Hyperspacing the Verb

The interplay between prosody, morphology, syntax and semantics in the Western South Slavic verbal domain

Boban Arsenijević
Franc Lanko Marušič
Stefan Milosavljević
Petra Mišmaš
Marko Simonović
Rok Žaucer

No series description provided

Hyperspacing the Verb

The interplay between prosody, morphology, syntax and semantics in the Western South Slavic verbal domain

Boban Arsenijević

Franc Lanko Marušič

Stefan Milosavljević

Petra Mišmaš

Marko Simonović

Rok Žaucer

Boban Arsenijević, Franc Lanko Marušič, Stefan Milosavljević, Petra Mišmaš, Marko Simonović & Rok Žaucer. 2025. *Hyperspacing the Verb: The interplay between prosody, morphology, syntax and semantics in the Western South Slavic verbal domain*. Berlin: Language Science Press.

This title can be downloaded at:
redefine\lsURL

ISBN: no digital ISBN
no print ISBNs!
no DOI
ID not assigned!

Cover and concept of design: Ulrike Harbort
Fonts: Libertinus, Arimo, DejaVu Sans Mono
Typesetting software: XeLaTeX

redefine \publisherstreetaddress
redefine \publisherurl

Storage and cataloguing done by redefine \storageinstitution

Contents

Acknowledgments v

Abbreviations 171

1 Introduction 1
- 1.1 Structure of the monograph . . . 2
- 1.2 Contributions made to the linguistic theory . . . 3
  - 1.2.1 The status of theme vowels . . . 4
  - 1.2.2 The inventory and the feature content of verbal suffixes 4
  - 1.2.3 The status of (Western South) Slavic prefixes . . . 5
  - 1.2.4 The nature of deverbal affixation . . . 7
  - 1.2.5 Multifunctionality in morphology . . . 7
  - 1.2.6 Verbal lexical prosody . . . 9
- 1.3 A note on examples and glosses . . . 9

2 Description of the database 11
- 2.1 Materials and infrastructure . . . 11
- 2.2 Verbal inflection . . . 12
  - 2.2.1 Theme vowels . . . 13
  - 2.2.2 Root allomorphy . . . 20
  - 2.2.3 Prosodic properties . . . 21
- 2.3 Verbal derivation . . . 28
  - 2.3.1 Roots . . . 28
  - 2.3.2 Prefixes . . . 37
  - 2.3.3 Compounds . . . 40
  - 2.3.4 Suffixes . . . 41
  - 2.3.5 Morphological complexity of the verb . . . 44
- 2.4 Deverbal derivation . . . 44
- 2.5 Verbal aspect . . . 47
- 2.6 Summary . . . 48

3 The domain of verbal inflection 49
- 3.1 Introduction . . . 49

3.2 Theme-vowel classes in WSS and interaction with further pieces of morphology . . . 51
3.3 Prosody of WSS verbs . . . 56
3.3.1 Prosody of Slovenian verbs . . . 59
3.3.2 Prosody of BCMS verbs . . . 61
3.4 Conclusions . . . 63

4 Properties of verb derivation **67**
4.1 Introduction . . . 67
4.2 Basic statistics and how it can be informative . . . 67
4.3 Quantitative insights into verbal prefixes . . . 73
4.3.1 Frequency of root-adjacent prefixes . . . 74
4.3.2 Prefix stacking . . . 75
4.3.3 Prefixes and other affixes . . . 80
4.4 Summary . . . 94

5 Deverbal derivation **99**
5.1 Introduction . . . 99
5.2 Method . . . 100
5.3 Results, BCMS . . . 103
5.4 Discussion, BCMS . . . 109
5.4.1 Theme vowels . . . 109
5.4.2 Frequency . . . 110
5.4.3 Compatibility with an accusative argument . . . 111
5.4.4 Meaning shift with the reflexive marking . . . 111
5.4.5 Imperfectivity . . . 112
5.4.6 Prefixed verbs . . . 112
5.4.7 The base verb includes a verbal suffix . . . 113
5.5 Results, Slovenian . . . 113
5.6 Discussion, Slovenian . . . 117
5.6.1 Theme vowels . . . 117
5.6.2 Frequency . . . 118
5.6.3 Compatibility with an accusative argument . . . 119
5.6.4 Meaning shift with the reflexive marking . . . 119
5.6.5 Imperfectivity . . . 120
5.6.6 Prefixed verbs . . . 121
5.6.7 The base verb is derived by suffixation . . . 121
5.7 Discussion, comparison between the two languages . . . 122
5.8 Overall discussion . . . 124

