable particle that syntactically depends on an interrogative pronoun and expresses the emotional or mental attitude of the Speaker in the situation in question—amazement, curiosity, disapproval, etc.—depending on the prosody. Both *èto*^{2} and *to*^{2} are colloquial. On the syntactic behavior of Serbian *to*^{2} see, in particular, Browne (1976: 200–01).

(38) a. Kuda→èto idët Ivan? ~ Èto←kuda idët Ivan?
lit. 'Where this ⟨This where⟩ goes Ivan?'

b. Začem [ty] èto vstaës´?
lit. 'Why you this stand up?'

(39) a. Ko→to tamo peva?
lit. 'Who that there sings?'

b. Ko ⟨Gde, Šta⟩ to?
lit. 'Who ⟨Where, What⟩ that?'

***èto*^{3}**

A focalization-marking clitic particle introduced into the SSynt-structure by a DSynt-rule as a marker of a focalized Rheme; the focalized rhematic expression is fronted and bears a phrasal stress and a rising contour.

(40) a. Èto^3–**IVÁN** moet posudu, a ne Petja.
'It is Ivan who is washing dishes, but not Pete.'

b. Èto^3–u **NÁS** muzyka igraet, a ne u sosedej.
'It is at our place that music is playing, but not at the neighbors.'

Serbian has no lexeme corresponding to Rus. *èto*^{3}, or any other "pure" segmental focalization marker (that is, a particle that has no propositional meaning and expresses exclusively focalization); in this language, focalization is expressed only by word order and/or prosody (which is rather the norm across the Slavic domain). Russian sentences featuring the focalizing *èto*^{3}, such as the one in (40a), can be translated into Serbian as shown in (41a–b); as an approximate translation, one could also use (41c), featuring the Claus-PresC headed by *to*^{1}**III**.

(41) a. Sudove pere **ÍVAN**, a ne Petar.
dish.PL.ACC wash.PRES.3SG Ivan.NOM but not Peter.NOM
'The dishes are being washed by Ivan, not by Peter.'

b. **ÍVAN** pere sudove, a ne Petar.
Ivan.NOM wash.PRES.3SG dish.SG.ACC but not Peter.NOM
'Ivan is (the one) washing the dishes, not Peter.'

c. To^1**III–copular-completive**–[**ÍVAN**]→pere sudove, a ne Petar.
lit. 'The situation is that Ivan is washing the dishes, not Peter.'

While in (41a–b) the primary focalization bears on a participant of the corresponding situation (the underlying question being 'Who does it?'), (41c) is about a situation within which a participant is the (prosodically) focalized secondary Rheme (with the underlying question 'What is this situation?').

5.3. The Lexicographic Entry for Rus. *vot*^{1}/Serb. *evo***1**

Rus. ***vot***^{1}/Serb. ***evo*1**: clausative, demonstrative, presentative, signalative.

Definition

'vot^1 Y/evo**1** Y-a' = 'I indicate here something that is Y.'

Government pattern

***vot*¹**	***evo*1**
'Y' ⇔ II	'Y' ⇔ II
1 **–cop-compl** → N_{NOM} 2. **–cop-compl** → $CLAUSE_{(pseudo\text{-}relative)}$[18] 3. **–cop-compl**→CLAUSE **obligatory**	1. **–cop-compl** → N_{GEN} 2. **–cop-compl** → $CLAUSE_{(pseudo\text{-}relative)}$ 3. **–cop-compl** → CLAUSE 4. **–cop-compl** → $PRON_{CLIT,\ GEN}$ N_{NOM} / PRON and N are coreferential **obligatory**

Syntactic properties

(i) Rus. *vot*¹/Serb. *evo***1** can express both the Sem-Theme and the Sem-Rheme of the sentence, but *evo***1** is more restricted in the role of Sem-Rheme (see examples (19–20) in §3.2, pp. 78–79).

(ii) Serb. *evo***1** cannot be the host for the verbal clitic *je* 'is' which is [$biti_{(locative)IND.PRES.3.SG}$], see examples (45b–c) below and the corresponding explanation.

Examples

(42) a. Vot Ivan ⟨moja komnata⟩. (Russian)
I.indicate.here Ivan ⟨my room⟩
'Here is Ivan ⟨my room⟩.' ≅
Ivan ⟨Moja komnata⟩ – vot.
'Ivan ⟨My room⟩ is here.'

[18] A pseudo-relative clause is a subordinated clause that has the structure of a relative clause (contains a WH-word) but is not a modifier of a noun (as a genuine relative is): a pseudo-relative clause is syntactically equivalent to a noun phrase and functions as a syntactic actant (e.g., *We will eat* ***what you brought***.); see Mel'čuk (2021: 249ff). In French and Italian examples (6c) and (7c), p. 70, the subordinated clauses *qui arrive/ che arriva* in the sentences *Le voici qui arrive/Eccolo che arriva* are pseudo-relatives.

(42) b. Vot to, čto ja tebe obeščal. (Russian)
I.indicate.here that what I to.you promised
'Here is (that) what I promised you.'

c. Vot gde ⟨s kem⟩ Ivan činit
I.indicate.here where ⟨with whom⟩ Ivan is.repairing
našu mašinu.
our car
'Here is where ⟨with whom⟩ Ivan is repairing our car.'

d. Vot Ivan opjat' $\varnothing^{\text{BYT'-cop}}\rightarrow$bolen.
I.indicate.here Ivan.NOM again ill
'Here is Ivan being ill again.'

e. Vot Ivan $\varnothing^{\text{BYT'-loc}}\rightarrow$v kuxne.
I.indicate.here Ivan.NOM in kitchen.SG.LOC
'Here is Ivan in the kitchen.'

(43) a. Evo Ivana ⟨moje sobe⟩. (Serbian)
I.indicate.here Ivan.GEN my room.SG.GEN
'Here is Ivan ⟨my room⟩.'

b. Evo ga ⟨njega⟩.
I.indicate.here he.CLIT.GEN he.FULL.GEN
'Here is he ⟨**HÉ**⟩.'

c. Evo ga ⟨*njega⟩ Ivan.
I.indicate.here he.CLIT.GEN he.FULL.GEN Ivan.NOM
'Here is he ⟨***HÉ**⟩ Ivan.'

d. Evo onoga što me najviše brine.
I.indicate.here that.SG.GEN what I.ACC most worry.PRES.3SG
'Here is (that) what worries me the most.'

e. Evo gde Ivan živi.
I.indicate.here where Ivan.NOM live.PRES.3SG
'Here is where Ivan lives.'

f. Evo Ivan je opet bolestan.
I.indicate.here Ivan.NOM $\text{be}_{(\text{cop})}$.PRES.3SG again ill
'Here is Ivan being ill again.'

(43) g. Evo Ivan je u kuhinji.
I.indicate.here Ivan.NOM be$_{(loc)}$.PRES.3SG in kitchen.SG.LOC
'Here is Ivan in the kitchen.'

h. Evo su deca u
I.indicate.here be$_{(loc)}$.PRES.3PL child.PL.NOM in
kuhinji.
kitchen.SG.LOC
'Here are the children in the kitchen.'

Some additional explanations are needed here.

The meaning of Rus. *vot*[1]/Serb. *evo***1** can be easily taken as an order/request with the imperative meaning; cf. the gloss 'look here/there', often used for both lexemes. But this is not the case. The meaning of Rus. *vot*[1] does not include the imperative semanteme 'I want you to…'. This is explicitly shown by the fact that *vot*[1] does not accept the particle *-ka*, which "mitigates" orders and instructions and can be attached to any verbal imperative form and to some particles that express incitement: *Na-ka!* 'Please take!' or *Nu-ka!* ≈ 'Please do!'. By analogy, we draw the same conclusion for *evo***1**. At the same time, *vot*[1]/*evo***1** are non-descriptive lexemes—their meanings must include the semanteme 'I'; the above definition—'I indicate here…'—does exactly this.

The sequence *vot gde* in (42c) should not be confounded with the sequence *vot gde* in the following sentence.

(44) Tak vot gde Ivan činit našu mašinu!
'So this is where Ivan is repairing our car!'

The meanings, prosodies, and SSynt-structures are completely different in these cases:

- *vot* in (42c) means 'I indicate here where…', while its formal counterpart in (44) has the meaning 'Now I know where…!'
- Sentence (42c) is declarative, *vot gde* is uttered with neutral (= flat) intonation, and the two lexemes can be separated by a parenthetical: *Vot, kak ty vidiš´, gde Ivan činit mašinu* 'Here is, as you can see, where Ivan is repairing the car'. In contrast, sentence (44) is exclamative, *vot gde* is uttered with a falling intonation, with strong phrasal stress on *vot*, and without a possible pause between *vot* and *gde*.
- While *vot* in (42c) is our clausative-presentative, serving as the syntactic head of the corresponding syntactic construction, *vot* in (44) is a modal particle that syntactically depends on a relative pronoun (*kto* 'who', *čto*

'what', *gde* 'where', *kogda* 'when', *začem* 'what for', etc.), with which it forms a phraseologized combination and means ≈ 'now I know…'; cf. respectively:

vot–**copular-completive**→*činit'*–**circumstantial**→*gde*
vs.
vot←**restrictive**–*gde*←**circumstantial**–*činit'*

As shown in examples (42d–e) for Russian and (43f–h) for Serbian, the copular and the locative verbs *byt'*/*biti* 'be' can appear as the syntactic head of the $C_{(\text{presentee})}$ clause that functions as the actant of *vot*¹/*evo***1** (albeit the corresponding construction is best suited for event-denoting verbs). In Serbian, this allows for a variation of the type illustrated in (45a), involving the locative *biti* 'be', with the following restriction: if the locative verb appears in the third person singular of the present tense, it cannot be hosted by *evo***1**. In such a case, if the structure of the sentence allows for it, *evo***1** must be skipped (i.e., not counted as a possible host of the locative *biti*) or else the construction with the pronoun in the genitive must be used. This is shown in (45b–c).

(45) a. Evo $\text{sam}_{(\text{V, loc})}$ (ja) u sobi. or Evo $\text{me}_{(\text{N, pron}) \text{ GEN}}$ u sobi.
'Here am (I) in the room.' lit. 'Here me in room.'

b. *Evo $\text{je}_{(\text{V, loc})}$ (on) u sobi. ⇒ Evo on $\text{je}_{(\text{V, loc})}$ u sobi.
'Here is (he) in the room.' 'Here he is in room.'
or Evo $\text{ga}_{(\text{N, pron})\text{GEN}}$ (on) u sobi.
lit. 'Here him (he) in room.'

c. *Evo $\text{je}_{(\text{V, loc})}$ (ona) u sobi. ⇒ Evo ona $\text{je}_{(\text{V, loc})}$ u sobi. or
Evo $\text{je}_{(\text{N, pron})\text{GEN}}$ (ona) u sobi.
'Here is (she) in the room.'

Serbian sentences with a locative *biti* in the $C_{(\text{presentee})}$—such as in (45a)—are colloquial and are deemed "substandard" by normative grammars.[19]

[19] The ungrammaticality of the starred sentences in (45b–c) is probably explainable by the fact that *je*, the expression ambiguous between the 3SG form of the locative verb *biti* 'be' and the genitive singular form of the pronoun *ona* 'she', is preferably interpreted by the Addressee as the latter, which may result in confusion. We have here a surprising example of a highly idiosyncratic constraint on "clitic ~ host" co-occurrence—imposed on a particular host by a particular clitic. More generally, we see to what extent clitics are important in syntax, even in cases such as ours, where they are not directly involved. (For a recent publication on Serbian clitics, see Milićević 2023.)

5.4. Other Rus. *vot*/Serb. *evo* Lexemes

We will limit ourselves to indicating some of the *vot*/*evo* lexemes whose uses can be confounded with *vot*¹/*evo*₁ in the Claus-PresC.

The narrative *vot*/*evo* ≈ 'I will describe what follows as if I were present there'

(46) Vot prixodit Ivan domoj i vidit…
'So Ivan comes home and sees…'

(47) I evo, posle mnogo godina, Ivan se vraća kući.
'And so, after many years, Ivan is returning home.'

The Actor's immediate action *vot*/*evo* ≈ 'right away…'

(48) a. Vot on zakončit pis´mo i pojdët.
'Now he'll.finish the.letter and will.go.'

b. Vot ja tebe zadam!
'Now I'll teach you a lesson!'

(49) a. Evo dolazi.
lit. 'Right.away (s/he's) coming.'

b. Evo, evo!
lit. 'Right.away, right.away!' = 'Coming, coming!'

The explanatory *vot*/*evo*

(50) a. Vot čto ja xotel tebe skazat´.
'Here is what I wanted to tell you.'

b. Vot kak ⟨počemu⟩.
'Here is how ⟨why⟩.'

(51) a. Evo šta sam hteo da ti kažem.
'Here is what I wanted to tell you.'

b. Evo kako ⟨zašto⟩.
'Here is how ⟨why⟩.'

The offering *vot*/*evo*

(52) Vot tebe kniga ⟨to, čto ja tebe obeščal⟩.
lit. 'Here to.you book ⟨that what I to.you promised⟩.'

(53) Evo ti knjiga ⟨ono što sam ti obećao⟩.
lit. 'Here to.you book ⟨that what I to.you promised⟩.'

The "temporal" *vot/evo*

(54) Vot uže tri dnja, kak idët dožd´.
'It's been raining for three days already.'

(55) Evo već tri dana kako pada kiša.
'It's been raining for three days already.'

The conjunction *kak/kako* is necessary: without it, we have the Claus-PresC with *vot*[1]; *uže/već* is desirable.

The phraseologized *vot* (*evo* has similar, albeit not as rich, phraseology):

- With an interrogative pronoun, the resulting combined meaning 'now I know where ⟨where from, etc.⟩': *vot* + *gde* 'where' ⟨*kuda* 'where to', *otkuda* 'where from', *kogda* 'when', *kto* 'who', *čto* 'what', etc.⟩.
- With a demonstrative locative-temporal adverbial pronoun, the resulting combination meaning 'exactly here ⟨to there, etc.⟩': *vot* + *zdes´* 'here', *tuda* 'to there', *ottuda* 'from there', *sejčas* 'this moment', *teper´* 'now', *togda* 'then'.
- In idioms such as ⌜*Vot i*⌝, 'Here is [X]', ⌜*Da vot*⌝, 'But that's how it is…', ⌜*Nu vot*⌝, 'Well then', ⌜*Tak vot*⌝, 'That's how it is', ⌜*Vot-te/Vot tebe na*⌝, 'Gee!', ⌜*Vot èto da*⌝!, 'Gee!', ⌜*Vot èto*⌝ [X]!, 'What a wonderful [X]!', etc.

This concludes the lexicographic description of the Russian and Serbian Claus-Pres lexemes.

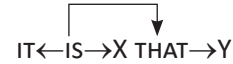
6. Conclusions

Two general results of our research are presented, followed by several specific remarks on the Claus-Pres lexemes analyzed.

6.1. General Remarks on the Slavic Clausative-Presentative Construction

Strictly speaking, the Claus-PresC is not a construction—at least, not in the same sense in which the cleft construction is. In today's linguistic literature, the quasi-term *construction* is liberally applied to any multiword expression having any special property. There is nothing wrong in that: such a "wild card" term proves convenient in many an informal discussion. However, at some point it is also necessary to precisely specify each object of our description which we provisionally call "construction".

Thus, the Claus-PresC is a phrase consisting of a Claus-Pres lexeme with its actant, the presentee clause or noun. To put it differently, *èto*[1]**III**/*to*[1]**III** + CLAUSE and *vot*[1]/*evo***1** + NOUN ⟨CLAUSE⟩ are the same type of linguistic entity as SAY + *that* CLAUSE or DEPEND + *on* N. As a result, the description of Slavic Claus-PresCs boils down to the elaboration of the lexical entries for the corresponding Claus-Pres lexemes. In contrast, the cleft construction is a genuine construction, or more precisely (as stated above), a surface-syntactic idiom—namely, a syntactic configuration of three grammatical lexemes:



A Claus-PresC has a rather unusual syntactic structure: it underlies a monoclausal subjectless sentence headed by a clausative lexeme that governs a copular complement—a clause. Such syntactic phenomena have not yet been sufficiently studied.

6.2. Specific Remarks on the Slavic Clausative-Presentative Construction

While Claus-Pres lexemes heading the Claus-PresC have similar semantics, *èto*[1]**III**/ *to*[1]**III** are linguistically more anchored and less "ostensive" than *vot*[1]/ *evo***1**. The latter lexemes are in a sense not fully linguistic: they represent a kind of gesture pointing to something that can be perceived here and now, taking place in the same space and contemporaneously with the utterance.

Russian and Serbian Clause-presentatives *èto*[1]**III**/*to*[1]**III** and *vot*[1]/*evo***1** turned out to have remarkably similar semantic and syntactic properties. (While we expected them to behave similarly, given the genetic relatedness of the two languages, this almost perfect parallelism did come as a surprise.) Among the few differences noted, we could mention the following:

- Less flexible linear positioning of *èto*[1]**III**, which, unlike *to*[1]**III**, must be sentence-initial (in neutral style).
- Unrestricted use of tenses other than the present in sentences headed by *èto*[1]**III** vs. low acceptability of the past and future tenses in sentences with *to*[1]**III** (probably due to clitic auxiliaries, whose adjacency to *to*[1]**III** may force its erroneous interpretation and result in confusion).
- More flexible word order in Russian in the presentee clause headed by both *èto*[1]**III.1** and *vot*[1].
- More versatile communicative properties of *vot*[1], functioning as either the Theme or the Rheme of the sentence, with *evo***1** being more appropriate in the communicative role of the Theme.

- More complex government of Serb. *evo***1**, which can take a clitic resumptive pronoun as an additional SSynt-dependent, alongside the noun co-referenced by the clitic.

A remarkable difference between Russian and Serbian, not directly linked to Claus-Pres lexemes but rather to their homophonous "partners", concerns the means of expressing focalization. While both languages employ to that effect both the prosody and the word order, Russian has an additional means—a dedicated focalization particle *èto*[3], with no counterpart in Serbian.

Abbreviations, Symbols, and Writing Conventions

ACC	accusative (grammeme of case)	LOC	locative (grammeme of case)
Aux	auxiliary verb	MASC	masculine (grammeme of adjectival gender)
CLIT	clitic (grammeme of tonicity)	neu	neuter (nominal gender)
Claus-PresC	clausative-presentative construction	NOM	nominative (grammeme of case)
cop	syntactic feature of the copular verb 'be'	P.PART	past participle
DAT.ETH	ethical dative (syntactic role)	PL	plural (grammeme)
DemClaus	demonstrative clausative lexeme	SemR	semantic representation
DSyntR	deep-syntactic representation	SG	singular (grammeme)
FULL	full (grammeme of tonicity)	SSyntR	surface-syntactic representation
GEN	genitive (grammeme of case)	Ø	zero (= phonologically null) wordform
L	the given lexeme	**I**, **II**, ...	deep-syntactic actant **I**, **II**, … (of a lexeme)
lit.	literal gloss (of an expression)	'x'	meaning of the expression *x*
loc	syntactic feature of the locative verb 'be'	*x*=*y*	boundary between host *x* and enclitic *y*

References

Browne, Wayles. (1976) "Two WH-fronting rules in Serbo-Croatian". *Južnoslovenski filolog* 32: 195–204.

Burukina, Irina, and Marcel den Dikken. (2020) "Russian *èto*-focus and *to*-topic sentences as elliptical question-answer pairs". *Proceedings of the 22nd Seoul International Conference on Generative Grammar (SICOGG 22)*, 12–14 August 2022, Gyeongsang National University, Jinju, South Korea, 63–79.

Duma, Yehor. (2022) "Structure of clefts in Ukrainian". Available at: https://www.researchgate.net/publication/364904045.

Grenoble, Lenore, and Matthew Riley. (1996) "The role of deictics in discourse coherence: French *voici/voilà* and Russian *vot/von*". *Journal of pragmatics* 25(6): 819–38.

Iomdin, Leonid. (2010) "Sintaksičeskie otnošenija". Jurij Apresjan, Igor´ Boguslavskij, Leonid Iomdin, and Vladimir Sannikov, *Teoretičeskie problemy russkogo sintaksisa. Vzaimodejstvie grammatiki i slovarja*. Moscow: Jazyki slavjanskix kul´tur, 21–43.

Junghanns, Uwe. (1997) "On the so-called ÈTO-cleft construction". Martina Lindseth and Steven Franks, eds. *Proceedings of the Sixth Annual Workshop on Formal Approaches to Slavic Linguistics*. Stanford: CSLI Publications, 166–90.

Kimmelman, Vadim. (2009) "On the interpretation of *èto* in so-called *èto*-clefts". Available at: https://vadimkimmelman.com/papers/Kimmelman%202009%20clefts.pdf.

Kordić, Snježana. (2002) *Riječi na granici punoznačnosti*. Zagreb: Hrvatska sveučilišna naklada.

Krylova, Tat´jana. (2014) Lexical entry for *vot*. Jurij Apresjan, ed. *Aktivnyj slovar´ russkogo jazyka. Tom 2. V – G.* Moscow: Jazyki slavjanskoj kul´tury, 284–88.

Lambrecht, Knud. (1988) "Presentational cleft constructions in spoken French". John Haiman and Sandra Thompson, eds. *Clause combining in grammar and discourse.* Amsterdam/Philadelphia: John Benjamins, 135–79.

Lambrecht, Knud. (2000) "Prédication seconde et structure informationnelle: La relative de perception comme construction présentative". *Langue française* 127: 49–66.

Li, Charles, and Sandra Thompson. (1989) *Mandarin Chinese: A functional reference grammar.* Berkeley/Los Angeles/London: University of California Press.

Mel'čuk, Igor. (1988) *Dependency syntax: Theory and practice*. New York: State University of New York Press.

————. (2001) *Communicative organization in natural language*. Amsterdam/Philadelphia: John Benjamins.

————. (2006a) *Aspects of the theory of morphology*. Berlin/New York: Mouton de Gruyter.

Mel'čuk, Igor. (2006b) "Parties du discours et locutions". *Bulletin de la Société de linguistique de Paris* 101(1): 29–65.

————. (2012–15) *Semantics: From meaning to text*. Vols. 1–3. Amsterdam/Philadelphia: John Benjamins.

————. (2021) *Ten studies in dependency syntax*. Berlin/Boston: De Gruyter Mouton.

Mel'čuk, Igor, and Jasmina Milićević. (2020) *An advanced introduction to semantics. A Meaning-Text approach*. Cambridge, UK: Cambridge University Press.

————. (forthcoming) "The cleft construction: A formal definition". *Russian journal of linguistics*.

Milićević, Jasmina. (2023) *Serbian clitics*. Amsterdam/Philadelphia: John Benjamins.

Morin, Yves-Charles. (1985) "On the two French subjectless verbs *voici* and *voilà*". *Language* 61(4): 777–820.

Padučeva, Elena. (1982) "Značenie i sintaksičeskie funkcii slova *èto*". Viktor Grigor´ev, ed. *Problemy strukturnoj lingvistiki 1980*. Moscow: Nauka, 76–91.

Peti, Mirko. (2005) "Bezlično-bezrodne zamjenice". August Kovačec, ed. *Zbornik o sedamdesetpetoj godišnjici akademika Dalibora Brozovića*. Zagreb: HAZU, 577–86.

Porhiel, Sylvie. (2012) "The presentative *voici/voilà*—Towards a pragmatic definition". *Journal of pragmatics* 44(4): 435–52.

Progovac, Ljiljana. (1998) "Event pronominal *TO*". *Journal of Slavic linguistics* 6(1): 3–39.