5.9 BCMS suffix *-aj* 124
5.10 Comparing results obtained by two different annotation methods 130
5.11 Conclusion 136

6 Multifunctional suffixes **139**
6.1 Introduction 139
6.2 Definition 139
6.3 Treatment of multifunctional affixes in *WeSoSlaV* 142
6.4 Three case studies 146
6.4.1 Example 1: The suffix *-ov* 146
6.4.2 Example 2: The suffix *-k* 157
6.4.3 Example 3: the suffix *-š* 162
6.5 Conclusion 165

7 Conclusion **167**

Abbreviations **171**

References **173**

Index **185**
Name index 185

Acknowledgments

The present monograph is a result of the project *Hyperspacing the Verb: The interplay between prosody, morphology and semantics in the Western South Slavic verbal domain* funded by the grant FWF I 4215 and ARRS N6-0113, which took place between October 2019 and September 2024 at the University of Graz and the University of Nova Gorica. The project, as indicated by the title, focused on a subpart of the verbal systems in two Slavic languages, Slovenian and Bosnian/Croatian/-Montenegrin/Serbian (BCMS). Part of the monograph resulted from the research within a follow-up project titled *Multifunctionality in morphology* (ARIS J6-4614, FWF I6258). The monograph also benefited from the research program funding scheme (ARIS P6-0382). The monograph has also greatly benefited from the contributions of Jelena Simić, Ema Štarkl, Predrag Kovačević, Katarina Gomboc Čeh, and the members of the reading group *From Morphophonology to Morphosyntax and back*.

... The reviewer who hates the phrase hotly debated.

Abbreviations

1	first person
2	second person
3	third person
ACC	accusative
ADJ	adjective
BCMS	Bosnian/Croatian/ Montenegrin/Serbian
DIM	diminutive
F	feminine
GEN	genitive
XXX^{IPFV}	imperfective
INF	infinitive
INS	instrumental
IPFV	imperfective
M	masculine
N	neuter
NEUT	neuter
NOM	nominative
XXX^{PFV}	perfective
PFV	perfective
PL	plural
PRS	present
PTCP	participle
REFL	reflexive
SG	singular
SI	secondary imperfective
S.INF	short infinitive
TV	thematic vowel
WSS	Western South Slavic

1 Introduction

Due to their high degree of grammaticalization of aspect, argument structure, and case, as well as elaborate verbal morphology, Slavic verbal systems have played a very prominent role in the research of the internal structure and semantics of the verb. In this monograph, we primarily focus on the quantification of the properties typically observed and reported in the theoretical literature on Slavic verbal systems with the goal of providing a clearer picture of the empirical magnitude of various phenomena that were instrumental in shaping the theory of verbal morphology, syntax, and semantics, in order to also better model their place in these theories. The central focus, which is at the same time at the core of a more general research program, is to investigate the interactions between semantics, morphology, and phonology through a systematic quantitative exploration of the (de)verbal domain in Western South Slavic languages (WSS).

In our research, we aim to address an array of long-established views in the linguistics of the (Slavic) verbal system that are often taken for granted (e.g., simplex verbs in Slavic are imperfective), to weigh in on often debated issues (e.g., how many and what types of prefixes do verbs in Slavic take), to contribute to ongoing debates (e.g., what is the nature of theme vowels, in Slavic and generally), to answer enduring empirical questions (e.g., are *-je* nominalizations only possible from imperfective verbs).

The implemented method was to establish a large database of verbs and their relevant properties and use it to observe systematic correlations between morphological (e.g., the continuum between inflection and derivation), semantic (e.g., interpretive transparency vs. idiosyncrasy), and prosodic properties (e.g., the surfacing of the prosodic prominence of the base, of an affix, or resorting to the default). One key output of this work is the database itself, which we refer to as the *Annotated Database of the Western South Slavic Verbal System*, shorter *WeSoSlaV*. The database, which is to the best of our knowledge the first richly manually annotated database of the morphological, syntactic, semantic, and prosodic properties of the verbal system of one or more languages, lends itself to the research on a variety of research questions and can be assessed by a broad range of quantitative tests and models, and is freely available online to all researchers in (Slavic) morphology. Based on the database, several publications that address specific

research questions (many of which are referred to throughout the monograph) were produced. However, these works only scratch the surface of the various dimensions along which the database can be used in research.

The monograph has two goals: 1) to describe WeSoSlaV and 2) to report the results of a collection of theoretical investigations of WSS verbal systems informed by a systematic quantitative empirical exploration of the database. In other words, the monograph aims to equip readers with the necessary knowledge to utilise WeSoSlaV in their research while simultaneously presenting the findings of the investigations conducted by the database's authors. In the remainder of this introductory chapter, we outline the structure of the monograph (Section 1.1), and the contributions it makes to the current linguistic theory (Section 1.2).

1.1 Structure of the monograph

The monograph is structured as follows. Chapter 2 presents the main properties of the *Database of the Western South Slavic Verb HyperVerb 2.0 – WeSoSlaV* (Arsenijević, Gomboc Čeh, et al. 2024, referred to as *WeSoSlaV* in what follows): the procedure and problems of selecting the verbs to be included, the technical issues in its development, the solutions opted for and the rationale behind them. It lists the annotated properties, outlines why some prominent verbal properties could not be included, and indicates properties that have not, but can be included in future work on the database.

The subsequent four chapters are dedicated to the four areas that the annotated properties are divided into: inflection, verbal derivation, deverbal derivation and multifunctional suffixes.

In Chapter 3, the properties of verb inflection of BCMS and Slovenian are assessed from the perspective of the database. After an overview of the field, the chapter focuses on the segmental content of the theme vowels and its interaction with the inflectional material, as well as on the prosodic properties of verbs in an attempt to answer the question of whether the verbs are prosodically specified on the root, on the theme vowel, on the inflection, on some combination of these morphological components, or by some superimposed procedure.

In Chapter 4, we present the verbal derivational system in BCMS and Slovenian as annotated in *WeSoSlaV*. We offer examples of basic descriptive statistics in the two languages and how they are or can be employed to draw theoretical conclusions. We eventually focus on the properties of WSS verbal prefixes as documented in *WeSoSlaV* and the theoretical potential of the quantitative observations in this domain.

Chapter 5 is concerned with the potential of *WeSoSlaV* to inform the investigation of deverbal derivations. We show one possible way of utilizing the database for this purpose, by reporting and discussing the results of an investigation involving the enrichment of *WeSoSlaV*, and focusing on the question what determines the capacity of a verb to undergo deverbal derivation by each from a selected set of suffixes. We also draw some conclusions informing future research and briefly summarize previous works that use *WeSoSlaV* for research on deverbal derivations.

Chapter 6 centers around the multifunctionality of suffixes. While it focuses on the multifunctionality as a property that was annotated in the database, the chapter primarily deals with properties of the suffixes that are already discussed in other chapters, but approaches them in light of multifunctionality. Three concrete items that were annotated as multifunctional in the database are presented, suffixes *-ov*, *-k* and -š, and their characteristics that figure in the perspective of multifunctionality, including issues that emerged in the process of annotation, are discussed in more detail.

Chapter 7 concludes the monograph.