————. (2005) *A syntax of Serbian*. Bloomington, IN: Slavica Publishers.

Raković, Sanja. (2017) "The role of clitics in Serbian presentative constructions". *Linguistica Brunensia* 65(2): 53–68.

Reeve, Matthew. (2008) "A pseudo-biclausal analysis of Slavonic clefts". *University College London working papers in linguistics* 20: 63–85.

————. (2012*) Clefts and their relatives*. Amsterdam/Philadelphia: John Benjamins.

Stevanović, Mihailo, and Ljudevit Jonke. (1967) *Rečnik srpskohrvatskoga književnog jezika.* Novi Sad/Zagreb: Matica srpska/Matica hrvatska.

Tesnière, Lucien. (1959) *Éléments de syntaxe structurale*. Paris: Klincksieck.

Vujanić, Milica. (2011) *Rečnik srpskoga jezika.* Novi Sad: Matica srpska.

Wood, Jim, and Raffaella Zanuttini. (2023) "The syntax of English presentatives". *Lingua* 99(3): 563–602.

Zanuttini, Raffaella. (2017) "Presentatives and the syntactic encoding of contextual information: Theory and description". Enoch Aboh, Eric Haeberli, Genoveva Puskás, and Manuela Schönenberger, eds. *Elements of comparative syntax: Theory and description*. Berlin/Boston: De Gruyter Mouton, 221–55.

Igor Mel'čuk
Département de linguistique et de traduction
Université de Montréal
Montreal, Quebec, Canada
igor.melcuk@umontreal.ca

Jasmina Milićević
Department of French
Dalhousie University
Halifax, Nova Scotia, Canada
jmilicev@dal.ca

Disassembling and Reassembling Pronouns: A Case Study of Bosnian/Croatian/Montenegrin/Serbian

Zorica Puškar-Gallien

Abstract: This paper explores the building blocks of personal pronouns in order to provide a unified model of the form, locus, and function of ϕ- and case features of pronouns that will account for their morphological distinctions and agreement properties. The proposal bears on the notion of hierarchy within the syntactic projections in the nominal domain, such that the base (*n*P) is dominated by ϕ-features (in the order: Person, Number, Gender), which are in turn dominated by a case hierarchy. The structure of pronouns proposed in this paper is shown to have consequences for pronominal morphology: third-person pronouns resemble nouns in that both consist of an *n*P base, dominated by number-, gender-, and case-bearing functional heads. First- and second-person pronouns, on the other hand, are also based on an *n*P, but they crucially lack grammatical gender. Both types of pronouns differ from nouns in lacking a lexical root (Moskal 2015b; Smith et al. 2019). The proposal for their morphological realization, based on the assumption that the *n*P and gender-bearing phrase are phases, will account for various types of suppletion found in their paradigms, as well as the similarities and differences in the spell-out of strong pronouns and clitics.

1. Introduction

This paper provides an in-depth investigation of the building blocks that construct the system of personal pronouns in Bosnian/Croatian/Montenegrin/Serbian (BCMS). It provides a unified analysis of the form, locus, function, and morphological realization of ϕ- and case features of pronouns that accounts for their similarities and distinctions. The model strives to integrate pronouns into a general theory of the morphosyntactic representation of nominal categories through capturing the hierarchies in their feature structures as well as those in their functional spine, ultimately offering a proposal on their spell-out and exploring its further consequences. Personal pronouns in BCMS show morphological distinctions along two dimensions: *local person* (first- and second-person) pronouns vs. third-person pronouns on one hand, and strong pronouns vs. clitics on the other. The form of local person pronouns varies depending on the person, number, and case of the pronoun. The shape of third-person pronouns varies depending on their gender, number, and case, while

Journal of Slavic Linguistics 32(1): 103–144, 2024.

they share the same base. The latter property is something that morphologically relates them to nouns, although they have different sets of inflectional endings, with the pronouns sharing the adjectival inflectional paradigm.

While local person pronouns show suppletion for number and case, third-person pronouns undergo only case suppletion. Assuming that pronouns are not simple bundles of features but may instead have a relatively complex internal structure, this indicates that there are differences between local person and third-person pronouns that should receive an adequate explanation. Another relevant point of their divergence relates both to their referential and morphological properties: Local person pronouns lack gender distinctions but can nevertheless control[1] gender agreement, while third-person pronouns do show gender distinctions and correspondingly control gender agreement.

Furthermore, clitics differ from strong pronouns in their morphological realization, as they basically present the reduced form of the pronoun, either lacking the base (third person) or lacking stress and an additional morpheme (local person). For instance, the accusative forms of third-person singular pronouns are *njega* '3SG.M.ACC', *nje* '3SG.F.ACC', and *njega* '3SG.N.ACC', while the corresponding clitics are realized by a portmanteau morpheme expressing gender, number, and case, omitting the base *n(je)-*, i.e., *ga*, *je*, and *ga*. In addition to this, clitics are more flexible than strong pronouns in their reference, as they allow for both animate and inanimate referents. Such differences in reference should ultimately also be taken into account.

In modeling the internal structure of a pronoun, the analysis proposed in this paper will rely heavily on the notion of syntactic *hierarchy*. Building on the insights of Déchaine and Wiltschko (2002); Weerman and Evers-Vermeul (2002); Neeleman and Szendrői (2007); Barbiers et al. (2009); Moskal (2015b); Smith et al. (2019); and van Urk (2018), I will argue that the hierarchy of pronominal extended projection encompasses three zones: a base (*n*P), followed by ϕ-features (PhiP), followed by case (KP). I will then depart from these proposals by arguing that the PhiP further decomposes into three projections, such that PERSON precedes NUMBER, which in turn precedes GENDER, reflecting the markedness hierarchy of these features (Greenberg 1963; Noyer 1992). This will be represented by person being the lowest and grammatical gender being the highest projection of the three. Within KP, following Caha (2009), the features stand in an entailment relationship such that UNMARKED (NOM) case precedes DEPENDENT (ACC, GEN) case, which is encoded before OBLIQUE (DAT) case.

[1] I use the term "to control agreement" in the sense of Corbett (2006: 4), who distinguishes between agreement *controllers*, elements that determine the agreement (e.g., subject NPs), and agreement *targets*, elements that change in form depending on the controllers' properties.

Finally, the highest projection in each of the (lower) zones, i.e., the *n*P and the gender phrase (cLP), will delimit a locality domain by being a phase.

The proposal will ultimately arrive at a unified account of nominal categories: They are built on a basic category *n*P. Nouns and pronouns differ in that the *n*P in nouns consists of a nominalizing head *n* that categorizes a lexical root, while the *n*P in pronouns consists of the nominalizer alone (combining the proposals of Moskal 2015a, 2015b; Moskal and Smith 2016; Smith et al. 2019; and van Urk 2018). Clitics lack *n*P altogether, and its phasal status is what this difference partially results from.

Additionally, the analysis will provide a means to distinguish between local person and third-person pronouns: Both categories encode person, number, gender, animacy, and humanness (without the need to phonologically realize them); however, they differ in their extended projections. The absence of a particular feature (e.g., gender on local person pronouns, or person on nouns) will be modeled by the lack of the relevant projection. A distinguishing feature of third-person pronouns is the *grammatical* gender they bear.

In addition to providing a spell-out model for strong pronouns and clitics, the analysis accounts for the suppletion patterns of pronouns by treating suppletion as contextual allomorphy in the sense of Moskal (2015a, 2015b) and Moskal and Smith (2016). I will, however, complement these accounts by integrating gender in this model and arguing that relativized locality should be abandoned in favor of strict cyclic domains such as those announced above. As a result, the locality domain delimited by the gender projection will account for the suppletion patterns of the third-person pronouns, in particular the distinction between their nominative and non-nominative forms.

This paper is structured as follows. Section 2 presents the paradigms of personal pronouns, scrutinizing their morphological and referential properties. Some challenges that this set of data poses to previous analyses of pronominal structure and contextual allomorphy are presented in section 3. The proposals on how pronouns should be disassembled to their atomic parts is provided in section 4, while section 5 outlines a proposal on their morphological realization. Subsequently, section 6 inspects some consequences of the analysis for clitics and demonstratives, and section 7 concludes.

2. Data

2.1. The Morphology of BCMS Pronouns

Let us start by considering the basic set of personal pronouns in BCMS in Table 1. Local person pronouns show person, number, and case distinctions.

Third-person (3π) pronouns share a common stem *on-*,[2] but their endings reveal a three-way gender distinction: masculine (M), feminine (F), and neuter (N).

Table 1. Personal pronouns in BCMS

	1SG	2SG	1PL	2PL	3SG.M/N	3SG.F	3PL
NOM	*ja*	*ti*	*mi*	*vi*	*on-Ø/-o*	*on-a*	*on-i*.M/*-e*.F/*-a*.N
GEN	*m-en-e*	*t-eb-e*	*na-s*	*va-s*	*nje-ga*	*nj-e*	*nj-ih*
DAT	*m-en-i*	*t-eb-i*	*na-ma*	*va-ma*	*nje-mu*	*nj-oj*	*nj-ima*
ACC	*m-en-e*	*t-eb-e*	*na-s*	*va-s*	*nje-ga*	*nj-u*	*nj-ih*
INST	*m-n-om*	*t-ob-om*	*na-ma*	*va-ma*	*nj-im*	*nj-om*	*nj-ima*
LOC	*m-en-i*	*t-eb-i*	*na-ma*	*va-ma*	*nje-mu*	*nj-oj*	*nj-ima*

Each nominative local person pronoun has a unique form; they have suppletive forms in the plural (here, I use the term *suppletion* to indicate a single lexical item associated with two phonologically unrelated forms; Moskal 2015a). In environments other than nominative singular, the first-person pronoun uses a suppletive stem (*m-*), while the stem of the second-person pronoun does not change in non-nominative environments. I will assume that the stem encodes person and number features. First- and second-person pronouns share the same set of case exponents, which is how I will interpret their affixes. Local person pronouns thus realize their base, consisting of person (π) and number (#),[3] by means of one morpheme, separately from their case features.[4]

[2] Also, the stem of distal demonstratives; see §6.2.

[3] Throughout the paper, I will be using the following abbreviations: 1 "first person", 2 "second person", 3 "third person", SG "singular", PL "plural", M "masculine gender", F "feminine gender", NOM "nominative", GEN "genitive", DAT "dative", ACC "accusative", INST "instrumental", LOC "locative", CLT "clitic", π "person", # "number", CL "class", HUM "human", ANIM "animate", INANIM "inanimate", PRTCPT "participant", SPKR "speaker", AUX "auxiliary", PRT "participle", UNM "unmarked", DEP "dependent", OBL "oblique", and DU "dual".

[4] I will consider the morphemes *-en-* and *-eb-* in the singular to be the so-called "support morphemes" (Cardinaletti and Starke 1999), which distinguish the "strong" pronoun forms from their clitic counterparts. The clitic forms of those pronouns are the simple *me* and *te*, without this extension.

The paradigm of the third-person pronouns is essentially adjectival, i.e., other than the stem *on-*, which they all share, their suffixes are the same as the suffixes of adjectives, realizing gender, number, and case features. The nominative suffixes on third-person pronouns are identical to the gender and number suffixes on nouns. Masculine nouns belonging to declension class I end with a consonant (presumably having the -Ø suffix), just like the masculine pronoun (cf. *on-Ø* '3-M.SG.NOM' vs. *dečak-Ø* *'boy*-M.SG.NOM'); feminine nouns typically end in *-a*, just like the feminine pronoun (cf. *on-a* '3-F.SG.NOM' vs. *devojk-a* 'girl-F.SG.NOM'); and neuter nouns end in *-o* (or *-e*, which will be put aside), just like the neuter pronoun (cf. *on-o* '3-N.SG.NOM' vs. *kril-o* 'wing-N.SG.NOM'). In the non-nominative cases, all third-person pronouns' stems undergo suppletion.

Finally, clitic forms of pronouns (represented in boldface in Table 2) are available in genitive, dative, and accusative case. They are essentially morphologically reduced forms of strong pronouns. This reduction is achieved in different ways with local person and third-person pronouns. Local person clitics realize person, number, and case, without the support morpheme. Third-person clitics are the spell-out of the gender, number, and case feature bundle (arguably also person, judging from the feminine form that includes the morpheme *-j-*), leaving the pronominal base out.

Table 2. Clitics vs. strong pronouns in BCMS

	1SG	2SG	1PL	2PL	3SG.M/N	3SG.F	3PL
NOM	*ja*	*ti*	*mi*	*vi*	*on-Ø/-o*	*on-a*	*on-i/-e/-a*
GEN	***m**-en-**e***	***t**-eb-**e***	***na-s***	***va-s***	*nje-**ga***	*nj-**e***	*nj-**ih***
DAT	***m**-en-**i***	***t**-eb-**i***	***na-ma***	***va-ma***	*nje-**mu***	*nj-**oj***	*nj-**ima***
ACC	***m**-en-**e***	***t**-eb-e*	***na-s***	***va-s***	*nje-**ga***	*nj-**u***	*nj-**ih***
INST	*m-n-om*	*t-ob-om*	*na-ma*	*va-ma*	*nj-im*	*nj-om*	*nj-ima*
LOC	*m-en-i*	*t-eb-i*	*na-ma*	*va-ma*	*nje-mu*	*nj-oj*	*nj-ima*

2.2. Additional Explanatory Desiderata: Agreement

Apart from the differences in their morphology, local person pronouns and third-person pronouns differ in how they control agreement. A peculiar property of local person pronouns in BCMS is that they can control natural

gender agreement. Specifically, agreement present at their agreement targets reflects the notional gender of their referent, as indicated in (1). This is true only of masculine and feminine genders. Neuter gender cannot be used in agreement with first and second person, (2).

(1) a. Ja sam došla / *došao.
1SG AUX.1SG came.F.SG came.M.SG
'I (female referent) came.'

b. Mi smo došle / *?došli.
1PL AUX.1PL came.F.PL came.M.PL
'We (female referents) came.'

(2) a. *Ja sam došlo.
1SG AUX.1SG came.N.SG
Intended: 'I (neuter) came.'

b. *Mi smo došla.
1PL AUX.1PL came.N.PL
Intended: 'We (neuter) came.'

In contrast, third-person pronouns control agreement in accordance with their grammatical gender, (3). We know that this gender is purely formal since a pronoun that refers to an inanimate entity can control agreement, (4). It should be noted that examples like (4) are quite marginal and only acceptable if the inanimate pronoun expresses some kind of focus or contrast, as is the case here; otherwise, strong pronouns generally almost exclusively allow animate interpretation (see §6.1).

(3) a. On je došao. b. Ona je došla.
3SG.M AUX.3SG came.M.SG 3SG.F AUX.3SG came.F.SG
'He came.' 'She came.'

(4) Ovo je moj novi bicikl.
this AUX.3SG my new bicycle.M.SG
On je mnogo brži od starog.
3SG.M AUX.3SG much faster.M.SG than old
'This is my new bicycle. It is much faster than the old one.'

The distinctions above pose the question of whether natural gender is a part of a featural representation of a strong pronoun. Depending on that, it calls for an investigation of how it participates in agreement. On the other

hand, the formal grammatical gender of the third-person pronouns also raises the issue of its formal representation as well as similarities to, and differences from, natural gender.

2.3. Summary and Main Questions

To sum up, although local person pronouns and third-person pronouns are similar in their feature inventories, encoding person, number, gender (overtly at least on third person), and case, they differ in the ways these features are morphologically represented. While number and case are present as categories on all of them, local person pronouns lack overt gender features. Number and person seem to be able to form a morphological unit to the exclusion of gender and case (pronominal base of local person pronouns), or form a unit together with case (resulting in a third-person clitic), while gender can form a morphological unit with number and case (and as such be realized either as a clitic or as an agreement affix). Finally, local person pronouns can control gender agreement without having an overt gender feature, which poses the question of how such features are encoded in the grammar.

The main question that the rest of the paper will aim to answer concerns the structural encoding of phi-features and case features on (pro)nouns and their realization, especially such that the patterns of suppletion can fall out from their internal syntactic structure. In particular, we want to distinguish between the number- and case-conditioned suppletion of local person pronouns and case-triggered suppletion only in non-nominative environments with third-person pronouns. If it is true that this can be made to follow from their syntactic structure, this should then straightforwardly capture the morphological realization of clitic forms as a natural consequence of the spell-out process. Finally, I also aim to uncover the structural encoding of properties responsible for allowing natural gender agreement.

3. Previous Literature

The bulk of previous literature has converged on the idea that, being nominal categories themselves, pronouns can project complex internal syntactic structure similar to that of regular noun phrases. For instance, for Postal (1969) and Elbourne (2005), pronouns realize a DP without a noun, which equates them to definite articles. The encoding of phi-features and case features on (pro)nouns and their realization has been subject to much debate. Some very specific proposals were advanced by Déchaine and Wiltschko (2002); Weerman and Evers-Vermeul (2002); Neeleman and Szendrői (2007); Barbiers et al. (2009); Moskal (2015b); Smith et al. (2019); van Urk (2018); and especially for Slavic, by Progovac (1998); Franks (2013); Despić (2017); Stegovec

(2019); Caha (2021); and Ruda (2021a). The consensus is mostly that a pronoun consists of a base, followed by ϕ-features, topped off by case projections.

(5) nominal base > phi-features > D > case

Yet, while Déchaine and Wiltschko (2002) do not really consider case and specifics of morphological realization, Weerman and Evers-Vermeul (2002), Neeleman and Szendrői (2007), and Barbiers et al. (2009) do so to some extent, utilizing the proposed pronominal skeleton to account for other phenomena, such as pro-drop or pronominal copying.

Based on the discussion on the presence of the DP layer in the nominal structure of BCMS in §3.1 below, I will argue that the D-layer from the structure in (5) need not be assumed for BCMS. I will otherwise be following the accounts listed above in terms of the general idea that the pronominal structure encompasses a pronominal base, phi-features, and case; however, a novel contribution is going to consist in the definition of locality domains that these delimit within the nominal structure. Another departure from this general approach is going to involve the way in which features are represented; namely, I will assume that they have complex internal structure in the form of feature hierarchies in the sense of Harley and Ritter (2002). Finally, I will analyze the suppletion patterns in terms of contextual allomorphy in the sense outlined in §3.2 below, arguing that the locality domains defined by the syntactic projections that build the nominal phrase play a vital role in deriving the resulting patterns.

3.1. The NP/DP Debate

Across the Slavic family, similarities and differences between pronominal elements have been addressed most actively within the debate on whether nominal categories project a DP (Progovac 1998; Cardinaletti and Starke 1999; Despić 2011; Arsenijević 2014; Runić 2014; Puškar-Gallien 2019; Ruda 2021a; Bešlin 2023; and Jovović 2024). Arguments have been advanced in favor of nominals being strictly NPs (Bošković 2008; Runić 2014), strictly DPs (Arsenijević 2014, 2018a), or for a parametrized view under which pronouns involve more structure as they include a DP layer, while lexical nouns are NPs (Bešlin 2023). I will take an intermediate position: Pronominal elements in BCMS are essentially what Déchaine and Wiltschko (2002) term *PhiP*s.

Let me briefly elaborate on why many currently available diagnostics for the categorial status of nominals are at best inconclusive for BCMS. Déchaine and Wiltschko (2002) argue that pronouns come in three sizes: NP, PhiP, and DP. In order for a pronoun to qualify as a Pro-DP, it must fulfill a particular set of criteria. First, a pronoun needs to allow overt lexical material to appear next to it, as in *we linguists* in English, where the noun *linguists* is arguably the

overt realization of the NP, and *we* of D. Pronouns in BCMS allow for other NP material to co-occur with them (e.g., *mi lingvisti* 'we linguists'). By analogy with English, this overt material can be treated as the overt realization of the NP part. However, overt material can also appear in front of the pronoun in BCMS. Bošković (2008), Runić (2014), and Arsenijević (2017) note that pronouns in BCMS can be modified by adjectives and other modifiers such as demonstratives.[5]

(6) a. Dobri ti me retko zove/zoveš.
good.M.SG 2SG me rarely call.3SG/2SG
'The good you call(s) me rarely.' (Arsenijević 2014)

b. Ja volim onog tebe kojeg poznajem.
1SG love.1SG that.M.SG you who know.1SG
'I love that you that I know.'
(cf. pronouns in Russian; Pereltsvaig 2007: 28)

As observed by Arsenijević (2017), the mere fact that pronouns can be modified in BCMS and in English is not enough to claim that there is indeed a DP layer involved in the structure. Under most of the proposals that postulate a DP layer in the structure of the pronoun, it is assumed that pronouns are generated in N and move to D, collecting the inflectional morphology along the way from the Agr(eement) projections that stand in between, e.g., Progovac (1998) (see also Franks and Pereltsvaig 2004). Furthermore, BCMS pronouns allow an overt noun to follow them, and assuming that the overt noun also originates in N, the two should be mutually exclusive. Conversely, overt material such as adjectives and demonstratives can precede the pronoun, (6), which would also be unexpected if pronouns are located in D.[6] Finally, a potential confound related to examples such as (6a) was recognized by Arsenijević (2014), who argues that there are also semantic restrictions on the adjectives that can modify pronouns. In particular, he argues that only non-restrictive adjectives can combine with pronouns. Adjectives that are used restrictively can combine with pronouns only if the pronouns themselves semantically shift in interpretation, acquiring the interpretation of nouns (i.e., from <e> to <e, t>, as is evident in the different agreement possibilities that such a pronoun can show demonstrated in (6a)).

Furthermore, a Pro-DP behaves as an R-expression, while a Pro-PhiP behaves as a bound variable, and as such it may license sloppy readings. Clitics

[5] See Pereltsvaig (2007, esp. p. 28) for Russian and see also Arsenijević (2017, 2018a) and references therein for the claim that Polish and Slovak disallow modification of pronouns.

[6] See Despić (2011: 201–22) for a detailed discussion.

in BCMS are argued to allow for sloppy identity readings (see Runić 2014), which would qualify them as PhiPs. There are nevertheless contexts in which strong pronouns can also be interpreted as bound variables, thus counting as PhiPs, e.g., focus in (7b).[7]

(7) Clitics and strong pronouns as bound variables

a. Svaki predsednik$_i$ misli da ga$_i$/??njega$_i$ svi vole.
every president thinks that him.CLT/him everyone love
'Every president$_i$ thinks that everyone loves him$_i$.'

b. Svaki predsednik$_i$ misli da samo njega$_i$/*ga$_i$
every president thinks that only him/him.CLT
svi vole.
everyone love
'Every president$_i$ thinks that everyone loves only him$_i$.'
(Despić 2011: 243)

The final diagnostic that should distinguish Pro-DPs from Pro-PhiPs is their distribution in a clause: a DP cannot be used as a predicate but only as an argument, while a Pro-PhiP can be either an argument or a predicate. In BCMS, both strong pronouns and clitics can be used as arguments, (8). Strong pronouns can also function as predicates, (9), which would make them PhiPs (see Ruda 2021a for Polish). However, note that the very claim that DPs cannot function as predicates, put forward by Longobardi (1994) and followed by Déchaine and Wiltschko (2002), has been disputed in the literature (see, for instance, Pereltsvaig 2007: 21f. and references therein for Slavic).

(8) Video sam tebe/te.
seen AUX.1SG 2SG.ACC/CLT.2SG.ACC
'I saw you.'

(9) Postala sam ti.
become.PRT.F.SG AUX.1SG 2SG.NOM
'I became you.'

I thus conclude that even though the diagnostics for the structural size of a nominal phrase proposed by Déchaine and Wiltschko (2002) do not conclusively uncover the category of BCMS pronominal elements, the closest generalization that the tests above offer is that the presence of the DP category cannot be safely

[7] See also Ruda (2021b) and Stegovec (2019) for additional conflicting data from Polish and Slovenian.

confirmed from them. In the absence of positive evidence for the DP layer, I will assume that it is absent, thereby treating pronominal elements as PhiPs.[8]

3.2. Allomorphy

While Moskal (2015b); Smith et al. (2019); and van Urk (2018) do specifically focus on morphological realization of pronouns and their suppletion patterns, I argue that they (i) do not provide sufficient detail about the nature of the pronominal base, (ii) rely on both categorical and relative locality, and (iii) cannot account for gender and its place in the structure. Moskal (2015a, 2015b); Moskal and Smith (2016); Smith et al. (2019); and McFadden (2018) reduce suppletion to the allomorphy of the stem. Allomorphy itself is considered to result from quite a local process in the sense that only the features that are somehow close to a node may affect the spell-out of that node (see, e.g., Moskal and Smith 2016 and references therein). Consider the following example of allomorphy in the nominal domain. It has been observed that both number and case can trigger suppletion of the stem of a pronoun (cf. *ja* vs. *na-* below). In contrast, nouns can supplete only for number (*človek* 'man, human' vs. *ljudi* 'people' below), but case-driven suppletion of a nominal root is virtually unattested in nouns.[9]

Table 3. Suppletion in the nominal domain

N	*ja*	*mi*
G	*mene*	***nas***
A	*mene*	***nas***
D	*meni*	***nama***

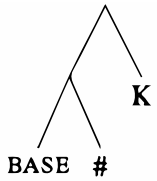
N	*človek*	***ljudi*** / **človeci*
G	*človeka*	***ljudi*** /**človeka*
A	*človeka*	***ljude***
D	*človeku*	***ljudima***

[8] An alternative way to approach this debate would be to apply the tests advanced by Cardinaletti and Starke (1999), who argue for a tripartite distinction between strong, weak, and clitic pronouns. Without going into further detail, I will note that their tests are also inconclusive and point the reader to Despić (2011, esp. 240f.); Ruda (2021a, 2021b); and Bešlin (2023) for detailed discussions on why this is the case.

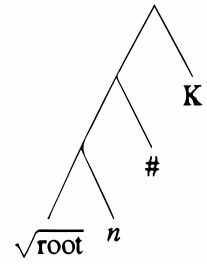
[9] Note that a challenge to this claim can be found in Slovenian. While the equivalent noun *človek* 'man' uses the equivalent suppletive stem *ljudje* in the plural, in the dual the non-suppletive stem is used in the nominative (*človeka* 'two men'). To make matters more complicated, the suppletive stem is, however, used in the dual in the genitive case (*ljudi*) and locative case (*ljudeh*), whereas in the other cases, the non-suppletive stem is used (*človeka*.ACC.DU, *človekoma*.DAT.DU, and *človekoma*.INST.DU). I thank the editors for pointing this out. For more information on the quirks of the Slovenian dual, see Marušić and Žaucer (2021).

In order to account for why this is the case, Moskal (2015a, 2015b) and Smith et al. (2019) assume the structural distinction in (10–11). Nouns (11) differ from pronouns (10) in having a lexical root and a nominal categorizing head *n*. They are both similar in having a number projection above their base (#) and the K(ase) head above it. The K-head of nouns is argued to be too far away to be able to affect the realization of the nominal root (11), while with pronouns it is sufficiently local to the pronominal base. This is implemented by the proposal that nouns include a locality boundary in their structure such that only number, but not case, can affect the realization of the root.[10] The #-head and K-head are sufficiently local to the *pronominal base* in order to be able to create a context for the insertion of its Vocabulary Item, while the realization of the *root of a noun* can only be affected by the #-head. Case is too far away.

(10) Pronouns (Smith et al. 2019):



(11) Nouns (Moskal 2015b):



The cause of the allomorphy is taken to be the cyclicity of the *n*-head. With nouns, *n* is a cyclic node, triggering the Vocabulary Insertion (VI) of the root. Since Vocabulary Insertion can be affected by the structure up to the next cyclic node and one node above that, the only other projection that is able to affect the VI of the nominal root is the first node above the cyclic node, i.e., the #-head. Pronouns, however, lack such a cyclic domain (having no root and no *n*), which makes both number and case local enough to be able to condition the realization of the pronominal base, which may result in both case- and number-driven suppletion. Furthermore, even though *n* will trigger insertion into its complement, the realization of the *n* node itself will be triggered by the next higher cyclic node, which is how McFadden (2018) models stem al-

[10] Here and below in my analysis, I will partially adopt the right-branching structures and notations from Moskal (2015a, 2015b); Moskal and Smith (2016); Smith et al. (2019); and McFadden (2018), which indicate the linear order of the realization of the individual morphemes. The type of operation responsible for the correct linear realization of the given morphemes (head-movement, roll-up movement, or morphological merger) is in no way crucial for the proposal below and, as such, will be left as a task for future work.

lomorphy in Tamil nouns (which only show allomorphy in nominative vs. non-nominative contexts, just like BCMS pronouns).

Within this line of thought, there is a notable lack of consensus on the nature of the pronominal base. Even though van Urk (2018) explicitly treats it as an *n*P, for Moskal (2015a) it is a D, while Smith et al. (2019) call it generally a root. For all of them, its exact nature is less important than the fact that this projection is the locus of person features. Even though he treats the pronominal base as *n*, van Urk (2018) has little to say about the parallelism between a pronoun and a noun, i.e., assuming that a noun would involve an additional lexical root, the *n* would still be able to bear person features, the consequences of which have not been further explored.

All of these proposals also assume some notion of dynamic determination of locality domains (Bobaljik and Wurmbrand 2005 and Bošković 2014). For instance Moskal (2015a, 2015b) focuses only on locality domains in the morphological realization of syntactic structures, arguing that each node is potentially cyclic, but whether or not it will become a cyclic node depends on the nodes that are introduced above it (or rather realized after it). Adopting a similar approach, van Urk (2018) and McFadden (2018) assume that the *n*-head is a categorical cyclic domain (so does Moskal), but in addition to that, the cyclicity of other nodes above it is still assumed to be dynamic. To this, McFadden also introduces an additional empty node in Tamil above the #-node, which is also argued to be cyclic. Thus, in this line of research, it seems that the determination of locality is not exactly unified, and as such still open to debate.

Finally, since these proposals focus on the interaction of person, number, and case, gender is largely left aside. Having seen above that in BCMS the realization of gender is also important for deriving the suppletion patterns of pronouns, the assumptions on its position and realization should be made more explicit. In my analysis below, I will adopt the basic premises of this strand of research in terms of how allomorphy functions. However, I will argue that the domains that determine the spell-out of syntactic nodes and affect morphological processes are definite, defined by specific syntactic heads (e.g., the *n*-head and the gender-feature-bearing head).

4. Disassembling Pronouns: The Internal Structure

This section introduces the idea that the complete structure of a (pro)noun in BCMS includes three general zones: nominal base > ϕ-features > case. I will argue that each of these contains additional structure within it. In particular, ϕ-features are going to further branch into person > number > gender, in that order, while case will include a sub-hierarchy of unmarked > dependent > oblique case. Finally, pronouns differ from nouns in lacking a lexical root, thus being a purely functional category.

4.1. Base

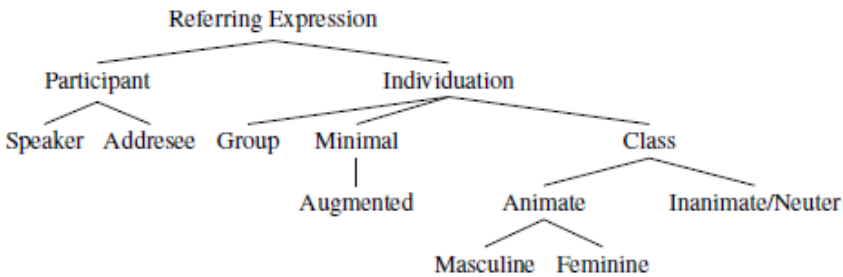
Following van Urk (2018), Déchaine and Wiltschko (2002), and van Koppen (2012), I take the base of the pronoun to be crucially nominal. Specifically, I will assume that the pronominal base is formed by the same nominalizing head *n* that builds nouns by categorizing a root (Marantz 2001, 2007; Arad 2003, 2005; Kramer 2015). The pronominal base will furthermore crucially differ from that of nouns in lacking a lexical root (see Moskal 2015a, 2015b; Smith et al. 2019). The pronominal *n*P thus consists solely of the categorizing head *n*.

4.2. ϕ-features

4.2.1. Feature Hierarchies

For the purposes of the formal representation of person, number, and gender, I will adopt the premises of Harley and Ritter's (2002) *feature geometry* approach, which views ϕ-features as consisting of hierarchically organized building blocks.[11]

(12) Structural hierarchy of ϕ-features (Harley and Ritter 2002: 486):



Accounts distributing these features across the nominal spine have mostly focused on two types of features, person and number, or number and gender (see Béjar and Řezáč 2009; van Koppen 2012; Puškar 2018; Puškar-Gallien 2019; and Caha 2021). I will offer a unified proposal for structural encoding of the hierarchy in (12) within the nominal phrase.

[11] Harley and Ritter (2002) draw an analogy with phonological features, arguing that the structured geometric representation of morphological features, modeled after that of the phonological ones, may help constrain pronoun and agreement systems. Preminger (2014) argues for a geometric representation of features in the syntax as a necessary alternative to the (un)interpretable features. See these works for more detail and further motivation.

Following Béjar and Řezáč (2009) (see also McGinnis 2005; Georgi 2012, 2013; Nevins 2007; Preminger 2014; Deal 2015; Kalin 2019), I assume that person features can be further decomposed such that the complexity of representation increases from the third towards the first person. Specifically, I assume that first person comprises the features [π, Participant, Speaker], (13), second person lacks the [Speaker] feature, (14), and third person is represented by the person [π] node alone, (15). An important property of these subfeatures is that they stand in an entailment relationship to each other, whereby having a [PRTCPT] or a [SPKR] node entails bearing the dominating [π] node as well.[12] Consequently, if the root node π is absent, the entire πP is absent from the structure, which will be the case with lexical nouns.

(13) $\begin{bmatrix} \pi \\ | \\ \text{PRTCPT} \\ | \\ \text{SPKR} \end{bmatrix}$ (14) $\begin{bmatrix} \pi \\ | \\ \text{PRTCPT} \end{bmatrix}$ (15) $[\pi]$

A similar manner of decomposition was applied to number by Harley and Ritter (2002), who represent its subparts by features such as [Group, Minimal, Augmented]. Since BCMS has a simple binary number system, for the purposes of developing an initial analysis, I will adopt the representation of the plural number proposed by Preminger (2014), as in (16). Singular will be treated as the absence of number (Nevins 2011 and Pesetsky 2013; see Despić 2017 for a claim that singular number is unmarked with respect to plural in Serbian). Technically, #P will be postulated only in case it specifies plural number, i.e., #P is not projected if the noun is singular (Kratzer 2007).

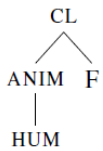
(16) $\begin{bmatrix} \# \\ | \\ \text{PL} \end{bmatrix}$

As the final member of the ϕ-set, I argue that gender features can receive a corresponding geometric treatment (see also Puškar 2018; Puškar-Gallien, forthcoming). Harley and Ritter (2002: 514) acknowledge that the internal structure and organization of gender would have to vary across languages, due to the great variation languages display in gender and class features in general. Adopting Harley and Ritter's intuition that gender features include animacy and humanness specification in their structure, I propose an adaptation

[12] Henceforth, the features will be presented by a variation of the following bracketed notations: [π[PRTCPT[SPKR]]].

of the hierarchy that will capture gender in BCMS. According to Willer-Gold et al. (2016) and Arsenijević (2018b), feminine is argued to be the most marked gender in BCMS, masculine being the semantically unmarked and neuter the syntactically unmarked one. Combining this with the feature geometry approach, I propose that gender in BCMS is represented in terms of a general gender node CLASS, a marked feminine value [F], and an animacy[13] and humanness specification, represented as [ANIM] and [HUM] nodes (see also Hammerly 2018; Foley and Toosarvandani 2019; Caha 2021; and Adamson and Anagnostopoulou 2024, 2025 for similar proposals for French, Zapotec, Czech, and Greek). My proposal for the hierarchy of gender, which is able to capture the distinction between natural and grammatical gender, is given in (17).

(17)



This approach to gender provides a direct link between gender and the features [ANIM] and [HUM] as subparts of its specification. The sub-hierarchy in (17) can be used to represent any of the three genders in BCMS and their nuances. For instance, nouns of feminine natural gender will involve all the available nodes in the hierarchy: [CL[ANIM[HUM]][F]], while grammatically feminine nouns will lack the animate and human specification, leaving them with [CL[F]]. Nouns of masculine grammatical gender will only involve the [CL] node, signaling that they carry an unmarked gender feature. Masculine natural gender will involve the [ANIM] and [HUM] features as well, accounting for the general bias in language under which the default referent of human nouns is male (see Arsenijević et al. 2022 for a justification of this claim based on experimental evidence). Finally, the absence of the [CL] node signals the absence of gender, thereby modeling neuter gender. Markedness of gender may thus be expressed in terms of the number of nodes it contains: feminine natural gender being the most marked one, grammatical masculine the least.

Differences between natural and grammatical gender thus fall out from their internal feature structure: they both involve the general [CL] node but

[13] Based on syncretism in inflectional paradigms and certain agreement properties, two *subgenders have been identified* in BCMS within the category of masculine: animate and inanimate (see, e.g., Corbett 1991 for an overview), which justifies positing [Animate] as a subfeature of gender. The difference in animacy among masculine nouns in BCMS leads to genitive-accusative syncretism in animate masculine nouns and nominative-accusative syncretism in inanimate nouns and to differences in agreement with nominal modifiers and relative pronouns.

differ in the rest of the feature inventory. Another difference between them will be in their syntactic distribution, as outlined in the following section.

4.2.2. Distribution of Features across the Nominal Spine

With respect to how morphosyntactically or semantically marked they are, ϕ-features have been argued to align according the implicational hierarchy given in (18), where the degree of markedness increases towards the right.[14]

(18) Implicational hierarchy of ϕ-features (Greenberg 1963; Noyer 1992):
Person > Number > Gender

I propose that (18), combined with Harley and Ritter's (2002) geometry in (12), translates into a hierarchy of syntactic projections, such that each feature type projects an independent XP. Person and number features have been argued to reside on two separate projections, such that person is lower than number (Moskal 2015b; Harbour 2016; Smith et al. 2019; and van Urk 2018). This is advocated particularly strongly by Harbour (2016), who argues that encoding person higher than number makes wrong predictions for possible and impossible pronoun inventories, both when it comes to their morphology and their interpretation.[15] Morphological evidence presented by Noyer (1992); Trommer (2002); Harbour (2007, 2008, 2016); and Arregi and Nevins (2012) indicates that, if pronouns can be morphologically decomposed into person, number, and case, number comes in between person (pronominal base) and case. Under a Mirror-Theoretic view of the interaction of syntax and

[14] Noyer (1992) argues for this hierarchy of ϕ-features based on Impoverishment patterns that morphemes in certain languages show. For instance, he proposes the following hierarchy for Arabic: 1 > 2 > PL > DUAL > F, based on which features get deleted first in the case of markedness accumulation (Noyer 1992: 46). Specifically, Arabic shows gender distinctions on pronouns and agreement affixes in local person in the second person but not in the first. Since it is presumably the feminine feature that gets deleted in the context of first person (if it were the other way round, first-person feminine would be syncretic with third-person feminine), Noyer's interpretation of this is that markedness filters, which determine what combinations of features Impoverishment rules will apply to, consider the features not on an individual basis but on the basis of their position in the markedness hierarchy.

[15] As the focus of this paper is the morphological encoding of person features, their semantics will largely be put aside. If semantics were to be taken into consideration, the property of *person* would quickly expose its further complexity. See Gruber (2013); Ackema and Neeleman (2013, 2018); and Harbour (2016) for different proposals. These argue that mapping between the morphological and semantic realization of person features is not always direct and, as such, it will be left for further research.

morphology (Baker 1985; Brody 2000; Brody and Szabolcsi 2003), this indicates a lower base position of person with respect to number.

Based on this, following recent proposals of Ruda (2021a) for Polish and Stegovec (2019) for Slovenian, I take person to head its own projection, πP, above the *n*P. I propose that number is then introduced by a further projection that I will label as #P.[16]

As grammatical gender forms a portmanteau with number and case, I assume it is local to these two features (for more on case, see the following section). Additional evidence for the separation and ordering of person and gender comes from Slovenian first-person plural pronouns, which have a feminine (*m-e* '1-F.PL.NOM') and a masculine (*m-i* '1-M.PL.NOM') version. As the final item on the scale in (18), I thus propose that grammatical gender is represented by a phrase above #P, the CLP. If this phrase includes only the CL node, the result is a masculine pronoun, but if the [F] sub-node is present, a feminine pronoun will result.

Recall that I argued that BCMS makes a distinction between natural and grammatical gender, based on the agreement patterns of local person pronouns, which do not show gender distinctions but nevertheless control gender agreement. Assuming that their lack of overt gender distinctions indicates a lack of grammatical gender, and assuming that natural gender is also represented syntactically in the same form of gender hierarchies as in (17), the question arises where natural gender is located. I argue that the locus of natural gender is the *n*P.

For Harley and Ritter (2002), a pronoun is essentially a "Referring Expression", as the root node of their tree suggests. Given that I take the *n*P to be the root node, I equate that with the projection responsible for referentiality as well as individuation of a (pro)noun. Individuation is necessary for reference taking and quantification (see, e.g., Sichel and Toosarvandani 2021), which differentiates nouns from other lexical categories (Baker 2003: 94–189). Under the Distributed Morphology assumptions, what additionally differentiates nouns from other lexical categories is the categorizing head *n* (as opposed to *v* or *a* for verbs and adjectives). Assuming that individuation is somehow connected to it, we should define how this property can be structurally represented.

Sichel and Toosarvandani (2021, 2024) take individuation to be introduced by a separate head σ. This head is a precondition for having person features, since their role is also inevitably connected to referentiality, as well as animacy and humanness. As argued above, properties such as human and animate do not necessarily depend on the kind of person that is present in a pronoun in BCMS, but they have more to do with the expression of natural gender. I therefore propose that individuation is basically an abstract prop-

[16] See, e.g., Ruda (2021a, 2021b) for arguments that PersP can be taken to introduce definite reference in languages without articles, such as Polish and BCMS.

erty of the *n*P and is tied to having the features [anim] and [hum]. Since these cannot appear without being connected to class in my system, I assume that a cl node can optionally be attached to the *n*P.[17] This results in the possibility of having natural gender on the *n*P.[18]

Both grammatical and natural gender thus represent types of gender available in the language. Both have geometric representation and can be targeted by Agree operations.[19] Grammatical gender is simpler and less marked by virtue of including fewer features. Finally, grammatical gender gets targeted by spell-out rules on nominal elements, while natural gender can only be spelled out as a result of agreement.

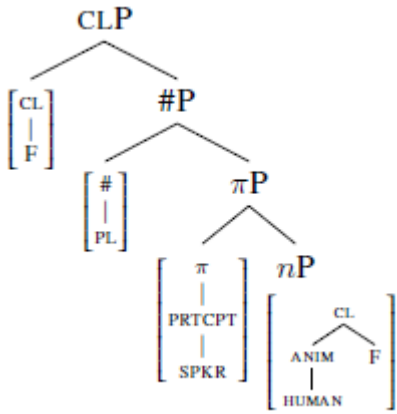
To sum up, (19) represents the complete structure of a pronoun in BCMS in the most complex case (1pl with a natural feminine gender). This structurally encodes the hierarchy from (18) above, with an additional benefit of providing a way to distribute the Harley and Ritter (2002) hierarchy across the pronominal spine (see van Koppen 2012 and Fassi Fehri 2000).

[17] Formally, under the assumption that grammatical gender as a head in the syntactic structure projects a gender phrase (clP), natural gender can be assumed to be an adjunct to the *n*P.

[18] Gender as a category can be dispersed across the nominal spine (Steriopolo and Wiltschko 2010; Pesetsky 2013; Landau 2016; Kučerová 2018; Steriopolo 2018a, 2018b; Fassi Fehri 2018; and Puškar 2018; but see Arsenijević 2021 for an alternative view). Here I follow Puškar (2018) and the argumentation therein for a low position of natural gender and depart from Kramer (2015), who places both natural and grammatical gender on *n*.

[19] Since the process of agreement is not the main focus of the paper, I will note that I assume that agreement is carried out by the operation Agree (Chomsky 2001) in the standard Minimalist terms, where an unvalued Probe searches for a valued Goal to satisfy its missing features. An additional assumption that I make, following Béjar and Řezáč (2009), is that ϕ-features can be probed for separately, to which I add the proposal that the Probe for gender can be parametrized such that it does not only look for gender features in general, but for gender features of a particular kind: natural gender (in which case the Probe will search for both gender and animacy and humanness), or grammatical gender (in which case it will not look for features [anim] and [hum], but only for [cl]). As a consequence, the Probe will be able to agree with both types of gender proposed in (17). For pronouns, this will have the consequence that the Probe will be able to agree both in the natural gender of local person pronouns and with the grammatical gender of third-person pronouns. Further details of this approach and its consequences for patterns of hybrid agreement are explored in Puškar (2018) and Puškar-Gallien (2019).

(19)



Finally, the absence of a root-level feature (π, #, CL) is going to result in the absence of a feature-bearing phrase. As a consequence, singular number will be treated as the absence of number. Technically, #P will be postulated only in case it specifies plural number. Similarly, the absence of gender will be encoded as the absence of CLP, yielding neuter gender. Missing features will be realized by means of default exponents, as further elaborated in §5.[20]

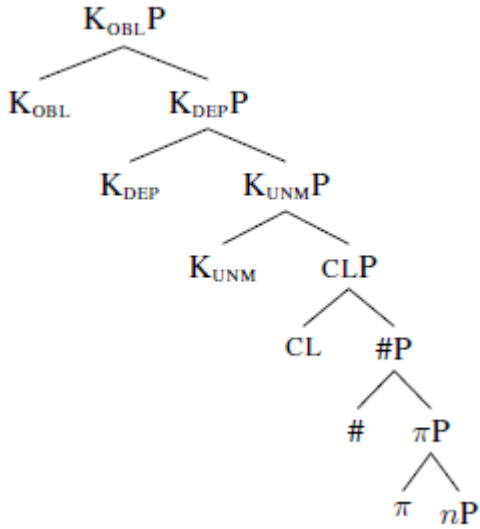
4.3. Case

Following Bittner and Hale (1996), Caha (2009), Neeleman and Szendrői (2007), Moskal (2015a, 2015b), and Smith et al. (2019), I assume that Case is introduced by a separate projection K(P), on top of the ϕ-feature bearing projections. K can have a complex structure that encodes Caha's (2009) *Case Hierarchy*: NOMINATIVE > ACCUSATIVE > GENITIVE > DATIVE > INSTRUMENTAL > COMMITATIVE. Smith et al. (2019) collapse this into a distinction between the *dependent case* (DEP, here encompassing ACC and GEN) and the *oblique case* (OBL,

[20] A reviewer wonders about feature co-occurrence restrictions, e.g., why person and grammatical gender do not co-occur. I envisage two possibilities. It may be assumed that the universal structure proposed in (19) is generally available, but not all languages will make use of all possibilities. For instance, while BCMS does not show gender distinctions on local person, and presumably lacks CLP with local person, Slovenian does contain this phrase and consequently distinguishes between masculine and feminine first person. Alternatively, we may assume grammatical gender to be universally present, but that gender gets deleted under Impoverishment in local person contexts, as suggested by Noyer (1992) for Arabic (see fn. 11 above). The latter option would have the benefit of accounting for the nominative vs. non-nominative suppletion on local person pronouns in the same way as third-person pronouns are accounted for in §5.2 below.

here DAT).[21] Nominative is modeled as the absence of case by Andrews (1982); Falk (1991); Bittner and Hale (1996); Taraldsen (1996); Neeleman and Weerman (1999); de Hoop and Malchukov (2008); McFadden and Sundaresan (2009); and Kornfilt and Preminger (2015).[22] I will follow McFadden's (2018) implementation, where nominative amounts to the absence of the case-bearing projection(s). This eliminates K_{UNM} from (19).

(20)



4.4. Interim Summary: Disassembled Pronouns

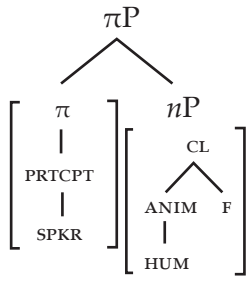
The complete structure of a pronoun given in (20) offers possibilities for parametrization, as not all pronouns will include all of the available nodes. Local person pronouns lack CLP in general, which models the lack of grammatical gender. Their singular forms also lack #P. The πP is projected, since they must have at the minimum the [PRTCPT] feature. The structures in (21–22) represent the first-person pronouns in the nominative case (hence the

[21] I will exclude instrumental and locative for the purposes of the current discussion. Locative is always syncretic with dative in BCMS, with the difference that it must be preceded by a preposition, hence it can be analyzed simply as a PP. An argument was advanced for instrumental by Milićev and Bešlin (2019). See Puškar-Gallien (forthcoming) for further consequences of this idea, among others that the PP introduces another locality boundary.

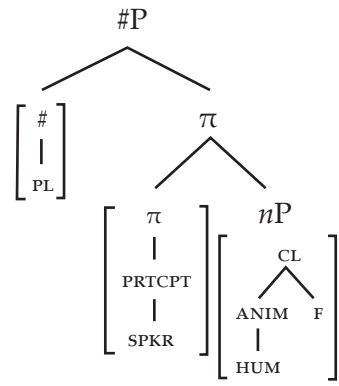
[22] Thanks to an anonymous reviewer for drawing my attention to this work.

lack of KP). The second person will differ from first person in lacking the feature [SPKR].[23]

(21) Singular 1st-person pronoun:



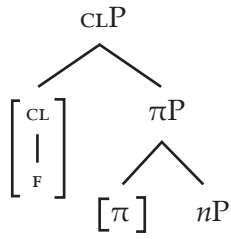
(22) Plural 1st-person pronoun:



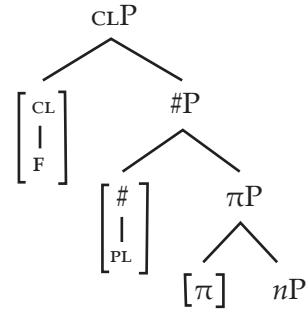
The proposed structures for third-person pronouns are presented in (23–24). As number is absent, in the singular their *n*P will be dominated by πP and CLP, which bears the [F] node for grammatically feminine nouns or just the [CL] node for masculine ones. In the plural, the CLP will be projected above the #P. The combination of these two phrases will define the inflectional affixes of the pronouns. The *n*P lacks features if the pronoun denotes an inanimate entity. With an animate (or human) referent, the *n*P will bear natural gender and number in the same manner it does with local person pronouns.

[23] A reviewer wonders about the relationship between features, i.e., how the model accounts for the plurality of local person, where first person plural is not exactly a plurality of persons, but rather a group formed around the speaker. One way of dealing with this would be to adopt Harley and Ritter's representation of number as including the features Minimal and Group, instead of simply # and PL. This would enable us to indicate the difference between a single referent (Minimal) and multiple referents (Group), and addition of other features such as Augmented would enable representing other options for quantification, such as dual, mass, collective, etc. Nothing in the account would change if this more precise denotation were accepted; however, the features # and PL were chosen for the sake of simplicity in the representation of the basic patterns.

(23) Singular 3rd-person pronoun:



(24) Plural 3rd-person pronoun:



5. Reassembling Pronouns: Morphological Realization

Adopting the general Distributed Morphology premise that syntactic nodes are realized postsyntactically by corresponding Vocabulary Items in the process of Vocabulary Insertion, two additional sets of assumptions necessary for a proposal on the morphological realization of pronouns must be briefly introduced, namely locality considerations and conditions on suppletion. As far as locality domains in the nominal phrase are concerned, I assume that the categorizing head *n* is a phase-head (Marantz 2001, 2007; Embick and Marantz 2008; and Embick 2010, 2021). I propose that an additional locality domain in the nominal phrase in Slavic is defined by the clP, as the final phrase that demarcates the ϕ-domain. These two phrases will trigger the spell-out of their complements. Otherwise, the derivation proceeds until all the numeration is spent and only then is the structure spelled out. As for suppletion, I will largely rely on the analysis of contextual allomorphy proposed by Moskal (2015a, 2015b), Moskal and Smith (2016), Smith et al. (2019), and McFadden (2018), who argue that it is reducible to the allomorphy of the stem. I will follow them in assuming that outward-sensitive allomorphy can only be triggered by nodes up until the next cyclic node and one node beyond that. I depart from them in assuming that cyclic nodes are phase nodes, thus eliminating the need for postulating separate syntactic and morphological locality domains. Inward-sensitive allomorphy will also play a role in the realization of case features (see Gribanova and Harizanov 2017 for the directionality of grammatically-conditioned allomorphy).

To sum up, the internal functional spine of pronouns involves three locality domains, defined by the nominalizing head *n* and the grammatical-gender introducing head cl. These also delineate the three domains of the nominal phrase: the lexical domain, ϕ-feature domain, and case domain. In the remainder of this section, we will examine how the morphology deals

with the output of the syntax in realizing the structures proposed above. As a reminder, Table 4 repeats the pronominal paradigms.

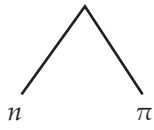
Table 4. Personal pronouns in BCMS

	1SG	2SG	1PL	2PL	3SG.M/N	3SG.F	3PL
NOM	*ja*	*ti*	*mi*	*vi*	*on-Ø/-o*	*on-a*	*on-i/-e/-a*
GEN	*m-en-e*	*t-eb-e*	*na-s*	*va-s*	*nje-ga*	*nj-e*	*nj-ih*
DAT	*m-en-i*	*t-eb-i*	*na-ma*	*va-ma*	*nje-mu*	*nj-oj*	*nj-ima*
ACC	*m-en-e*	*t-eb-e*	*na-s*	*va-s*	*nje-ga*	*nj-u*	*nj-ih*
INST	*m-n-om*	*t-ob-om*	*na-ma*	*va-ma*	*nj-im*	*nj-om*	*nj-ima*
LOC	*m-en-i*	*t-eb-i*	*na-ma*	*va-ma*	*nje-mu*	*nj-oj*	*nj-ima*

5.1. Spelling out Local Person Pronouns

As local person pronouns in their strong form carry pitch accent, I will assume that this is due to a lexical high tone (H), carried by their *n*, resulting in a falling accent (see Talić 2018 and references therein). This *n* is otherwise realized by a phonologically null exponent, unless in non-nominative case, where it is spelled out as the support morpheme *-en-/-eb-*.[24] As a result, singular nominative local person pronouns will be realized as in (25–28). Their *n*-node will carry the high tone, while the person features' realization depends on their internal complexity. Plural is realized by its exponent *-i*, and it triggers stem allomorphy.

(25) 1st SG NOM



[PRTCPT, SPKR]
ja

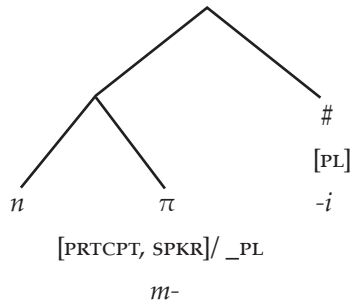
(26) a. [π, PRTCPT, SPKR] ⇔ *ja*

b. [π, PRTCPT] ⇔ *t*(*i*)

c. *n* ⇔ H

[24] I will also include the assumption that the *n* of local person pronouns must carry [HUM] and [ANIM] features. This can be thought of as an interface requirement, i.e., the derivations in which the *n* lacks these features will be filtered out at the interfaces by corresponding licensing conditions, in the sense of Kramer (2015).

(27) 1st pl nom



(28) a. [π, prtcpt, spkr] ⇔ *m-* / __X][25]

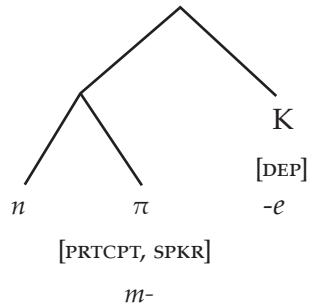
b. [π, prtcpt] ⇔ *v-* / __ #

c. [#] ⇔ *-i*

d. *n* ⇔ H

The presence of case other than nominative is implemented as the presence of the K-head above #, which can in turn affect the realization of the nodes below it. Example (29) illustrates a first-person pronoun in acc/gen (dependent) case in the singular. Recall that the second-person pronoun's base does not supplete for case. Its base will thus be pronounced by the exponent *t-* presented in (26b), to which the case suffix *-e* will be attached.

(29) 1st sg acc/gen



(30) a. [π, prtcpt, spkr] ⇔ *m-* / __X]

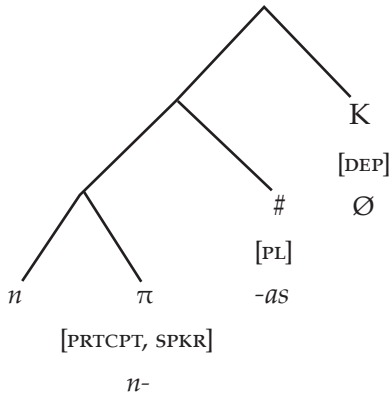
b. [dep] ⇔ *-e*

c. [obl] ⇔ *-i*

Last but not least, if the # head is present above π, both will be realized in the same cycle as case. I assume that K-projections trigger allomorphy on the number node, as in (32).

[25] The context for the first-person base allomorph *m-* is presented as a generalized "X]" in order to capture the idea that the allomorphy of this exponent is simply triggered by any XP above Pers, be it number or case (see McFadden 2018 for similar cases in Tamil). Indeed, the same allomorph is found in the context of dependent and oblique case in the singular (see (30a) and Table 1).

(31) 1st PL ACC/GEN



(32) a. [π, PRTCPT, SPKR] ⇔ *n-* / __ #, DEP

b. [#] ⇔ *-as* / __ DEP

c. [#] ⇔ *-am* / __ OBL

d. [K_{DEP}] ⇔ Ø / __ #

To sum up, (33–36) present the full list of VI rules for local person pronouns:

(33) a. *n* ⇔ H

b. *n* ⇔ *-en-* + H / __ SPKR, DEP

c. *n* ⇔ *-eb-* + H / __ PRTCPT, DEP

(34) a. [π, PRTCPT, SPKR] ⇔ *ja*

b. [π, PRTCPT, SPKR] ⇔ *m-* / __X]

c. [π, PRTCPT, SPKR] ⇔ *n-* / __ #, DEP

d. [π, PRTCPT] ⇔ *t(i)*

e. [π, PRTCPT] ⇔ *v-* / __ #

(35) a. [#] ⇔ *-i*

b. [#] ⇔ *-as* / __ DEP

c. [#] ⇔ *-am* / __ OBL

(36) a. [DEP] ⇔ *-e*

b. [OBL] ⇔ *-i*

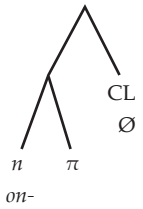
c. [DEP] ⇔ -Ø / __ #

d. [OBL] ⇔ -Ø / __ #

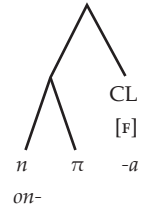
5.2. Third-Person Pronouns and Stem Suppletion

Recall that third-person pronouns include a gender phrase above the #P. Another difference from local person pronouns is the overt exponent of *n*, realized as the base *on-*. As a result, I postulate two different nominalizers, one deriving local person and the other one deriving third-person pronouns, which will be subject to corresponding licensing conditions at the interfaces, following Kramer (2015).[26] Focusing for now only on masculine and feminine pronouns, let us consider their postsyntactic content, represented in (37) and (38). In addition to their *n*, CL is also a cyclic node. It will be realized according to the gender features it bears, cf. (37), where it only contains the gender node of the hierarchy in (17), and (38), which contains an additional [F] node.

(37) 3rd M SG NOM:



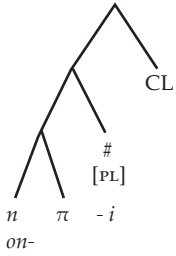
(38) 3rd F SG NOM:



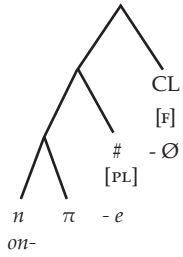
As for the plural, the base will retain its form, while the plural exponent will be affected by the presence of the [F] gender. I will assume that the markedness constraints active in the language delete the [F] gender feature in the presence of person and number, yielding a null feminine plural exponent (see Despić 2017).

[26] I assume that the nominalizer for third person can only be licensed under the presence of the CL-node (see Kramer 2015 for more detail on licensing conditions on nominalizers).

(39) 3rd M PL NOM:

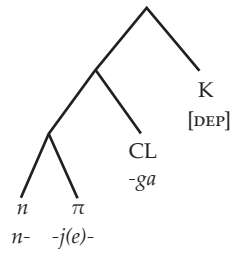


(40) 3rd F PL NOM:

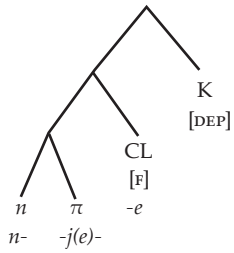


Adding the K head in order to introduce case features places this head into a position from which it can trigger stem allomorphy. Since CL is a cyclic node, the first node above is still available for morphological operations. It can thus create a context for the suppletion of the *n*-base. As a result, only K_{DEP} can affect its spell-out, but not any head above it. Thus, the base *n* will be realized as the allomorph *n-* in the presence of the first case-bearing head, K_{DEP}. Due to containment, if any additional case projection is present, it will inevitably require the presence of K_{DEP}, hence the context for allomorphy will always be available, but any other head would be too far away from *n*. This is why we only have the nominative vs. non-nominative allomorphy of the third-person pronominal base (similar to Tamil nouns investigated by McFadden 2018).[27]

(41) 3rd M SG ACC/GEN

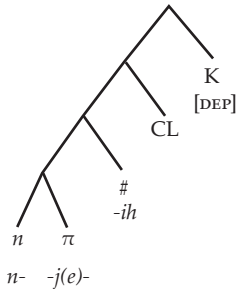


(42) 3rd F SG ACC/GEN

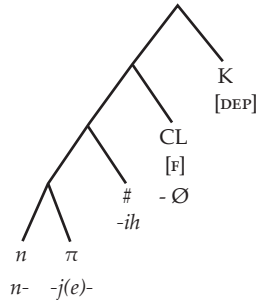


[27] The realization of K features in (41–42) will be left for further investigation. In principle, CL can be assumed to cause the null spell-out of the case projections and thereby block their realizations, as in (36).

(43) 3rd m pl acc/gen



(44) 3rd f pl acc/gen



To sum up the discussion thus far, examples (45–49) present a full list of exponents that realize third-person pronouns in BCMS.

(45) a. *n* ⇔ *on-*

b. *n* ⇔ *n-* / __ dep

(46) a. [π] ⇔ Ø

b. [π] ⇔ *-j* / __ dep

(47) a. [cl] ⇔ Ø

b. [cl] ⇔ *-ga* / __ dep

c. [cl] ⇔ *-mu* / __ obl

(48) a. [f] ⇔ *-a*

b. [f] ⇔ *-e* / __ dep

c. [f] ⇔ *-oj* / __ obl

d. [f] ⇔ *-Ø* / __ π, #

(49) a. [#] ⇔ *-i*

b. [#] ⇔ *-e* / __ f

c. [#] ⇔ *-ih* / __ dep

6. Discussion and Broader Implications of the Analysis

6.1. The Realization of Clitics

Recall from Table 2 that third-person clitics are simply the spell-out of the inflectional information without the base, while local person clitics spell out person, number, and case without the support morphemes found in strong pronouns. Under the current account, the realization of third-person clitics would amount to the spell-out of the π # CL K sequence without the *n*P, which is governed by the VI rules in (46–49). Similarly, local person clitics can be thought of as realized by the rules in (34–36), without the support morpheme, which is assumed to realize the base *n*.

Recall that third-person pronouns have adjectival endings (also present on other pronominal modifiers such as possessives and demonstratives). The realization of ϕ- and case features as clitics has the additional benefit of making the exponents more universally applicable in realizing agreement morphology on nominal modifiers. The only difference between them may be that nominal agreement morphology does not involve person agreement (Baker 2008). And omitting *n* in pronouns leaves us with a π-projection intact. Feminine clitics provide a window into how the spell-out of clitics and agreement affixes may be differentiated. Feminine clitics *je*.CL.F.SG.GEN, *joj*.CL.F.SG.DAT, and *ju*.CL.F.SG.ACC contain an additional *-j-* that adjectival agreement affixes lack (*-e*.F.SG.GEN/*-oj*.F.SG.DAT/*-u*.F.SG.ACC), and it is precisely this morpheme that is argued to be the realization of the π-node in the context of case, (46).[28]

Apart from the immediate morphological consequences that the non-realization of the *n*P base has, its absence also accounts for further differences between strong pronouns and clitics, namely their animacy restrictions. Specifically, strong pronouns must refer to animate/human entities, whereas clitics allow inanimate referents, (50).

(50) Clitics vs. pronouns, animacy/humanness (Despić 2011: 240)

a. Čuo sam **je**.
heard.M.SG AUX.1SG **CLT.3.F.SG.ACC**
'I heard her.' [+HUM] [-HUM]

b. Čuo sam **nju**.
heard.M.SG AUX.1SG **3.F.SG.ACC**
'I heard her.' [+HUM] *?[-HUM]

[28] I thank an anonymous reviewer for this insight.

The structures proposed above provide a handle on (50) by locating animacy and humanness features on the *n*P, as a part of natural gender. Specifically, since clitics lack animacy determination, by virtue of lacking the *n*P, they are compatible with any interpretation, i.e., there is no strict requirement that they be animate.

The outstanding question is what ensures the non-pronunciation of the *n*P base. A possible solution may employ *n*P deletion, as applied by van Urk (2018), based on a peculiar pronoun copying pattern in Dinka Bor (Nilotic), represented in (51). Both examples involve an overt copy of a fronted object pronoun, realized as the 3PL *kêek*. This pronoun thus matches the fronted pronoun only partially—in number, but not in person.

(51) a. wɔ̂ɔk ć̤i bôl kêek/*wɔ̂ɔk tî̤iŋ.
1PL PRF.OV Bol.GEN 3PL/1PL see.INF
'Us, Bol has seen.'

b. wêek ć̤i bôl kêek/*wêek tî̤iŋ.
2PL PRF.OV Bol.GEN 3PL/2PL see.INF
'You all, Bol has seen.' [Dinka Bor] (van Urk 2018: 940)

Van Urk (2018) argues that the sentences in (51) involve pronoun copying followed by partial deletion of the material in the lower copy. Whether a complete or an incomplete copy will be pronounced depends on whether a deletion operation has taken place within the DP.[29] Such operations may delete parts of the DP, with the restriction that the deleted bits be phasal units, leaving the remnant to be spelled out. In Dinka, he argues that *n*P is a phase, which makes it eligible for deletion. Recall that in his account, the *n*P also carries person features. If *n*P undergoes deletion, only the higher phrases in the DP are eligible for spell-out, resulting in the realization of a pronoun that only carries [pl] number, and no person (i.e., a 3PL pronoun). This can be straightforwardly extended to my account, due to the same treatment of pronominal *n*P as a phase. With third-person pronouns, the *n*P undergoes deletion before Transfer, thus leaving the $\pi > \# > \gamma >$ K phrases to be spelled out. The copy deletion analysis of van Urk may further capture the realization of resumptive pronouns as in (52). Assuming that resumption involves DP copying and the overt spell-out of the lower copy, the clitic *ga* in (52) may result from the same process as the copied pronouns in Dinka Bor; DP copying followed by copy deletion. The latter would delete the *n*P (and the root of the noun in its

[29] Van Urk (2018) tentatively proposes that such deletion operations may be implemented as non-Transfer, whereby Transfer would apply to an entire phase, not just the phase domain.

complement), leaving the $\pi > \# > \gamma > K$ sequence to be spelled out as a resumptive clitic.[30]

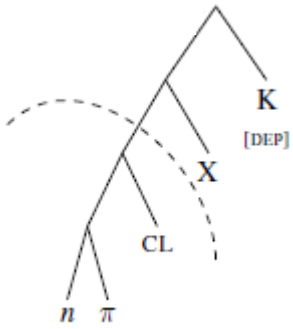
(52) **čovek** što sam **ga** upoznao
man.M.SG that AUX.1SG CLT.M.SG met.M.SG
prošle godine
last year
'a man whom I met last year'

In conclusion, clitics and strong pronouns start out as same abstract syntactic structures, differing only in their spell-out, which does not include the *n*P with clitics.

6.2. Additional Evidence: Demonstratives

The proposal above makes the prediction that the suppletion of third-person pronouns should be impossible if something blocked the K_{DEP} from creating a context for it, as illustrated in (53). In this section, I argue that demonstratives present exactly the case in question.

(53) X blocking allomorphy



BCMS has three types of demonstratives, which I will classify as proximal to the speaker (*ovaj*), proximal to the communication situation (*taj*), and distal (*onaj*), following Arsenijević (2018a). Abstracting away from their further spatial, temporal, discourse organizational, and epistemic interpretation (see Arsenijević 2018a for a detailed overview), for our purposes it suffices to note that the reference of the proximal demonstrative *ovaj* is connected to speaker-oriented deixis. The proximal demonstrative *taj*'s interpretation includes the

[30] As nouns are assumed not to involve a πP, additional analytical steps would have to be invoked to ensure that the resumptive pronoun does contain one.

proximity to the hearer, i.e., "the referent is present in the physical context of the communication and spatially proximal to the hearer" (Arsenijević 2018a: 166). Finally, the distal demonstrative includes the pronominal base of the third-person pronouns, *on-*, and is compatible with uses in which the referent is distal to the speech situation. Table 5 provides an overview of the morphological paradigms of demonstratives in BCMS.

Table 5. BCMS *on*-demonstratives

	SG			PL
	M	F	N	M/F/N
NOM	*ov-/t-/on-aj*	*ov-/t-/on-a*	*ov-/t-/on-o*	*ov-/t-/on-i/-e/-a*
ACC	*ov-/t-/on-og(a)*	*ov-/t-/on-u*	*ov-/t-/on-og*	*ov-/t-/on-e*
GEN	*ov-/t-/on-og(a)*	*ov-/t-/on-e*	*ov-/t-/on-og*	*ov-/t-/on-ih*
DAT	*ov-/t-/on-om(e)*	*ov-/t-/on-oj*	*ov-/t-/on-om*	*ov-/t-/on-ima*
INST	*ov-/t-/on-im*	*ov-/t-/on-om*	*ov-/t-/on-im*	*ov-/t-/on-ima*

What can be observed from Table 5 is that the inflectional endings of demonstratives are the same as those of third-person pronouns. The difference is in the nominative suffix *-aj*, instead of -Ø, which also carries an additional vowel length, both in the nominative and other cases. Given that demonstratives include the same set of inflectional endings as third-person pronouns, this indicates a shared internal structure. In fact, the distal demonstrative even shares the same base, with only the difference in the masculine nominative ending and vowel length on the final syllable (present in feminine and neuter as well). It is important to note that the stem in this case does not undergo suppletion, which would indicate that the phrase otherwise responsible for suppletion (assumed K_{DEP}) is either absent, or is present but too far away. I will argue for the latter option.

I assume that demonstratives are built the same way as pronouns, with the *n* π # CL K sequence, including an additional deictic phrase layer between the CLP and KP. In order to account for the internal structure of this deictic layer, I will borrow an assumption from Wiland (2018) (building on Lander and Haegeman 2016), who proposes the following hierarchy:

(54) $[_{KP}$ K $[_{DistP}$ Dist $[_{MedP}$ Med $[_{ProxP}$ Prox $[_{PersonP}$ Person $[_{NP}$ N]]]]]]

Taking the tripartite distinction from above, I assume that proximal demonstratives *ov-* and *t-* include the ProxP in their structure, which is then followed by DistP, which yields a distal demonstrative *on-*. This yields the functional sequence in (55).

(55) $[_{KP}$ K $[_{DistP}$ Dist $[_{ProxP}$ Prox $[_{clP}$ CL $[_{\#P}$ # $[_{\pi P}$ π $[_{nP}$ *n*]]]]]]]

The analysis introduced above predicts particular consequences for the structure in (55). Since K_{DEP} is no longer the first phrase above the cyclic node CL, it should not be able to trigger the allomorphy of the pronominal stem. And in fact, this is what we observe, especially with distal demonstratives (which retain the base *on-* throughout).

Furthermore, due to containment of deictic projections (see Wiland 2018), we do not expect that DistP will affect the realization of the *n*-head either. A Prox-head must be present in order for the Dist one to be introduced, hence an intervening head will always be there, over which Dist will be too far away from the *n*P base (not the first head after the cyclic node). Case is introduced above this head and may still affect the spell-out of the inflectional endings, hence yielding the same set of exponents, but crucially leaving the base intact.

This analysis would still require some revision in order to account for Arsenijević's (2018a) claims that the particular properties of demonstratives can be used as proof of the existence of the D-layer in their structure, which he then uses to account for their behavior. I will leave this issue for further research. To that, it is also claimed that proximal *ov-* demonstratives include the representation of first person in their structure, while the hearer-proximal demonstratives include the representation of the hearer, also noticeable in the stem syncretic to that of second-person pronouns (*t-*). One way to deal with this would be to assume that proximal demonstratives are built on the local person nominalizer, while distal demonstratives are built on the third-person nominalizer. Or alternatively, that proximal pronouns also include the [SPKR] and [PRTCPT] features in their representation, which would then require stricter licensing conditions on their realization. One could also do away with this by following Gruber (2013) in assuming that there is no strict mapping between spacio-temporal dimensions of person and its morphological representation. These issues will be left for further research.

7. Summary and Conclusion

This paper has proposed a decompositional model of the internal structure of pronominal categories in BCMS. It was argued that a pronoun includes three internal zones: the *n*P base, the ϕ-features, and case. The base is argued to consist only of the *n*-head, which distinguishes pronouns from nouns, which, in addition to this head, include a category-free root. Phi-features are argued

to have a hierarchical internal structure, where they stand in an entailment relation to each other, such that the basic node (π, #, CL) projects the corresponding syntactic phrase, and the sub-nodes define the kind of person, number, or gender that a pronoun bears. Finally, case is represented by means of a case hierarchy, in which nominative is absent, and dependent case is entailed in oblique case.

The proposed hierarchical structure of the nominal phrase offers a possibility to model markedness based on the feature inventory and their structural organization. This is particularly useful in the representation of gender, where it was argued that the most marked gender in BCMS (feminine natural gender) also has the most complex internal structure, involving the most nodes in the gender hierarchy. Conversely, a default feature may be presented either as the root node only (e.g., π for third person or CL for masculine gender), or as an absence of features (absence of π resulting in lexical nouns or absence of CL resulting in neuter gender).

Finally, I have argued against dynamic phase determination in the internal structure of nominal categories by arguing that the *n*P and CLP are phases. This was shown to have particular consequences on allomorphy patterns, as well as the realization of clitics. In particular, I argued that the phasehood of the CL node accounts for the suppletion pattern under which suppletion is sensitive only to the nominative/non-nominative distinction, but not to any case beyond nominative. This was argued to follow from the entailment of case projections: Only the dependent-case-bearing projection is close enough to the pronominal stem to be able to trigger allomorphy; all the other case projections (which also entail K_{DEP}) are too far away. The phasehood of the *n* head has another consequence, namely this head can undergo deletion and thereby leave the ϕ- and case projections as remnants to be spelled out as a clitic.

The analysis has provided a unified structure for all pronominal elements, i.e., strong pronouns and clitics, deriving the structure of nouns as its direct consequence. I have shown how the analysis may extend to other pronominal elements, such as demonstrative pronouns, which opens an avenue for further research and inspection of other categories such as possessives (which would require more elaboration in order to capture two sets of gender/number affixes). In addition to providing a unified structure of the nominal phrase, the proposal thus has the benefit of providing a means to account for the morphological, syntactic, and referential behavior of nominals, under a single analysis.

References

Ackema, Peter, and Ad Neeleman. (2013) "Person features and syncretism". *Natural language and linguistic theory* 31: 901–50.

————. (2018) *Features of person*. Linguistic Inquiry Monographs. Cambridge, MA: MIT Press.

Adamson, Luke J., and Elena Anagnostopoulou. (2024) "Interpretability and gender features in coordination: Evidence from Greek". Robert Autry, Gabriela de la Cruz, Luis A. Irizarry Figueroa, Kristina Mihajlovic, Tianyi Ni, Ryan Smith, and Heidi Harley, eds. *Proceedings of the 39th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 11–20.

————. (2025) "Gender features and coordination resolution in Greek and other three-gendered languages: Implications for the crosslinguistic representation of gender". *Linguistic inquiry* (2025): 1–50. DOI 10.1162/ling_a_00543.

Andrews, Avery D. (1982) "The representation of case in modern Icelandic". Joan W. Bresnan, ed. *The mental representation of grammatical relations*. Cambridge, MA: MIT Press, 427–503.

Arad, Maya. (2003) "Locality constraints on the interpretations of roots: The case of Hebrew denominal verbs". *Natural language and linguistic theory* 4(21): 737–79.

————. (2005) *Roots and patterns: Hebrew morpho-syntax*. Dordrecht: Springer.

Arregi, Karlos, and Andrew Nevins. (2012) *Morphotactics: Basque auxiliaries and the structure of spellout*. Dordrecht: Springer.

Arsenijević, Boban. (2014) "The agreement of adjectives with the honorific pronoun in Serbo-Croatian". *Philologia mediana* 6(6): 39–50.

————. (2017) "What the prosody of Serbian short and long form adjectives tells us about the nominal structure". Gordana Djigić and Sofija Miloradović, eds. *Godišnjak za srpski jezik* [Yearbook for Serbian language]. Niš, Serbia: Faculty of Philosophy, University of Niš, 9–21.

————. (2018a) "Atypical demonstratives in an articleless language". Marco Coniglio, Andrew Murphy, Eva Schlachter, and Tonjes Veenstra, eds. *Atypical demonstratives: Syntax, semantics and pragmatics*. Berlin: De Gruyter. [Linguistische Arbeiten, vol. 568]

————. (2018b) "Gender, like classifiers, specifies the type of partition: Evidence from Serbo-Croatian". *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 52: 21–37.

————. (2021) "'No gender in 'gender agreement': On declension classes and gender in Serbo-Croatian". *Balcania et Slavia* 1(1): 11–46.

Arsenijević, Boban, Ivana Mitić, and Zorica Puškar-Gallien. (2022) "*Judge* vs. *fool*: Restrictive use of a noun predicts hybrid agreement". Slides from a talk presented at the Formal Approaches to Slavic Linguistics 31, McMaster University, Hamilton, Ontario, Canada, 24 June 2022.

Baker, Mark C. (1985) "The mirror principle and morphosyntactic explanation". *Linguistic inquiry* 16(3): 373–415.

————. (2003) *Lexical categories: Verbs, nouns, and adjectives*. Cambridge, UK: Cambridge University Press.

————. (2008) *The syntax of Agreement and Concord*. Cambridge, UK: Cambridge University Press.

Barbiers, Sjef, Olaf Koeneman, and Marika Lekakou. (2009) "Syntactic doubling and the structure of *wh*-chains". *Journal of linguistics* 46(1): 1–46.

Béjar, Susana, and Milan Řezáč. (2009) "Cyclic agree". *Linguistic inquiry* 40(1): 35–73.

Bešlin, Maša. (2023) "DP in a model NP language: Evidence from Serbo-Croatian personal pronouns". Noga Syon and Nir Segal, eds. *Proceedings of IATL 2021.* Cambridge, MA: MIT Working Papers in Linguistics.

Bittner, Maria, and Ken Hale. (1996) "The structural determination of case and agreement". *Linguistic inquiry* 27(1): 1–68.

Bobaljik, Jonathan, and Susi Wurmbrand. (2005) "The domain of agreement". *Natural language and linguistic theory* 23: 809–65.

Bošković, Željko. (2008) "What will you have, DP or NP?". Emily Elfner and Martin Walkow, eds. *NELS 37: Proceedings of the thirty-seventh annual meeting of the North East Linguistic Society.* Vol. 1. Amherst: GLSA, Department of Linguistics, University of Massachusetts, 101–14.

————. (2014) "Now I'm a phase, now I'm not a phase: On the variability of phases with extraction and ellipsis". *Linguistic inquiry* 45(1): 27–89.

Brody, Michael. (2000) "Mirror theory: Syntactic representation in perfect syntax". *Linguistic inquiry* 31(1): 29–56.

Brody, Michael, and Anna Szabolcsi. (2003) "Overt scope in Hungarian". *Syntax* 6(1): 19–51.

Caha, Pavel. (2009) *The nanosyntax of case*. PhD thesis, University of Tromsø, Norway.

————. (2021) "Modeling declensions without declension features. The case of Russian". *Acta linguistica academica* 68(4): 385–425.

Cardinaletti, Anna, and Michael Starke. (1999) "The typology of structural deficiency: A case study of the three classes of pronouns". Henk van Riemsdijk, ed. *Eurotyp.* Vol. 5, *Clitics in the Languages of Europe*. Part 1. Berlin: Mouton de Gruyter, 145–233.

Chomsky, Noam. (2001) "Derivation by phase". Michael Kenstowicz, ed. *Ken Hale: A life in language*. Cambridge, MA: MIT Press, 1–52.

Corbett, Greville. (1991) *Gender*. Cambridge, UK: Cambridge University Press.

————. (2006) *Agreement*. Cambridge, UK: Cambridge University Press.

de Hoop, Helen, and Andrej L. Malchukov. (2008) "Case-marking strategies". *Linguistic inquiry* 39(4): 565–87.

Deal, Amy R. (2015) "Interaction and satisfaction in ϕ-agreement". Thuy Bui and Deniz Özyildiz, eds. *NELS 45: Proceedings of the forty-fifth annual meeting of the North East Linguistic Society*. Vol. 1. Amherst: GLSA, Department of Linguistics, University of Massachusetts, 179–92.

Déchaine, Rose-Marie, and Martina Wiltschko. (2002) "Decomposing pronouns". *Linguistic inquiry* 33(3): 409–42.

Despić, Miloje. (2011) *Syntax in the absence of determiner phrase*. PhD thesis, University of Connecticut in Storrs.

————. (2017) "Investigations in mixed agreement: Polite plurals, hybrid nouns and coordinate structures". *Morphology* 27(3): 253–310.

Elbourne, Paul. (2005) *Situations and individuals*. Cambridge, MA: MIT Press.

Embick, David. (2010) *Localism versus globalism in morphology and phonology*. Cambridge, MA: MIT Press.

————. (2021) "The motivation for roots in distributed morphology". *The annual review of linguistics* 7: 69–88.

Embick, David, and Alec Marantz. (2008) "Architecture and blocking". *Linguistici inquiry* 39(1): 1–53.

Falk, Yehuda N. (1991) "Case: Abstract and morphological". *Linguistics* 29(2): 197–230.

Fassi Fehri, Abdelkader. (2018) *Constructing feminine to mean: Gender, number, numeral, and quantifier extensions in Arabic*. Lanham, MD: Lexington Books.

————. (2000) "Distributing features and affixes in Arabic subject verb agreement paradigms". Jacqueline Lecarme, Jean Lowenstamm, and Ur Shlonsky, eds. *Research in Afroasiatic grammar*. Amsterdam: John Benjamins, 79–100.

Foley, Steven, and Maziar Toosarvandani. (2019) "Agreement, locality, and the syntax of pronouns: The person-case constraint and beyond". Manuscript, University of California, Santa Cruz.

Franks, Steven. (2013) "Orphans, doubling, coordination, and phases: On nominal structure in Slovenian". *Slovenski jezik / Slovene linguistic studies* 9: 55–92.

Franks, Steven, and Asya Pereltsvaig. (2004) "Functional categories in the nominal domain". Olga Arnaudova, Wayles Browne, Maria L. Rivero, and Dejan Stojanović, eds. *Annual Workshop on Formal Approaches to Slavic Linguistics 12. The Ottawa meeting 2003*. Ann Arbor: University of Michigan Press, 109–28.

Georgi, Doreen. (2012) "A uniform analysis of global and local argument encoding patterns: A local and cyclic approach". Camelia Constantinescu, Bert L. Bruyn, and Kathrin Linke, eds. *Proceedings of ConSOLE XVII*. ConSOLE XVII, Nova Gorica, Slovenia, January 2009. Leiden, The Netherlands: Leiden University Centre for Linguistics, 137–61.

Georgi, Doreen. (2013) "A relativized probing approach to person encoding in local scenarios". *Linguistic variation* 12(2): 153–210.

Greenberg, Joseph H. (1963) "Some universals of grammar with particular reference to the order of meaningful elements". Joseph H. Greenberg, ed. *Universals of language*. Cambridge, MA: MIT Press, 73–113.

Gribanova, Vera, and Boris Harizanov. (2017) "Locality and directionality in inward-sensitive allomorphy: Russian and Bulgarian". Vera Gribanova and Stephanie Shih, eds. *The morphosyntax-phonology connection: Locality and directionality at the interface*. Oxford, UK: Oxford University Press, 61–90.

Gruber, Bettina. (2013) *The spatiotemporal dimensions of person: A morphosyntactic account of indexical pronouns*. Utrecht: LOT.

Hammerly, Christopher. (2018) "Limiting gender". Éric Mathieu, Myriam Dali, and Gita Zareikar, eds. *Gender and noun classification*. Oxford, UK: Oxford University Press.

Harbour, Daniel. (2007) "Against PersonP". *Syntax* 10(3): 223–42. DOI 10.1111/j.1467-9612.2007.00107.x.

————. (2008) "Discontinuous agreement and the syntax-morphology interface". Daniel Harbour, David Adger, and Susan Béjar, eds. *Phi theory: Phi-features across modules and interfaces*. Oxford, UK: Oxford University Press, 185–220.

————. (2016) *Impossible persons*. Cambridge, MA: MIT Press.

Harley, Heidi, and Elizabeth Ritter. (2002) "Person and number in pronouns: A feature-geometric analysis". *Language* 78(3): 482–526.

Jovović, Ivana. (2024) "Condition B and other conditions on pronominal licensing in Serbo-Croatian". *Linguistic inquiry* 55(2): 402–21.

Kalin, Laura. (2019) "Nominal licensing is driven by valued (phi-)features". Gillian Ramchand and Peter Svenonius, eds. "GLOW Short Reports", special issue. *Nordlyd* 43(1): 15–29.

Kornfilt, Jaklin, and Omar Preminger. (2015) "Nominative as no case at all: An argument from raising-to-acc in Sakha". Andrew Joseph and Esra Predolac, eds. *Proceedings of the 9th Workshop on Altaic Formal Linguistics (WAFL 9)*. Cambridge, MA: MITWPL, 109–20. [MIT Working Papers in Linguistics 76.]

Kramer, Ruth. (2015) *The morphosyntax of gender*. Oxford, UK: Oxford University Press.

Kratzer, Angelika. (2007) "On the plurality of verbs". Johannes Dölling, Tatjana Heyde-Zybatow, and Martin Schäfer, eds. *Event structures in linguistic form and interpretation*. Berlin, Boston: Mouton de Gruyter, 269–99.

Kučerová, Ivona. (2018) "Phi-features at the syntax-semantics interface: Evidence from nominal inflection". *Linguistic inquiry* 49(4): 813–45.

Landau, Idan. (2016) "DP-internal semantic agreement: A configurational analysis". *Natural language and linguistic theory* 34(3): 975–1020.

Lander, Eric, and Liliane Haegeman. (2016) "The nanosyntax of spatial deixis". *Studia linguistica* 71: 1–66.

Longobardi, Giuseppe. (1994) "Reference and proper names: A theory of N-movement in syntax and logical form". *Linguistic inquiry* 25(4): 609–65.

Marantz, Alec. (2001) "Words". Manuscript, Massachusetts Institute of Technology, Cambridge, MA.

Marantz, Alec. (2007) "Phases in words". Sook-hee Choe, ed. *Phases in the theory of grammar*. Seoul: Dong-In Publishing Company, 191–222.

Marušič, Franc, and Rok Žaucer. (2021) "Dual in Slovenian". Patricia Cabredo Hofherr and Jenny Doetjes, eds. *The Oxford handbook of grammatical number*. Oxford: Oxford University Press, 428–44.

McFadden, Thomas. (2018) "*ABA in stem-allomorphy and the emptiness of the nominative". *Glossa* 3(1): 1–36.

McFadden, Thomas, and Sandya Sundaresan. (2009) "Subject distribution and finiteness in Tamil and other languages: Selection vs. case". *Journal of South Asian linguistics* 2: 5–34.

McGinnis, Martha. (2005) "On markedness asymmetries in person and number". *Language* 81: 699–718.

Milićev, Tanja, and Maša Bešlin. (2019) "Instrumental case: Why it is absent from the clitic system in Serbian-Croatian". Vesna Lopičić and Biljana Mišić Ilić, eds. *Jezik, književnost, teorija* [Language, literature, theory]. Niš, Serbia: Faculty of Philosophy, University of Niš, 153–68.

Moskal, Beata. (2015a) *Domains on the border: Between morphology and phonology*. PhD thesis, University of Connecticut.

————. (2015b) "Limits on allomorphy: A case study in nominal suppletion". *Linguistic inquiry* 46(2): 363–75.

Moskal, Beata, and Peter W. Smith. (2016) "Towards a theory without adjacency: Hyper-contextual VI-rules". *Morphology* 26: 295–312.

Neeleman, Ad, and Kriszta Szendrői. (2007) "Radical pro drop and the morphology of pronouns". *Linguistic inquiry* 38(4): 671–714.

Neeleman, Ad, and Fred Weerman. (1999) *Flexible syntax: A theory of case and arguments*. Vol. 47 of *Studies in natural language and linguistic theory*. Dordrecht: Springer.

Nevins, Andrew. (2007) "The representation of third person and its consequences for the person–case constraint". *Natural language and linguistic theory* 25(3): 273–313.

————. (2011) "Multiple Agree with clitics: Person complementarity vs. omnivorous number". *Natural language and linguistic theory* 29(4): 939–71.

Noyer, Robert. (1992) *Features, positions, and affixes in autonomous morphological structure*. PhD thesis, MIT, Cambridge, MA.

Pereltsvaig, Asya. (2007) *Copular sentences in Russian: A theory of intra-clausal relations*. Dordrecht: Springer.

Pesetsky, David. (2013) *Russian case morphology and the syntactic categories*. Cambridge, MA: MIT Press.

Postal, Paul. (1969) "On so-called 'pronouns' in English". David Reibel and Sanford Schane, eds. *Modern studies in English: Readings in transformational grammar*. Englewood Cliffs, NJ: Prentice-Hall, 201–24.

Preminger, Omer. (2014) *Agreement and its failures*. Cambridge, MA: MIT Press.

Progovac, Ljiljana. (1998) "Determiner phrase in a language without determiners". *Journal of linguistics* 34: 165–79.

Puškar, Zorica. (2018) "Interactions of gender and number agreement: Evidence from Bosnian/Croatian/Serbian". *Syntax* 21(3): 275–318. DOI 10.1111/synt.12154.

Puškar-Gallien, Zorica. (2019) "Resolving polite conflicts in predicate agreement". *Glossa* 4(1): 33. DOI 10.5334/gjgl.587.

————. (forthcoming) "Morphosemantic mismatches with pronouns as a consequence of their internal structure". Berit Gehrke, Denisa Lenertová, Roland Meyer, Daria Seres, Luka Szucsich, and Joanna Zaleska, eds. *Advances in formal Slavic linguistics 2022*. Berlin: Language Science Press.

Ruda, Marta. (2021a) "Polish personal pronouns: [$_{persp}$ pers [$_{nump}$ num [n]]] and [$_{nump}$ num [n]]". *Studies in Polish linguistics* 16(1): 23–40.

————. (2021b) "Strict and sloppy readings of pronominal objects in Polish". *Studies in Polish linguistics* 16(2): 121–44.

Runić, Jelena. (2014) *A new look at clitics, clitic doubling, and argument ellipsis: Evidence from Slavic*. PhD thesis, University of Connecticut.

Sichel, Ivy, and Maziar Toosarvandani. (2021) "Nominal intervention and the extended person domain". Robert Autry, Gabriela de la Cruz, Luis A. Irizarry Figueroa, Kristina Mihajlovic, Tianyi Ni, Ryan Smith, and Heidi Harley, eds. *Proceedings of the 39th West Coast Conference on Formal Linguistics*. University of Arizona, 8–11 April 2021. Somerville, MA: Cascadilla Proceedings Project, 662–69.

————. (2024) "The featural life of nominals". *Linguistic inquiry* (2024): 1–56. DOI 10.1162/ling_a_00517.

Smith, Peter W., Beata Moskal, Ting Xu, Jungmin Kang, and Jonathan D. Bobaljik. (2019) "Case and number suppletion in pronouns". *Natural language and linguistic theory* 37: 1029–101. DOI 10.1007/s11049-018-9425-0.

Stegovec, Adrian. (2019) "Crop to fit: Pronoun size and its relation to strict/sloppy identity". Handout from a talk presented at the 93rd Annual Meeting of the Linguistic Society of America (LSA), New York, NY, 3–6 January 2019.

Steriopolo, Olga. (2018a) "Mixed gender agreement in the case of Russian hybrid nouns". *Questions and answers in linguistics* 5(2). DOI 10.2478/qal-2018-0001.

————. (2018b) "Morphosyntax of gender in Russian sex-differentiable nouns". *Journal of Slavic linguistics* 26(1): 307–36.

Steriopolo, Olga, and Martina Wiltschko. (2010) "Distributed gender hypothesis". Gerhild Zybatow, Philip Dudchuk, Serge Minor, and Ekaterina Pschehotskaya, eds. *Formal studies in Slavic linguistics: Proceedings of FDSL 7.5*. New York: Peter Lang, 155–72.

Talić, Aida. (2018) "Spelling out enclitics and giving their tone a voice: Cyclic clitic incorporation in BCS and breaking the cycle". *The linguistic review* 35(2): 307–70.

Taraldsen, Knut T. (1996) "Reflexives, pronouns, and subject/verb agreement in Icelandic and Faroese". James R. Black and Virginia Motapanyane, eds. *Micorparametric syntax and dialect variation*. Amsterdam: John Benjamins, 189–211.

Trommer, Jochen. (2002) "The interaction of morphology and syntax in affix order". Geert Booij and Jaap van Marle, eds. *Yearbook of morphology 2002*. Dordrecht: Kluwer, 283–324.

van Koppen, Marjo. (2012) "The distribution of phi-features in pronouns". *Natural language and linguistic theory* 30(1): 135–77.

van Urk, Coppe. (2018) "Pronoun copying in Dinka and the copy theory of movement". *Natural language and linguistic theory* 36(3): 937–90.

Weerman, Fred, and Jacqueline Evers-Vermeul. (2002) "Pronouns and case". *Lingua* 112: 301–38.

Wiland, Bartosz. (2018) "A note on lexicalizing 'what' and 'who' in Russian and in Polish". *Poznan´ studies in contemporary linguistics* 54(4): 573–604.

Willer-Gold, Jana, Boban Arsenijević, Mia Batinić, Nermina Čordalija, Marijana Kresić, Nedžad Leko, Franc L. Marušič, Tanja Milićev, Nataša Milićević, Ivana Mitić, Andrew Nevins, Anita Peti-Stantić, Branimir Stanković, Tina Šuligoj, and Jelena Tušek. (2016) "Morphosyntactic production of coordination agreement in South Slavic: From theory to experiments". *Journal of Slavic linguistics* 24(1): 187–224.

Zorica Puškar-Gallien
Leibniz-Zentrum Allgemeine Sprachwissenschaft
(Leibniz-Centre General Linguistics)
Berlin, Germany
puskar@leibniz-zas.de

Reviews

Ludmila Veselovská. *Wh-questions: A case study in Czech.* Olomouc, Czech Republic: Palacký University, 2021. 283 pp. ISBN 978-80-244-5965-3 (print), ISBN 978-80-244-5966-0 (online PDF).

Reviewed by Julia Bacskai-Atkari

The volume, written by Ludmila Veselovská, takes a look at the syntax of *wh*-questions, with a particular focus on Czech. The author adopts a generative approach and aims at reconsidering the previous approaches to *wh*-fronting in the literature in light of the Czech data, as contrasted especially with English. Apart from discussing data obtained via grammaticality judgments, the book presents the results of an extensive corpus study. The volume contains altogether seven chapters. Chapters 1–2 discuss the earlier literature on *wh*-movement in general, while chapters 3–4 are dedicated to earlier analyses focusing on multiple *wh*-movement; Chapter 5 summarizes some novel approaches related to Czech multiple *wh*-fronting. Chapter 6 presents the results from the corpus study, while Chapter 7 is an appendix containing the actual data.

Chapter 1 introduces the basics regarding the notion of movement. The discussion is largely theory-neutral and explains the reasons for assuming movement in standard generative grammar in such a way that it should also make the motivations clear for readers coming from different frameworks. In the following example (p. 15, (1.4b)), for instance, the *wh*-element corresponds to an argument of the lexical verb, as simply leaving out the same argument is not permitted in other contexts; it seems straightforward that *what* should be present in the derivation at some point as a verbal argument.

(1) [$_{NP}$ **What**] did Tom put [$_{NP}$ <**t**>] on the table?

Apart from subcategorization, Veselovská summarizes standard arguments from case and the binding of anaphors as well as the English-specific "wanna" contraction. As mentioned in the introduction (p. 12), the arguments are based on standard works such as Haegeman and Guéron (1999) and Adger (2003) for English; the literature review here would have benefited from more

specific references in the individual sections, if previous knowledge is not assumed on the reader's part anyway. The argumentation is clearly presented, though the line of reasoning citing case (pp. 15–17) works better for Czech, where morphological case is indicative of function, while in English this is not so (see (1)); for this, a more detailed discussion of English would have been necessary. The rest of the chapter is concerned with briefly presenting other structures showing *wh*-movement, much in the spirit of Chomsky (1977); the data here are given for English only—at least a brief note regarding to what extent the relatedness of the constructions is tenable for Czech would have been welcome.

Chapter 2 offers a contrastive overview of the basic English and Czech data regarding the constraints on *wh*-movement, such as pied-piping, long-distance movement, islands, and the Left Branch Condition. The chief merit of the chapter lies in a meticulous comparison of the two languages; such neat and detailed comparisons are an enrichment to the existing literature. Some of the differences appear relatively obvious; for instance, since Czech does not allow preposition stranding, it is straightforward that extraction out of PPs (pp. 34–36) will be impossible. In other cases, the empirical data are less clear and the conclusions are less robust. For instance, while long-distance movement is grammatical with various English verbs (traditionally called "bridge verbs"), this option is marked as ungrammatical in Czech prescriptive grammars (p. 29, (2.12c)):

(2) [$_{CP}$ **Co** si Máša myslí, [$_{CP}$ *(že) Tom koupil <**t**>]].
what.ACC REFL Mary thinks that Tom bought
'What does Mary think (people say) Tom bought?'

Confusingly, while Veselovská also marks (2) as ungrammatical, she adds (p. 29) that such sentences with the complementizer *že* are actually acceptable in colloquial language and interpreted as reported speech. Given that the monograph is a formal, generative study, it is somewhat unfortunate that prescriptive judgments are mixed with notions of grammaticality, as the intended audience of the book is undoubtedly more interested in the latter. Likewise, novel experimental data on adjunct island violations are mentioned only in a footnote (p. 37) but not incorporated into the investigation. As the monograph should cover a gap in the existing literature in terms of the fine-tuned empirical description mentioned above, going beyond the limits of what is accepted by traditional grammars would have been desirable.

Chapter 3 presents some basic considerations regarding multiple *wh*-movement. The phenomenon is illustrated for Czech on the following page (p. 70, (3.2)).

(3) a. **Kdo** **co** koupil?
who.NOM what.ACC bought
'Who bought what?'

b. **Co** **kdo** koupil?
what.ACC who.NOM bought
'Who bought what?'

As can be seen, more than one *wh*-element can be fronted in Czech; further, the order of these is not fixed, as the subject either precedes or follows the (direct) object. This contrasts with English, which shows superiority effects and is generally assumed to ban multiple fronting. The Czech data thus present at least two major challenges: (i) how cross-linguistic differences can be accounted for and (ii) how multiple *wh*-elements can be accommodated at the front of the clause. Chapter 3 reviews the available literature in this respect, and it also clearly shows that while the issue is evidently at the heart of generative investigations, specifically the Czech data did not receive ample attention previously. Most analyses tried to relate cross-linguistic differences to certain parameters, such as the possibility of multiple filled specifiers (Rudin 1988). At the same time, it also emerges clearly that most of the challenges for the analysis stem from the assumption that a given projection (such as the CP) can only have a single specifier position so that multiple *wh*-elements are either stacked into a single specifier or they require separate projections, or at least the non-first *wh*-element should be attached as an adjunct. While several aspects of the cited analyses, as Veselovská also points out, count nowadays as outdated (such as the use of the S and COMP labels), the insights are still highly relevant for the current theory as well, and it is therefore particularly welcome that the book presents such a succinct and clear overview.

Following this line of thought, Chapter 4 is dedicated to the analysis of Czech multiple *wh*-questions. While examples like (3) may suggest that the fronting operation targets essentially the same position (or at least the same projection), other examples clearly show that there is an asymmetry between the first *wh*-element and all the other ones. Consider the following sentence (p. 58, (4.3a)).

(4) **Co** by **komu** **kdy** udělala Monika?
what.ACC AUX.3 whom.DAT when done.S.F Monika.NOM
'What would Monica do to whom when?'

In (4), the auxiliary intervenes between the first *wh*-element *co*, and the second and third *wh*-elements *komu* and *kdy*. Importantly, *komu* and *kdy* are not in situ, as they precede the subject (*Monika*); if they remained in a postverbal

position, they would instead be interpreted as indefinites. In other words, examples like (4) clearly show that all *wh*-phrases undergo fronting, though not all of them target the same position. The auxiliary is a second-position clitic that is associated with the C position: this indicates that the first *wh*-element indeed targets [Spec,CP], while the other *wh*-phrases target lower positions, which Veselovská, following the original proposal of Rudin (1988) for Czech and Polish, identifies as adjunction to IP. This kind of asymmetry appears to be systematic; as Veselovská convincingly shows, the same can be observed with the unmarked complementizer *že* 'that' and the second-person auxiliary affix *-s/jsi*. While the trigger for the movement of the non-first *wh*-phrases is not clearly fleshed out, the empirical data altogether strongly support a multiple constituent analysis of Czech multiple *wh*-fronting and the summary in this chapter provides the basis for further investigations.

Chapter 5 continues discussing the previous literature; while a major focus is still on multiple *wh*-movement, many of the cited analyses only marginally touch upon this issue. The works discussed here are rather divergent, and the chapter creates the impression of a series of review articles that fail to provide a coherent picture when put together. There are two main aspects introduced in this chapter that distinguish newer approaches from more traditional ones: the availability of a more complex left periphery and the consideration of information-structure notions. Citko (1998) and Bošković (1997), as well as subsequent works by the same authors, are summarized adequately in this respect, and the immediate connection to multiple *wh*-movement is also clear. By contrast, the summaries in §5.3 are only loosely connected, and in some cases, not adequately represented. For instance, Veselovská claims that Bacskai-Atkari (2018) offers a "cross-linguistic perspective in multiple *Wh*-questions" (p. 135), while the cited article in fact deals with doubly filled COMP effects of the sort given in (5) below (p. 136, (5.41)).

(5) Kdo-že přijel pozdě?
who-that arrived late
'WHO arrived late?'

For cases like (5), which have an echo reading, Veselovská wrongly cites that the decisive feature is the leftmost one, when it is in fact claimed to be the rightmost one, as is also evident from the table cited in (5.40). Likewise, Veselovská seems to suggest that feature stacking (as also in Sturgeon 2008) and the lack of information-structure features in the narrow syntax (as also in Kaspar 2017) are shortcomings, while this is in fact a standard assumption in non-cartographic approaches (see Fanselow and Lenertová 2011).

Chapter 6 presents novel corpus data based on the Czech National Corpus, examining the question of whether superiority effects can be detected in multiple *wh*-fronting. Superiority effects in Czech were investigated exper-

imentally by Meyer (2004), who showed that superiority effects can be detected in Czech, while these effects are not categorical (as claimed for English) but rather scalar (across Slavic). He further observed a reverse animacy effect, which means that when an animate subject is combined with an inanimate object, no superiority effects arise. An example like (6) below (p. 157, (6.13c)), which demonstrates a counter-hierarchical order in terms of animacy, is just as acceptable as the opposite (hierarchical) order.

(6)	**Co**	**kdo**	řikal	o	mý	ruce?
	what.ACC	who.NOM	said	about	my	hand

'Who said what about my hand?'

These findings are also corroborated by the results presented by Veselovská. Further, her study clearly shows that certain combinations are more frequent than others; interestingly, while the effects regarding superiority seem to occur not only with two fronted *wh*-elements with no intervenor but also with two (fronted) coordinated *wh*-elements, the preferences for the combinations of pronominal and adverbial *wh*-elements point to a clear split between these two constructions. While the title of the chapter and several points in the text refer to a statistical analysis of the corpus results, no such analysis has actually been carried out, as only the raw frequencies and the percentages are presented. For a fine-tuned analysis, especially in terms of being comparable to the experimental findings, the use of proper statistical tests would have been more than desirable.

Despite some of the concerns mentioned above, I believe that *Wh-Questions: A Case Study in Czech* is an excellent volume that provides novel insights into the syntax of Czech *wh*-questions, and it also offers a fresh look at the earlier analyses that have not been discussed extensively in any comparable work. The text is well written and the ideas of the author are conveyed in a clear way. At the same time, there are unfortunately quite a few issues that indicate the lack of professional proofreading, the burden of which lies primarily with the publisher. For instance, section heading 2.1 is entirely missing. There are several typos and quite a few copy-and-paste errors in the example glosses (as on p. 156, where *co* is translated twice as 'who' in (6.12c) and (6.12d) and *koho* is translated as 'what' in (6.12c)).

Even though such minor errors may occasionally disturb the reader, they do not undermine the general high quality of the book, which I found to be a lucid and thought-provoking read.

References

Adger, David. (2003) *Core syntax: A minimalist approach*. Oxford, UK: Oxford University Press.

Bacskai-Atkari, Julia. (2018) "Doubly filled COMP in Czech and Slovenian interrogatives". Denisa Lenertová, Roland Meyer, Radek Šimík, and Luka Szucsich, eds. *Advances in formal Slavic linguistics 2016*. Berlin: Language Science Press, 1–23. DOI 10.5281/zenodo.2545509.

Bošković, Željko. (1997) "Superiority effects with multiple *wh*-fronting in Serbo-Croatian". *Lingua* 102(1): 1–20.

Chomsky, Noam. (1977) "On *wh*-movement". Peter W. Culicover, Thomas Wasow, and Adrian Akmajian, eds. *Formal syntax*. New York: Academic Press, 71–132.

Citko, Barbara. (1998) "On multiple WH movement in Slavic". Željko Bošković, Steven Franks, and William Snyder, eds. *Formal Approaches to Slavic Linguistics: The Connecticut Meeting, 1997*. Ann Arbor, MI: Michigan Slavic Publications, 97–114.

Fanselow, Gisbert, and Denisa Lenertová. (2011) "Left peripheral focus: Mismatches between syntax and information structure". *Natural language & linguistic theory* 29(1): 169–209.

Haegeman, Liliane, and Jacqueline Guéron. (1999) *English grammar: A generative perspective*. Oxford, UK: Blackwell.

Kaspar, Jiri. (2017) *Syncretism: A case study of the particle že in Czech*. PhD dissertation, University College London.

Meyer, Roland. (2004) *Syntax der Ergänzungsfrage: Empirische Untersuchungen am Russischen, Polnischen und Tschechischen*. Munich: Sagner.

Rudin, Catherine. (1988) "On multiple questions and multiple WH fronting". *Natural language & linguistic theory* 6(4): 445–501.

Sturgeon, Anne. (2008) *The left periphery: The interaction of syntax, pragmatics and prosody in Czech*. Amsterdam: John Benjamins.

Julia Bacskai-Atkari
Department of Modern Foreign Languages and Cultures
University of Amsterdam
Amsterdam, Netherlands
j.bacskaiatkari@uva.nl

Janina Mołczanow. *Interactions of vowel quality and prosody in East Slavic*. Sheffield, UK: Equinox Publishing Ltd, 2022. xii + 203 pp. [Series: Advances in Optimality Theory.] ISBN-13 9781800502345 (hbk), ISBN 9781800502352 (ebk).

Reviewed by Darya Kavitskaya

Janina Mołczanow's book *Interactions of Vowel Quality and Prosody in East Slavic* develops an account of vowel reduction in East Slavic languages, such as standard Belarusian and standard Russian, as well as many dialects that exhibit many different patterns of vowel reduction.[1] Vowel reduction in East Slavic is extremely complex. Traditionally, one distinguishes moderate reduction, which applies in the syllable immediately preceding the stressed one (called "first pretonic" or just "pretonic" in the Slavic literature), and extreme reduction, which applies to vowels in other unstressed syllables in a word. While Mołczanow adopts an existing account of extreme reduction, she proposes a novel treatment of moderate reduction. The claim of the book is that all the different patterns of moderate vowel reduction attested in East Slavic can be unified because all vowel alternations that are the result of this type of reduction are due to the presence of High tone in the head foot. Mołczanow states that "different patterns of reduction in pretonic syllables arise due to the interaction of phonological tone with segmental quality" (p. 2), thus making an assumption about the presence of phonological tone in East Slavic.

The book consists of seven chapters (and the concluding remarks). Chapter 1 presents the theoretical basis of the book, which is the standard Optimality Theory (OT) model as applied to vowel reduction. Mołczanow argues that OT is the best theoretical apparatus to account for the East Slavic vowel reduction since OT is a typologically oriented theory that allows for an analysis of many related, distinct, but slightly varying patterns of alternations. The OT account proposed in the book relies on universally fixed constraint hierarchies with respect to the sonority hierarchy and the syllable peak and margin prominence scales, and harmonic alignment is used for scale combination. All the OT mechanisms employed in the analysis are general and developed by other researchers.

Chapter 2 outlines the basic generalizations with respect to vocalic neutralizations in East Slavic and provides a summary of the types of reduction

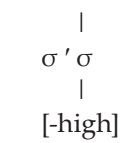
[1] Ukrainian does not have phonological vowel reduction.

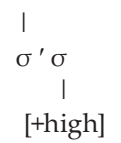
present in East Slavic dialects. Mołczanow takes into consideration several types of dissimilative reduction, paying attention to the data not only from the most well-known Žizdra, Obojan', and Don dialects, but also to the compound dissimilative reduction (backness harmony) types, such as those attested in Ščigry, Sudža, Mosal'sk, and Dmitrov dialects, as well as to the assimilative-dissimilative (height harmony) types, such as those in Novoselki, Kidusovo, Orexovo, Kultuki, and Bel'sk dialects. The chapter illustrates the types of non-dissimilative and dissimilative vowel reduction on the basis of the data from the existing descriptions, such as Avanesov (1974); Kasatkin (2005); Kasatkina (2000, 2005); Knjazev (2000); Požarickaja (2005); Vajtovič (1968); and Vysotskij (1984), among others.

Chapter 3 addresses the East Slavic metrical system, mostly on the basis of the facts of Russian morphologically-governed stress. Following Halle and Vergnaud (1987), Mołczanow assumes the presence of right-headed (iambic) feet in Russian (contra a later amendment to the proposed model by Idsardi 1992 and Halle and Idsardi 1995, among others, who assume left-headed (trochaic) feet in Russian). Mołczanow points out that while the phonetic and phonological data are not sufficient to argue for a type of foot in Russian, language games appear to provide enough evidence for an iambic foot.[2] Foot is thus the key domain in the proposed model of East Slavic stress. The main assumption is that there is a High tone associated to the head foot (due to HDFT = H constraint). The head foot carries both metrical prominence, as it is the site of stress, and tonal prominence due to the presence of High tone. Metrical and tonal prominence may or may not coincide, depending on the dialect. In the former situation, High tone is linked to the head syllable, and the resulting foot is unary. The latter situation happens when High tone shifts one syllable to the left of the stressed syllable, creating a disyllabic iambic foot. This allows for an account of moderate reduction in standard Russian and of dissimilative reduction in many East Slavic dialects. To illustrate, (1) shows that in the dialects with dissimilative reduction, the presence of H in the head foot is dependent on the quality of the stressed vowel. High tone is linked to the non-high stressed vowel in (1a) and surfaces on the immediately pre-stress vowel if the stressed vowel is high.

[2] Since the language game data are only available for standard Russian, it is unclear how the argument would develop for standard Belarusian or the dialects of either Russian or Belarusian.

(1) Distribution of H tone in the dissimilative dialects (p. 67)

a.


b.


Chapter 4 presents an account of the extreme reduction of vowels in non-pretonic positions, which is driven by the absence of stress. Following Crosswhite (2001) and De Lacy (2006), the extreme reduction is accounted for as prominence reduction in prosodically weak positions. There are environments that are exempt from the extreme reduction in Russian, specifically, vowels in absolute phrase-initial positions (as in /ogo'rod/ 'vegetable garden', pronounced as [aga'rot] or [ɐga'rot], but not *[əga'rot]) and in hiatus positions (as in /sootno'ʃenʲije/ 'proportion', pronounced as [saatnɐ'ʃenʲɪə] or [sɐɐtnɐ'ʃenʲɪə], but not [səətnɐ'ʃenʲɪə]). These exceptions are treated as gaps in the distribution of schwa. Following Anderson (1982) and Oostendorp (1995), among others, Mołczanow assumes that schwa "is devoid of a melodic content" (p. 75) and that an empty vowel cannot occur in an onsetless syllable. The quality of the resulting vowel in such positions is identical to the outcome of moderate reduction but cannot be attributed to the presence of High tone. An additional constraint is needed to derive the quality of the vowels in these positions, which, admittedly, makes the analysis seem a bit disjointed.

Chapter 5 presents the main point of the book, an account of tone-driven moderate vowel reduction in the first pretonic position. Mołczanow's analysis of pretonic reduction in East Slavic is built on the analytical assumptions made by Bethin (2006) and Dubina (2012). Bethin (2006) proposes that the High tone, which is associated with contrastive stress, is assigned to the immediately pretonic syllable. The accommodation of the pitch rise from the pre-stress to stress syllable is responsible for the pretonic lengthening in some dialects of East Slavic. Dubina (2012) goes on to extend Bethin's analysis to standard Belarusian and to propose that the increased prominence of pretonic vowels is due to the anticipatory spreading of High tone from the stressed syllable to the preceding one.

Mołczanow makes a further assumption, extending Bethin's tonal analysis to all East Slavic dialects. Under the present analysis, the phonological High tone can be correlated with segmental properties (vowel quality) and suprasegmental properties (vowel duration). An OT account is then developed for moderate non-dissimilative and dissimilative types of reduction.

Non-dissimilative reduction is illustrated on the basis of standard Russian, where the pretonic vowel is argued to be a host for High tone. The facts of standard Russian non-dissimilative reduction are derived by "conflating

the sonority scale with the tonal prominence scale" (p. 89), which results in markedness constraints of the type **H/i, *H/i;ə, *H/i;ə;i,u, *H/i;ə;i,u;e,o,* etc. The interaction of these markedness constraints with metrical structure constraints, as well as with faithfulness constraints on vowel quality, is responsible for the outcome of moderate vowel reduction, which reduces /o/ to an [a]-type vowel in non-palatalized contexts—the simplest case at hand. Lowering of /o/ to [a] makes the vowel longer and thus a better tone-bearing unit, while the higher ranked faithfulness to [+high] prohibits lowering of high vowels in the same environment.

Dissimilative vowel reduction refers to vowel reduction in which the quality of the reduced vowel depends on the quality of the vowel in the following stressed syllable. The patterns of dissimilative vowel reduction are derived by the ranking permutations between the sonority/tonal prominence constraints and HEAD=H constraint that requires the stressed (head) syllable to be associated with High tone.

Chapter 5 also presents an analysis of the dialects with pretonic vowel length. These are the dialects in which the lengthening was attributed to the presence of phonological tone by Bethin (2006). However, even in the pretonic length dialects, the presence of the phonetic correlate of the tone is inconsistent. Following Bethin (2006), Mołczanow assumes the presence of phonological High tone, adopts Bethin's analysis that links the immediately pretonic syllables in the lengthening dialects to the LH contour (LH=LONG), and introduces additional constraints that require head feet to be associated with a contour tone (HDFT=LH).

In the interim conclusion, Mołczanow acknowledges that the proposed analysis cannot resolve a long-standing debate about the origin of vowel reduction. The prediction of OT is that both patterns (dissimilative and non-dissimilative) are equally possible as an earlier stage given the proposed constraint and their possible re-rankings.

Chapter 6 turns to the harmony patterns in vowel reduction where in addition to dissimilation, there exists some feature harmony between the vowel in the pretonic position and the stressed vowel. After reviewing the facts of backness harmony (compound dissimilative patterns) and height harmony (assimilative-dissimilative patterns) presented in Chapter 2, Mołczanow proceeds to account for the dialects with these types of reduction in sections 6.2 and 6.3, respectively.

The difference between dissimilative and compound dissimilative patterns lies in the behavior of the reduced vowels before the stressed mid vowels. Compound dissimilative patterns of reduction only occur in the environment of palatalized consonants. Mołczanow points out that the five attested systems of compound dissimilative reduction can be viewed as "a combination of dissimilative reduction with assimilation in backness between vowels in two neighboring syllables" (pp. 121–22). The assimilative-dissimilative

patterns exhibit "reduction to [a] after palatalized consonants if the following tonic syllable contains high and low vowels. Before mid vowels, either [i] or [a] is attested, depending on a dialect" (p. 137). To account for the harmony, Mołczanow adopts Alignment constraints (such as Align-L([back], Ft) for compound dissimilative reduction and Align-L([low], Ft) for assimilative-dissimilative reduction) that operate within a foot, once again making the foot a crucial prosodic domain for the application of reduction and harmony.

Certain compound dissimilative patterns are opaque and thus present a problem for a standard OT model. In order to account for these patterns (e.g., the difference between the surface realizations of the pretonic vowels in the examples like [vʲiˈdʲɔm] 'lead.ɪPL.PRES' and [mʲidʲvʲaˈdʲɔm] 'bear.INSTR.SG' in Sudža and Ščigry dialects, where the stressed vowels are phonetically identical but come from different historical sources), Mołczanow needs to adopt a version of serial OT that allows for multiple levels (Bermúdez-Otero 1999 and Kiparsky 2000, among others).

Chapter 7 looks at a pattern where the presence of High tone fails to trigger vowel lowering in the pretonic position, which goes against the main argument of the book that High tone is best accommodated by more sonorous vowels. The lowering is blocked in the environment of palatalized consonants as well as after non-palatalized stridents [ʃ], [ʒ], and [ts]. The former case is attested in standard Russian as well as in several southern Russian and Belarusian dialects. The latter case is in effect in most East Slavic dialects.

Mołczanow treats the blocking of vowel lowering in the environment of palatalized consonants as assimilation, or agreement in backness with palatalized consonants in the cases where non-high back vowels /a/ and /ɔ/ surface as [e], and agreement in height in the cases where the non-high vowels surface as [i]. A further complication is required to account for the dialects such as Čuxloma, where [i] only surfaces if both preceding and following consonants are palatalized. To account for such alternations, a further Agree constraint is proposed, specifically referring to the environment (AGREE-C^{j}VC^{j}[+high]), which assigns a violation for every non-high vowel that occurs between two palatalized consonants. As this appears to be a case of *phonological teamwork*, an analysis along the lines of the one proposed by Lionnet (2016, 2017) would probably prove more insightful than a direct stipulation.

Finally, the inconsistency of reduction after non-palatalized stridents—cf. [ˈʃar] 'ball', [ʃaˈrɨ] 'ball.NOM.PL'; [ˈʃɛstʲ] 'six' (also pronounced [ˈʃesʲtʲ], e.g., by the present reviewer), [ʃɨˈstʲi] 'six.GEN'—is accounted for by the proposed constraint Hard-C that stipulates that anterior affricates and non-anterior coronal constituents cannot be [-back] (*tsʲ, *ʃʲ, *ʒʲ). This constraint is low ranked at Level 1, which allows for the palatalization of stridents, and high ranked at Level 2, which chooses unpalatalized stridents as the optimal output (/ʃɛstʲ-ˈi/ → *Level 1 output* ʃʲiHˈstʲi → *Level 2 output* ʃiHˈstʲi).

To summarize, the book argues that in East Slavic dialects, tone (i) is phonological, which amounts to saying that it is synchronically present and (ii) directly interacts with vowel quality, which implies that there is some interaction of the phonological tone with the properties of vowels that are phonetically realized.

However, while the analysis hinges on the assumption that High tone is present in all the relevant dialects, the phonetic realization of tone is not uniform throughout East Slavic: LHL contour is indeed attested in some dialects but not in others. An important fact that Mołczanow is well aware of is that only the "pretonic length" East Slavic dialects, such as the Vladimir-Volga Basin and the Malyja Aucjuki/Upper Snov dialect types (Bethin 2006), exhibit non-contrastive tonal contours over the pretonic vowel, while in others, such as standard Belarusian, "tone is employed as a phonological construct not translatable into phonetic pitch" (Mołczanow, p. 88). Thus, while it is quite plausible that High tone was present historically in East Slavic, the analysis that argues that High tone is the phonological reality of the dialects in question is rather abstract in those dialects where it is not present phonetically. The assumption of the phonological High tone unifies all the instances of moderate vowel reduction, and the remaining question is how much discrepancy between the phonological analysis and the phonetic substance are we willing to bring into a phonological analysis.

A completely different approach to the problem is developed by Iosad (2012), who posits that the difference between "moderate" and "radical" vowel reduction stems from the difference in duration between the vowels in the syllables that participate in reduction, and that the apparent sonority-driven effects are epiphenomenal (Iosad 2012: 521). This account is incompatible with Mołczanow's account and does not rely on abstract assumptions about tone in Russian. While Mołczanow acknowledges the existence of Iosad's "no substance" analysis, she does not address it in any substantial detail.

The book provides an internally consistent working analysis of complex patterns of East Slavic vowel reduction, covering the data from many dialects described in the literature. OT is used to derive the typology of vowel reduction, without under- or over-generating the patterns, employing the generally accepted OT constraints and using universally fixed rankings to express implicational universals. The book develops a type of study that Optimality Theory was designed to do and is particularly good at: an account of a phenomenon attested in closely related dialects that differ in certain parameters.

References

Anderson, Stephen R. (1982) "The analysis of French schwa: Or, how to get something for nothing". *Language* 58: 534–73.

Avanesov, Ruben Ivanovič. (1974) *Russkaja literaturnaja i dialektnaja fonetika*. Moscow: Prosveščenie.

Bermúdez-Otero, Ricardo. (1999) *Constraint interaction in language change: Quantity in English and Germanic*. PhD dissertation, University of Manchester.

Bethin, Christina Y. (2006) "Stress and tone in East Slavic dialects". *Phonology* 23: 125–56.

Crosswhite, Catherine M. (2001) *Vowel reduction in Optimality Theory*. London/New York: Routledge.

de Lacy, Paul. (2006) *Markedness: Reduction and preservation in phonology*. Cambridge, UK: Cambridge University Press.

Dubina, Andrei. (2012) *Toward a tonal analysis of free stress*. Utrecht: LOT Publications.

Halle, Morris, and William Idsardi. (1995) "General properties of stress and metrical structure". John A. Goldsmith, ed. *The handbook of phonological theory*. Cambridge, MA: Blackwell Publishers, 403–43.

Halle, Morris, and Jean-Roger Vergnaud. (1987) *An essay on stress*. Cambridge, MA: MIT Press.

Idsardi, William. (1992) *The computation of prosody*. PhD dissertation, MIT.

Iosad, Pavel. (2012) "Vowel reduction in Russian: No phonetics in phonology". *Journal of linguistics* 48: 521–71.

Kasatkin, Leonid L. (2005) "Fonetika". Leonid L. Kasatkin, ed. *Russkaja dialektologija*. Moscow: Akademija, 22–85.

Kasatkina, Rozalia F. (2000) "Južnorusskoe narečie. Novye dannye". *Voprosy jazykoznanija* 2000.6: 98–109.

_____. (2005) "Moskovskoe akan'e v svete nekotoryx dialektnyx dannyx". *Voprosy jazykoznanija* 2005.2: 29–45.

Kiparsky, Paul. (2000) "Opacity and cyclicity". *The linguistic review* 17: 351–67.

Knjazev, Sergej V. (2000) "K voprosu o mexanizme vozniknovenija akan'ja v russkom jazyke". *Voprosy jazykoznanija* 2000.1: 75–101.

Lionnet, Florian. (2016) *Phonological teamwork: A typology and theory of cumulative effects in phonology*. PhD dissertation, University of California, Berkeley.

_____. (2017) "A theory of subfeatural representations: The case of rounding harmony in Laal". *Phonology* 34: 523–64.

Oostendorp, Marc van. (1995) *Vowel quality and phonological projection*. PhD dissertation, Tilburg University.

Požarickaja, Sof'ja Konstantinovna. (2005) *Russkaja dialektologija*. Moscow: Akademičeskij Proekt, Paradigma.

Vajtovič, T. N. (1968) *Nenaciskny vakalizm narodnyx havorak Belarusi*. Minsk: Navuka i Texnika.

Vysotskij, S. S. (1984) "O zvukovoj structure slova v russkix govorax". J. S. Azarx, S. V. Bromlej, and L. N. Bulatova, eds. *Issledovanija po russkoj dialektologii*. Moscow: Nauka, 17–41.

Darya Kavitskaya
Department of Slavic Languages and Literatures
University of California Berkeley
Berkeley, CA, USA
dkavitskaya@berkeley.edu

Jasmina Milićević. *Serbian clitics*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2023. xviii + 166 pp. [Studies in Language Companion Series (SLSC), 229.][1] ISBN 9789027212979 (hbk), ISBN 9789027254641 (ebk).

Reviewed by Zrinka Kolaković

1. Introduction

1.1. Key Points

Serbian Clitics presents a comprehensive and theoretically based, formalized description of second-position clitics (henceforth 2P CLs) in standard Serbian. This monograph from the John Benjamins publishing house consists of seven chapters that previously appeared as individual papers in scholarly journals and edited volumes (Milićević 1999, 2005, 2007, 2009a, 2009b, 2014, 2019a, 2019b).

The seven chapters offer new, theoretically grounded insights into both morphology and syntax of 2P CLs, primarily focusing on the written standard Serbian variety. Given that there are no similar works that investigate Serbian 2P CLs within the theoretical framework of the Meaning-Text Theory (henceforth MTT) (Mel'čuk 1993, 1994, 1996, 1997, 1999), this is a highly noteworthy work. The approach taken sets this monograph apart from most contemporary studies on Slavic CLs because of its clear focus on syntactic dependencies and its orientation toward speech production (p. xiii).

Thus, being a reference work par excellence for all those interested in an approach to CLs that takes syntactic dependencies into account, the monograph will appeal not only to theoretical linguists interested in formal approaches to (Balto-)Slavic morphology and syntax but also to general linguists, typologists, Slavists, and advanced learners of Serbian.

[1] English proofreading funding was provided by the Faculty of Humanities of the University of Klagenfurt.

1.2. Language, Terminology, Bibliography, and Supplementary (Re)sources

The monograph is written in a precise and clear style. The terminology adopted fully aligns with current standards in linguistics, particularly those used in dependency-based theory and MTT. The author alerts the reader whenever the terms used in the book diverge significantly from other uses of the same expression (e.g., on pp. 2, 21, 31, 62, 77, 89, 94, 103). Moreover, the monograph contains an index of terms (pp. 165–66), allowing quick navigation through the text and access to key passages.

The text is enriched with carefully enumerated and captioned tables and figures, rendering it reader-friendly and easy to follow. The text is additionally supplemented with remark boxes containing more detailed comments and explanations of phenomena directly related to those discussed in the main text. Examples are carefully presented, numbered, translated, and glossed according to the Leipzig rules (https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf). Moreover, the reader can find an exhaustive explanation of glossing procedures, including comments on pronunciation and tonal accents (pp. xxii–xxiii) and explanations of symbols, abbreviations, and writing conventions, i.e., comments on which fonts are used in which function (pp. xix–xi). The author also provides the reader with an index of languages mentioned apart from Serbian (p. 163). This feature might be especially useful to those interested in typological research on CLs. Last but not least, the text is accompanied by ten pages of references (pp. 153–62).

2. Methodology and Data

2.1. Theoretical Framework

Although Milićević's monograph diverges from mainstream approaches to the Bosnian/Croatian/Montenegrin/Serbian (BCMS) CLs, the author herself characterizes it as mixed, assigning a greater role to syntax than to prosody (p. 22). Milićević's work on Serbian CLs was most strongly influenced by the least theory-dependent contributions, which feature authentic examples of CL usage more diverse than those usually found in the literature. These are Browne's (1974, 1975) two seminal papers written within the early generative stream and the detailed theory-neutral description of CL placement in literary Serbo-Croatian in Popović (1997: 22).

As there are only a handful of studies dedicated to CLs and CL linear placement based on syntactic dependencies (e.g., Iordanskaja 1982; Čamdžić and Hudson 2002; Hana 2007; Gross 2014), Milićević's comprehensive account of Serbian CLs was awaited with enthusiasm. For those unfamiliar with the

applied theoretical framework of MTT, Milićević (pp. 22–29) provides concise technical explanations in §3 of the first chapter, keeping the explanations as reader-friendly as possible by first giving an informal presentation, and following that, a formal description supported by abundant examples presented in graphic form. The author recommends that readers become acquainted with the technical details of MTT. However, a fair picture of the Serbian CL system can be obtained even if the reader entirely skips this part (p. xiii).

Unlike phrase-structure approaches to CLs, which strive for a good fit with the observables, the aim of MTT descriptions is to achieve a good fit with linguistic rules. An MTT account of CL placement begins with a linearly unordered syntactic dependency tree in which CLs have syntactic roles of actants or governors. CL linear positioning within (a morphological structure of) the clause takes place under the synthesis operation, which is governed by rules that take into account syntactic dependencies between clause elements and their syntactic/prosodic properties. In contrast to MTT, approaches to phrase structure rooted in classic generative syntax try to deduce CL "structural positions" in the linearly ordered phrase-structure tree, taking the possible linear arrangements in the clause as the starting point. These differences make the formalizations of the two mentioned approaches incommensurable (p. xiii).

Milićević (p. xiii) argues that it is the synthesis orientation of the MTT approach and its reliance on syntactic dependencies that makes it possible to gain essential insights into the Serbian CL system (and 2P CLs in general). The author claims that the use of MTT and syntactic dependencies permitted the role that syntactic and prosodic factors play in CL linear placement to be elucidated. Their interplay is characterized by syntactic relations that operate "behind the scenes" and shape the prosodic features of sentence elements to which CL placement is sensitive. Also, in Milićević's opinion, the approach taken allows one to pinpoint the differences between 2P and ad-verbal CLs with respect to the mechanism called clitic climbing (the author uses the term clitic pseudo-climbing, henceforth CPC). In contrast to ad-verbal CLs, 2P CLs, which are syntactically dependent on an infinitive outside the infinitive phrase, only pseudo-climb, merely displaying "climbing effects" in the matrix clause.

2.2. Language Data

The author focuses exclusively on CLs in contemporary standard Serbian, leaving out their diachronic and dialectal aspects. Some frequent divergent characteristics of CL usage in the press and colloquial registers are sporadically considered (p. xiv).[2] Still, the monograph analyzes contemporary

[2] For more information, see sections 4.5 and 5.

standard Serbian CLs within a broader BCMS context and draws attention to the major differences between the CL systems in these three closely related languages. Moreover, the author also offers a brief comparison with CL systems in other Slavic languages, arguing that BCMS CLs are the most representative example of the classical Wackernagel system (p. 152).

Milićević adopts a methodological approach regularly applied by linguists in formal theoretical and typological linguistic research: her data was gathered primarily through existing linguistic literature on Serbian CLs and on CLs in other languages of interest, the Corpus of Contemporary Serbian (Vitas et al. 2000; Utvić 2011; and www.korpus.matf.bg.ac.rs), the Serbian electronic and printed press, the Internet, examples provided by BCMS native speakers, and self-constructed examples (p. xv). The sources of linguistic data are clearly indicated only for languages other than BCMS. For BCMS examples, sources are provided only if Milićević's acceptability judgment differs from the acceptability judgment in the original source. Otherwise, there is no trace of data sources (pp. xv, xvii). This might be somewhat problematic since scholarly work is both incremental and collaborative, demanding that both data and analysis procedures are (re)traceable (Stefanowitsch 2020: 133–34).

The possible biases and issues that might have been caused by the methodology adopted were not mentioned explicitly. However, diverging opinions on the acceptability of specific structures are discussed on the margin and partly ascribed to the infrequency of the structures in question (p. 92). Although the methodology adopted is in line with the current state of the art in theoretical and typological linguistic research, follow-up studies embracing alternative methodological approaches based either on big data from corpora (spoken, written, or web-mediated) or on behavioral data (elicited production or acceptability) would be more than welcome in the future. This could be done with respect to selected issues raised in the monograph and possibly within a collaboration, e.g., with corpus linguists or psycholinguists.

3. Organization of the Monograph

The following sections each present one of the seven chapters in the monograph.

3.1. Chapter 1: Clitics and Syntactic Dependencies

Chapter 1 (pp. 1–29), offering a general introduction to the phenomenon of CLs, sets the scene for a more detailed presentation of issues closely related to Serbian CLs (chapters 2–7). The general concept of the CL described in §1 (pp. 1–9) is followed in §2 (pp. 9–22) by basic facts about the Serbian CL system. The final §3 (pp. 22–29) introduces the fundamentals of dependency syntax and morphology in MTT, i.e., the theoretical framework used in the work

to describe Serbian CLs.[3] Using Mel'čuk's (1993: 225) work as a point of departure, the author states that "[a] wordform w of language L is called clitic if and only if it lacks the prosodic features that characterize (almost) all wordforms of L" (p. 2). Moreover, she explicitly asserts that the prosodic deficiency of CLs is a language-specific feature (p. 3).

3.2. Chapter 2: Cliticization of Serbian Personal Pronouns and Auxiliary Verbs

This chapter (pp. 31–39) directly builds on the phenomenon, presented in Chapter 1, §2.1.1 (pp. 10–11), of Serbian personal pronouns and auxiliary verbs[4] having both full (tonic, stressed) and CL (atonic, unstressed) forms. Section 2 of Chapter 2 (pp. 33–37) examines in some detail the syntactic and communicative (information structural) factors that influence cliticization, i.e., determine whether the choice of CL over full forms of pronominal and verbal lexemes during sentence synthesis is obligatory or free.[5] Section 3 (p. 38) presents the corresponding rules for cliticization within an MTT model of Serbian.

According to the author, the CL forms of Serbian personal pronouns and auxiliary verbs are the default (pp. 32, 36). In contrast, full forms are either triggered by specific prosodic environments or syntactic configurations (including prosodic breaks, coordination, after prepositions, presence of a specific (co)-dependent), or freely chosen to express marked Focalized or Emphatic values of communicative oppositions (pp. 33–36). In §4, which contains a summary, the author argues that tonicity as a syntactic inflectional category has been "enlisted" to express information structure, as gender conversion is used to express derivational meanings (e.g., in Spanish) and a change of nominal class, to express plurality (e.g., in Bantu languages) (p. 39).

3.3. Chapter 3: Serbian Clitic Cluster

This chapter (pp. 41–60) also expands directly on Chapter 1: in this case, §2.1.3 (pp. 12–16). First, in §1, Milićević (pp. 41–45) provides a detailed description of the CL cluster in Serbian. Section 2 discusses the general and specific syntactic

[3] Section 3.1 of this review contains a brief overview of the benefits of the theoretical model applied for describing Serbian CLs.

[4] The author often uses auxiliary verbs as a cover term for both auxiliary and copular/locative verbs (see p. 31).

[5] Cliticization conditions for the interrogative particle *da li* were not considered in the monograph, whereas its dated CL form *li* is examined in Chapter 3 (pp. 43, 56–57). The emphatic particle *li* is not discussed in this chapter either since it is a CL lexeme without a corresponding full form (pp. 10, 25, 31).

and morphonological constraints on the co-occurrence of CL1 ~ CL2 within the cluster (pp. 46–56), and specific restrictions placed on some CL ~ host combinations (pp. 56–58). The author explains in what contexts dative and genitive CLs, genitive and accusative CLs, and accusative CL and reflexive CL *se* can, or cannot, appear together in a CL cluster and models restrictions as filter rules (pp. 46–56). Moreover, she also formulates rules for detecting and repairing illicit combinations. Regarding restrictions on CL ~ host combinations, Chapter 3 brings insights into which linguistic forms cannot host the interrogative particle CL *li* and the auxiliary CL *hteti* (pp. 56–58). The penultimate §3 (pp. 58–60) proposes examples of CL cluster building rules within MTT, whereas §4 (p. 60) provides a summary of the whole chapter.

3.4. Chapter 4: Linear Placement of Serbian Clitics

While §2.1.3.2 in Chapter 1 (pp. 14–16) presents a preliminary and concise description of the linear placement of Serbian CLs within the clause, this chapter offers a comprehensive presentation of the issue, including specific rules formulated within the MTT framework. More precisely, in Chapter 4 the author deals with the placement options that are, in principle, available to all clausal CLs.[6]

Section 1 (pp. 61–63) gives an overview of Serbian CL linear placement and introduces some considerations relevant to the dependency-oriented approach. A detailed informal description of CL placement within the clause follows in §2 (pp. 64–76), and §3 (pp. 76–86) presents an MTT-based description of the syntactic operations that drive CL placement in a clause.

The host of a CL in the morphological structure of a clause does not have to (and consequently often does not) correspond to its governor in the syntactic structure. In other words, Serbian 2P CLs do not have to be placed adjacent to their syntactic governors (p. 63). Consequently, all the CLs that appear in one clause and are linearly grouped in a cluster can have different governors in the syntactic structure, or can themselves be governors like, for instance, finite auxiliary/copular CL verbs (pp. 27, 87). Still, Milićević (pp. 82, 87) considers that syntactic dependencies in interplay with communicative oppositions play a crucial role in determining constituents and their properties necessary for CL placement. More specifically, they control all the properties of constituents on which CL placement depends, including prosodic properties (pp. 62, 67–68, 70–71, 78–79). In other words, although only indirectly, CL placement as a syntactic operation is essentially governed by syntactic dependencies. The role of prosodic factors is reflected merely in the prosodic features of sentence

[6] Special placement options (i.e., CPC out of dependent infinitive phrases and non-standard clause-initial and post-prosodic-break CL placement) are examined more thoroughly in Chapter 5.

elements (e.g., phonological heaviness, capacity to induce a prosodic break, contrastive stress) that affect those elements' capacity to host the CLs (p. 87). A prosodic break introduced by a constituent, as well as its prosodic contour, indicates its specific syntactic/communicative role, whereas the prosodic heaviness of a constituent results from its syntactic composition (pp. 67–69, 83–84).

Since various syntactic/communicative underlying factors can be reduced to the same prosodic expression, the author tries to capture and model them as the features [±heavy], [±detached], and [±contrastive] (p. 67). In variable CL placement, a constituent's suitability as a host depends both on the syntax and on the prosodic traits mentioned. Nevertheless, according to the author, the features [±heavy], [±detached], and [±contrastive] alone do not play a deciding role in CL placement. First, in the case of fixed CL placement, only syntactic factors, that is, the syntactic class of a constituent, are relevant (p. 66). Moreover, as Milićević (p. 68) argues, even in the case of variable placement the role of the features [±heavy], [±detached], and [±contrastive] can be overridden by syntactic factors. According to the author, some instances of variable CL placement in which syntax supersedes prosody occur when:

1. the syntactic role prevents skipping of the contrastive constituent, e.g., one that functions as (the lexical part of) the Main Verb of the clause (pp. 67–72);
2. the syntactic role and syntactic composition of a constituent are the dominant factors in making its splitting obligatory or blocking it (pp. 72–75);
3. there are syntax constraints on hosts of clusters containing the interrogative particle *li*: a non-finite verb cannot take such a role (p. 85).

Finally, some hosts must be lexically marked in order to act as hosts (p. 68).

To conclude, this chapter provides arguments for the crucial role of syntactic dependencies and communicative oppositions in the placement of Serbian 2P CLs. The author recognizes that factors expressed by prosody, such as the features [±heavy], [±detached], and [±contrastive], play a significant part in CL placement. However, in Milićević's opinion (pp. 62, 69, 88), CL placement cannot be entirely reduced to prosodic factors.

3.5. Chapter 5: Special Issues in the Linear Placement of Serbian Clitics

Chapter 5 (pp. 89–111) focuses on two unusual cases of CL linearization, each in its own section. Section 1 (pp. 89–108) deals with placement options available to CLs governed by an infinitive that is syntactically dependent on another clause element (verb, noun, or adjective), i.e., with CPC.

Section 2 (pp. 108–11) examines the atypical tendency seen in colloquial and journalistic registers of some Serbian 2P CLs to appear as PROclitics, both clause initially and after a post-prosodic break.

Unlike in Romance languages, Serbian infinitive CLs do not have to change their surface-syntactic governor in order to be placed non-locally. Since Serbian 2P CLs only display climbing effects and not genuine clitic climbing, Milićević (p. 103) argues that in their case, CPC is a more accurate term. Several factors play a role in resolving whether CPC will be obligatory, optional, or blocked. These are: the linear position of the infinitive phrase and the syntactic role of V_{INF}, prosodic features of clause elements, type and number of CLs involved, whether CPC will result in semantic ambiguity, whether CPC will result in a high number of CLs, whether CPC will result in case and case-person incompatibilities, and whether CPC will result in the sequences *[*se je*] (REFL AUX.3SG) or *[*se se*] (REFL REFL) (pp. 99–102). After a detailed discussion of these factors, the author presents linearization rules for infinitive CLs within MTT (pp. 105–08). Although CPC is obligatory if V_{INF} functions as the object of so-called restructuring verbs (e.g., modal and emotional state verbs) and if there are no prosodic factors at work, Milićević (pp. 90, 111) does not consider the phenomenon itself to be significant in Serbian. This is because the completive *da*-clause is usually used to replace the infinitive.

Clause-initial and post-prosodic break CL placement, often criticized by purists, is restricted to the Serbian substandard. Milićević (pp. 108–09) considers the first to be more common in polar questions in highly colloquial speech and in social media exchanges, attributing the latter to the language of the press. According to the author, post-prosodic break CL placement is primarily encountered for auxiliary/copular verb forms and the reflexive marker *se* (p. 108). Milićević (pp. 109–10) points out that in Serbian, the acceptability of this substandard CL position is a matter of degree, and that clause-initial CL placement is probably a result of the deletion of the clause-initial interrogative particle *da li* (colloquial *dal'* and *jel'*). As similar placement patterns are found in standard Slovene and colloquial Czech, i.e., languages that do not have classic Wackernagel CLs (Toman 1986: 127), the author speculates that the same fate might await Serbian 2P CLs (p. 111).

3.6. Chapter 6: Morphological Makeup of Serbian Clitics

Chapter 6 (pp. 113–27) offers a formal description of the morphology of Serbian pronominal and verbal CLs within MTT. The central MTT morphological concepts introduced in Mel'čuk (1993, 1994, 1996, 1997, 1999, 2006) are taken as a given and are not explained in detail. Needless to say, the author understands the term morphology as referring to the composition of CL wordforms, which are non-elementary linguistic signs, in terms of quasi-elementary and elementary signs (radical and affixal morphs, megamorphs)

(p. 113). Milićević (p. 113) uses "formal" to refer to establishing correspondences between "reasonable" deeper representations of the CLs and their phonemic signifiers (and vice versa) in such a way that they can be manipulated by a logical device (e.g., a computer program) and ultimately used in automatic text synthesis/analysis.

Pronominal CLs are dealt with in §1 (pp. 114–22), whereas the focus of §2 (pp. 122–26) is on verbal CLs. For both CL types, the author provides information on inflectional categories and the corresponding inflectional values (pp. 114–16, 123–24), a morphological description of the CLs (pp. 119–22, 125–26), including a discussion of their unique morphological properties (pp. 116–19).

3.7. Chapter 7: Serbian Clitics as a Challenge for Linguistic Theory

In the seventh and final chapter (pp. 129–50), the author raises two theoretically challenging questions concerning Serbian verbal CLs. Section 1 (pp. 129–37) deals with their unclear syntactic status as either heads or dependents, while the focus of §2 is Serbian future-tense markers that behave as both wordforms and affixes (pp. 137–49). They are followed by a short summary in §3 (pp. 149–50).

Early in Chapter 7, Milićević (p. 129) admits that the answers to the questions are already evident from how the CLs were treated (as lexemes functioning as clausal heads) throughout the monograph and that the two discussions presented in that chapter are merely demonstrations of appropriate diagnostics in ambiguous cases such as those.

The author starts the discussion of the first problem with the observation that, by default, an auxiliary appears in the clause as a CL, i.e., as a prosodically deficient element (p. 129). It is placed with respect to a host and seemingly lacks a key property of clausal heads: the ability to serve as the reference point for fully fledged clause elements (pp. 130, 149). However, a CL auxiliary does not differ from a full form as regards controlling the distribution of the phrase corresponding to a surface-syntactic subtree $V_{(Aux)FIN}$—$V_{(Lex)}$ (pp. 131–33). Moreover, it interacts morphologically with elements external to the phrase just like a finite lexical verb does (pp. 135–37). Milićević (p. 149) concludes that according to the formal criteria and informal considerations generally applied (Mel'čuk 1988: 129–31, 138–40; Mel'čuk 2009: 28–35), only the finite auxiliary in its full or CL form, and not the participle (the non-finite lexical verb), can be the governor and, therefore, the syntactic head of the entire clause. In other words, the discussion of the first theoretical problem demonstrates that CL auxiliaries have all the properties of syntactic heads.

To answer the second question of whether Serbian future tense auxiliary CLs are wordforms or affixes, Milićević (pp. 137, 142) uses MTT (Mel'čuk 1993, 1994, 1996, 1997, 1999) as a starting point and additionally applies morphological

criteria proposed by Zwicky and Pullum (1983). While Zwicky and Pullum's (1983) morphonological criteria did not lead to conclusive answers, syntactic criteria provided compelling evidence that Serbian future tense markers are CLs (pp. 149–50). It goes without saying that, for this question, the syntactic criteria which confirmed the CL status of future tense auxiliary CLs are more valuable since the CLITIC ~ AFFIX opposition is more of a syntactic one (p. 150).

4. Language Variation

The value of the presented monograph lies not only in its examination of Serbian CLs within MTT but also in its approach to CLs that recognizes the issue of variation. While the primary focus of the book is on the contemporary standard variant, some non-standard features of Serbian CLs receive special attention from the author. For instance, the post-prosodic break and clause-initial placements increasingly present in the language of the press and everyday colloquial speech are very extensively dealt with in Chapter 5 (pp. 108–10). Furthermore, some aspects of language change that affect Serbian CLs, such as further grammaticalization of the conditional[7] and future tense auxiliaries, are also analyzed in the monograph (pp. 123–24). These variation phenomena indicate that Serbian 2P CLs may be moving away from a purely Wackernagel system (p. 152).

In Chapter 2, Milićević (p. 31) openly admits that she does not examine cliticization conditions for the interrogative particle *da li* since, in her opinion, its CL form *li* is becoming outdated in standard Serbian. According to the author, the usage of the interrogative *li* in contemporary Serbian is more "a matter of register and/or individual style" (p. 31). Additionally, in the same chapter, while presenting the rule that the full form of *biti* 'be' is obligatory after an internal prosodic break/punctuation, Milićević (p. 34) comments that it is not always obeyed in journalistic and informal styles. As may be seen from the above, even though she does not directly or exhaustively deal with issues of variation, a reader who is interested in the topic can still obtain valuable information.

In Chapter 5, while discussing local vs. CPC placement, the author directly asserts that she will analyze preference rules that emerge as a function of syntactic context, i.e., only linguistic preferences proper. In contrast, variation emerging from the style/register or sociolinguistic factors (dialectal/idiolectal), are not examined (p. 107).[8] Milićević (p. xiv) comments that CLs in dialects have a different morphological makeup and that some dialects ex-

[7] More precisely, the conditional auxiliary is slowly becoming an invariable particle *bi* in colloquial Serbian (p. 137).

[8] In the Preface, the author emphasizes that she will not examine CLs in dialects (cf. Milićević 2023: xiv).

hibit CL doubling, which is not present in standard Serbian (Mišeska-Tomić 2008; Runić 2014).[9] Once again, the author openly indicates the focus of research, while giving useful hints to those who are interested in variation in the Serbian CL system.

In the Preface, Milićević (p. xiv) clearly states that she will abstain from discussing whether Bosnian, Croatian, Montenegrin, and Serbian are distinct languages or variants of the same language. The author points out that they are all mutually intelligible, comparing them to American and British English. However, she also emphasizes that differences between these four languages may grow in the future due to the current social and political climate (cf. p. xiv). Milićević (p. xiv) underlines that the Serbian CL system has many commonalities with CL systems in Bosnian, Croatian, and Montenegrin, and thus many of the reported findings may also be valid for CLs in these closely related South Slavic languages. Still, she also specifies that some significant differences exist and promises to refer to them where relevant (p. xiv). Milićević (p. xiv) accurately identifies one of the problems stemming from the close similarity of the CL systems in question: at times, native speakers can be uncertain as to which structures with CLs are acceptable or preferable in which CL system. Accordingly, these discrepant acceptability judgments "spill over" into linguistic literature (p. xiv).

Indeed, the author stays true to her promise: CL phenomena that exhibit variation within BCMS standard varieties are dealt with throughout the book, either in the main text, in special remark boxes inserted into the main text, or in footnotes. For readers who are highly interested in variation within BCMS only as it concerns CL phenomena, the remaining part of this section summarizes the crucial observations made by Milićević.

According to Milićević (pp. 72, 74, 152) the most salient differences between CL systems in the closely related South-Slavic languages Bosnian, Croatian, Montenegrin, and Serbian are the types of insertable hosts and preferences regarding their usage (Ćavar 1996: 58; Popović 1997: 306; Kolaković et al. 2022: 417). The author also emphasizes that many types of split phrases, now obsolete in contemporary Serbian, can be found in older Serbian texts, for instance in the literary style (pp. 72, 74).

In Chapter 3 on the Serbian CL cluster, Milićević (pp. 41–60) touches on several CL combinations possible in Croatian that are either not present in standard Serbian at all or are found in very limited contexts only. One of them is the combination CL_{ACC} + *se* (REFL), which exists in Bosnian and Croatian beyond CPC contexts. This is due to a special grammatical voice present in these languages, the subjectless suppressive from transitive verbs, marked with the reflexive adjunct *se* (pp. 49–50; Mel'čuk 2006: 181–263), not present in

[9] See Mrazović and Vukadinović (2009: 345) for more information on the emphatic expressions that are the closest to CL doubling in standard BCMS variants.

contemporary standard Serbian. The author also indicates that some Serbian native speakers find the [*se je*] (REFL AUX.3SG) CL sequence acceptable. However, she emphasizes that contemporary standard Serbian greatly inclines toward its circumvention. In contrast, the CL sequence [*se je*] is common in contemporary standard Croatian (p. 56). In Chapter 3, Milićević also acknowledges that in Croatian, unlike in standard Serbian, the accusative feminine singular CL *ju* can appear in contexts other than the allomorph (suppletion) rule context triggered by the third-person singular verbal CL *je* directly following it (p. 53; Kolaković, Jurkiewicz-Rohrbacher and Gradischnig 2023).

In Chapter 5, the author states that, unlike Croatian and Bosnian, Serbian exhibits more restricted use of the infinitive, preferring the competing *da*-complement (p. 111). Milićević (pp. 94–96) argues that in Serbian, V_{INF} is most naturally used to complement (quasi-)modal, phase, and emotional state verbs. In contrast, its application in Serbian with verbs from other classes, particularly object control, causative, and motion verbs, is dated. Moreover, nowadays it triggers a "Croatian feel", thus making the *da*-complement more preferred. This has tremendous consequences on CPC, which, though obligatory in restructuring contexts, is not a significant phenomenon in Serbian as compared to Croatian (pp. 90, 94–96, 111). It has to be pointed out that, according to the author's intuition, CPC out of *da*-complements is only marginally possible in Serbian literary and journalistic registers, and presumably more common in Bosnian (p. 97). The reader is offered two Serbian examples of the phenomenon, one from the literary register and the other from the press, and another from Bosnian, spontaneously produced in a conversation (p. 97).

In Chapter 6, in the subsection describing the morphology of personal pronouns, Milićević (p. 119) comments that unlike in Serbian, the dative reflexive CL form *si* is still used in Croatian. Nevertheless, she acknowledges that she cannot give any information on how frequent it is. In the same chapter, in the subsection dealing with forms of auxiliary verbs, using Belić's (1962: 6) and Radanović-Kocić's (1988: 47) observations as a point of departure, the author notes that, unlike in Croatian, the use of the third-person singular auxiliary stressed form *jè* is possible in Serbian only in polar questions marked with the CL interrogative particle *li* (p. 123). In colloquial Croatian, this form can also be used to answer polar questions. The standard Croatian third-person singular auxiliary form *jȅst*, specialized for this function, is present in Serbian only in some phraseological expressions, such as *to jȅst* 'that is' (p. 123).

5. Concluding Remarks on Milićević's (2023) Value and Originality

The volume of existing literature on CLs is immense: Nevis et al. (1994) list 1,550 bibliographical entries on CLs between 1892 and 1991. Moreover, research on CLs in world languages has been particularly intensive over the last 30 years, and so the number of studies must have greatly increased by now.

Needless to say, Serbian pronominal and verbal CLs occupy a salient place in this line of linguistic research. First and foremost, because they are of the Wackernagel, i.e., 2P CL, type (p. xi). As a brief overview of the existing body of research reveals, this CL subtype is especially interesting to theoretical linguistics. Nevertheless, as the reviewed monograph demonstrates, there is still much left to say about these not very "well behaved" words that are similar to affixes.

I personally consider Milićević's book a major contribution not only to the study of Slavic languages in particular but also to linguistics in general. As already stated at the beginning of this review, Milićević's MTT approach to morphology and syntax of Serbian CLs, and particularly its clear focus on syntactic dependencies and orientation toward speech production, is precisely what makes this monograph rare and exceptional. Consequently, *Serbian clitics* will be a primary reference source for researchers interested in studying CLs in the context of syntactic dependencies. Beyond that, the monograph will be thought-provoking for theoretical linguists specializing in formal descriptions of (Balto-)Slavic morphology and syntax as well as for general linguists, typologists, and Slavists. Although the author primarily examines the CL system in contemporary standard Serbian, it is safe to assume that many comments on various aspects of variation could arouse the interest of variationists. Finally, since CLs are challenging for language learners, Milićević's work might be useful to applied linguists and advanced learners of Serbian.

References

Belić, Aleksandar. (1962) *Istorija srpskohrvatskoga jezika*. Belgrade: Naučna knjiga.

Browne, Wayles. (1974) "On the problem of enclitic placement in Serbo-Croatian". Richard D. Brecht and Catherine V. Chvany, eds. *Slavic transformational syntax*. Ann Arbor: University of Michigan, 36–52. [Michigan Slavic Materials, 10.]

_____. (1975) "Serbo-Croatian clitics for English-speaking learners". Rudolf Filipović, ed. *Kontrastivna analiza engleskog i hrvatskog ili srpskog jezika*. Zagreb: Institut za lingvistiku Filozofskog fakulteta, 105–34. Reprinted in *Journal of Slavic linguistics* 12(1–2)(2004): 249–83.

Čamdžić, Aneta, and Richard Hudson. (2002) "Serbo-Croat-Bosnian clitics and word grammar". *UCL Working Papers in Linguistics* 14: 321–53.

Ćavar, Damir. (1996) "On cliticization in Croatian: Syntax or prosody?" *ZAS papers in linguistics* 6: 51–65.

Gross, Thomas. (2014) "Clitics in dependency morphology". Kim Gerdes, Eva Hajičová, and Leo Wanner, eds. *Dependency linguistics: Recent advances in linguistic theory using dependency structures*. Amsterdam/Philadelphia: John Benjamins, 229–53. [Linguistics Today, 215.]

Hana, Jiří. (2007) *Czech clitics in higher order grammar*. PhD dissertation, Ohio State University.

Iordanskaja, Lidija. (1982) "Le placement linéaire des clitiques pronominaux en français contemporain". *Lingvisticae Investigationes* 6(1): 145–88.

Kolaković, Zrinka, Edyta Jurkiewicz-Rohrbacher, Björn Hansen, Dušica Filipović Đurđević, and Nataša Fritz. (2022) *Clitics in the wild: Empirical studies on the microvariation of the pronominal, reflexive and verbal clitics in Bosnian, Croatian and Serbian*. Berlin: Language Science Press.

Kolaković, Zrinka, Edyta Jurkiewicz-Rohrbacher, and Jasmin Denise Gradischnig. (2023) "Akuzativne zamjeničke zanaglasnice *ju* i *je* u bosanskome, hrvatskome i srpskome: kontrastivno korpusno istraživanje standardnih i razgovornih idioma". *Jezikoslovlje* 24(1): 27–78.

Mel'čuk, Igor A. (1988) *Dependency syntax: Theory and practice*. Albany, NY: State University of New York Press.

_____. (1993) *Cours de morphologie générale.* Vol. 1, *Introduction et première partie: Le mot.* Montréal/Paris: Les Presses de l'Université de Montréal/CSRS Éditions.

_____. (1994) *Cours de morphologie générale.* Vol. 2, *Deuxième partie: significations morphologiques.* Montréal/Paris: Les Presses de l'Université de Montréal/CSRS Éditions.

_____. (1996) *Cours de morphologie générale.* Vol. 3, *Troisième partie: moyens morphologiques. Quatrième partie: syntactiques morphologiques.* Montréal/Paris: Les Presses de l'Université de Montréal/CSRS Éditions.

_____. (1997) *Cours de morphologie générale.* Vol. 4, *Cinquième partie: signes morphologiques.* Montréal/Paris: Les Presses de l'Université de Montréal/CSRS Éditions.

_____. (1999) *Cours de morphologie générale.* Vol. 5, *Sixième partie: modèles morphologiques, et septième partie: principes de la description morphologique.* Montréal/Paris: Les Presses de l'Université de Montréal/CSRS Éditions.

_____. (2006) *Aspects of the theory of morphology*. Berlin/New York: Mouton de Gruyter.

_____. (2009) "Dependency in natural language". Alain Polguère and Igor Mel'čuk, eds. *Dependency in linguistic description*. Amsterdam/Philadelphia: John Benjamins, 1–109. [Studies in Language Companion Series, 111.]

Milićević, Jasmina. (1999) "Pronominal and verbal clitics in Serbian: A morphological description". *Wiener Slawistischer Almanach* 43: 231–56.

_____. (2005) "Clitics or affixes? On the morphological status of the future tense markers in Serbian". Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer, and Franz Rainer, eds. *Morphology and its demarcations: Selected papers from the 11th Morphology Meeting, Vienna, February 2004*. Amsterdam/Philadelphia: John Benjamins, 39–52.

Milićević, Jasmina. (2007) "Co-occurrence of Serbian second-Position clitics: Syntactic and morphonological constraints". Nabil Nathout and Fabio Montermini, eds. *Morphologie à Toulouse. Actes du colloque international de morphologie 4e Décembrettes*. Munich: Lincom Europa, 99–121. [Lincom Studies in Theoretical Linguistics, 37.]

_____. (2009a) "Linear placement of Serbian clitics. A description within a dependency-based approach". Alain Polguère and Igor A. Mel'čuk, eds. *Dependency in linguistic description*. Amsterdam/Philadelphia: John Benjamins, 235–76. [Studies in Language Companion Series, 111.]

_____. (2009b) "Serbian auxiliary verbs—syntactic heads or dependents?" Wladyslaw Cichocki, ed. *PAMAPLA 31/ACALPA 31: Proceedings of the 31st Annual Conference of the Atlantic Provinces Linguistic Association*. Fredericton, New Brunswick, Canada: University of New Brunswick and Saint Thomas University, 43–53.

_____. (2014) "Ste normalni 'Are [you-pl] normal'? Or what happened to those enclitics?" Paper presented at the 9th Annual Meeting of the Slavic Linguistics Society (SLS 9), Seattle, Washington, 19–21 September 2014.

_____. (2019a) "Cliticization of Serbian personal pronouns and auxiliary verbs. A dependency-based account". Kim Gerdes and Sylvain Kahane, eds. *Proceedings of the 5th International Conference on Dependency Linguistics (DepLing/SyntaxFest, Paris, 26–30 August 2019)*. Paris: Association for Computational Linguistics, 60–69.

_____. (2019b) "Clitic pseudo-climbing from dependent infinitive phrases in Serbian". *Journal of Slavic linguistics* 27(1): 27–55.

Mišeska-Tomić, Olga. (2008) "Variation in clitic-doubling in South Slavic". Sjef Barbiers, Olaf Koeneman, Marika Lekakou, and Margreet van der Ham, eds. *Microvariation in syntactic doubling*. Bingley: Emerald, 443–68.

Mrazović, Pavica, and Zora Vukadinović. (2009) *Gramatika srpskog jezika za strance*. Sremski Karlovci/Novi Sad: Izdavačka kuća Zorana Stojanovića.

Nevis, Joel, Brian Joseph, Dieter Wanner, and Arnold Zwicky. (1994) *Clitics. A comprehensive bibliography 1892–1991*. Amsterdam/Philadelphia: John Benjamins.

Popović, Ljubomir. (1997) "Raspoređivanje zamenič kih i glagolskih enklitika". Ljubomir Popović, ed. *Red reči u rečenici*. Belgrade: Društvo za srpski jezik i književnost Srbije, 283–365.

Radanović-Kocić, Vesna. (1988) *The grammar of Serbo-Croatian clitics: A synchronic and diachronic perspective*. PhD dissertation, University of Illinois at Urbana-Champaign.

Runić, Jelena. (2014) *A new look at clitics, clitic doubling, and argument ellipsis: Evidence from Slavic*. PhD dissertation, University of Connecticut.

Stefanowitsch, Anatol. (2020) *Corpus linguistics*. Berlin: Language Science Press. [Textbooks in Language Sciences, 7.]

Toman, Jindřich. (1986) "Cliticization from NPs in Czech and comparable phenomena in French and Italian". Hagit Borer, ed. *The syntax of pronominal clitics*. New York: Academic Press, 123–45. [Syntax and Semantics, 19.]

Utvić, Miloš. (2011) "Annotating the corpus of contemporary Serbian". *INFOtheca: Journal of informatics & librarianship* 12(2): 36a–47a.

Vitas, Duško, Cvetana Krstev, and Gordana Pavlović-Lažetić. (2000) "Recent results in Serbian computational lexicography". Neda Bokan, ed. *Proceedings of the symposium "Contemporary Mathematics", 18–20 December 1998*. Belgrade: University of Belgrade, 113–30.

Zwicky, Arnold, and Geoffrey Pullum. (1983) "Cliticization vs. inflection: English *n't*". *Language* 59(3): 502–13.

Zrinka Kolaković
Alpen-Adria-Universität Klagenfurt
Institut für Slawistik
Klagenfurt am Wörthersee, Austria
Zrinka.Kolakovic@aau.at