1.2 Contributions made to the linguistic theory

The contribution of the monograph to the current linguistic theory is both empirical and theoretical. Empirically, it provides a powerful benchmark for testing and developing theoretical models of various aspects of WSS verbal systems. Introducing *WeSoSlaV* and basing our research on it enables a quantitative exploration of a wide range of linguistic questions using a representative and extensive body of manually annotated data. In particular, the database allows for a reliable estimation of the size, productivity and mutual interactions of various classes of verbs and morphological operations involving them. Quantification is enabled for universal, negative and implicational generalizations in this domain. Importantly, *WeSoSlaV* encompasses two closely related languages, thus also enabling multifaceted research in variation and microvariation. Many of these possibilities are illustrated throughout the monograph.

In the remainder of this section, we focus on some important findings that carry theoretical significance for the investigations of the verb in Slavic and beyond. These include the nature and theoretical status of theme vowels (Section 1.2.1), the inventory and feature content of verbal suffixes (Section 1.2.2), the status of prefixes, with a focus on prefix stacking (Section 1.2.3), the nature of deverbal affixation (Section 1.2.4), multifunctionality in morphology (Section 1.2.5), as well as novel theoretical findings regarding verbal lexical prosody (Section 1.2.6).

1.2.1 The status of theme vowels

The status of theme vowels (TVs) in the architecture of grammar is a matter of debate. In some approaches, they figure only at the level of morphology, i.e. more broadly at the Phonological Form, PF (e.g. Anderson 1992, Aronoff 1994, Oltra-Massuet 1999, 2020, Marvin 2002), and as such constitute one of the main pieces of evidence for morphology as an independent module of grammar in the so-called morphomic approaches (Aronoff 1994). By contrast, it has been argued in Svenonius (2004a), Jabłońska (2007), Biskup (2019), Milosavljević & Arsenijević (2022), Kovačević et al. (2024) that TVs do carry syntactic and/or semantic information, i.e. act as verbalizers. There are also approaches arguing that Slavic TVs do not form a homogeneous category, but can be split into verbalizers (or derivational affixes) and 'proper' TVs, which may be purely morphological items, or ornamental morphemes (Gribanova 2015, Simonović & Mišmaš 2022, Matushansky 2024).

Two chapters of the monograph argue in favour of treating TVs as morphemes that are not purely morphological. In Chapter 3, we show that there are significant correlations between TV classes and prosodic patterns, indicating that TVs interact very closely with the representation of roots and other morphemes—a type of behaviour that would not be expected if they were purely ornamental elements, often assumed to be adjuncts. This conclusion is supported both by divergent tendencies across different TV classes and by multiple cases of categorical incompatibility between particular TV classes and prosodic patterns, as demonstrated in the chapter. We also present a novel observation for BCMS: TV classes differ not only with respect to prominence, but also with respect to stem-final vowel length. Chapter 4 identifies significant correlations between TVs and both verbal prefixes and suffixes. Taken together, these correlations between TVs and other pieces of verbal morphology—alongside previous work largely building on *WeSoSlaV* (e.g. Milosavljević & Arsenijević 2022, Kovačević et al. 2024), which demonstrates semantic and syntactic correlations among different TV classes—provide strong support for analysing TVs as grammatical in nature, extending beyond a purely morpho(-phonological) role.

1.2.2 The inventory and the feature content of verbal suffixes

There are roughly three types of verbal suffixes in Slavic languages (as also annotated in *WeSoSlaV*; see Chapter 2): those that derive imperfective verbs from perfective ones, those that derive verbs from non-verbal bases (i.e. roots, nouns, adjectives, or phrases), and those that contribute additional content, such as

diminution. Focusing on their interaction with other morphological elements reveals that elements traditionally analysed as verbal suffixes can be decomposed into a TV and the material that precedes it. For instance, the traditional Slovenian imperfectivizer *-eva*, should be reanalyzed as *-ev* and the TV *-a*. Verbal suffixes are therefore morphologically, and arguably also syntactically and semantically, complex in their underlying structure. Such a decomposition has in fact been argued for in previous research, including that based on *WeSoSlaV*. Simonović et al. (2023), Arsenijević et al. (2025) show that the productive imperfectivizing suffixes *-iva* and *-ava* in BCMS both begin and end in vowels that match the two most productive TVs in BCMS, *i* and *a*. Consequently, they argue, that the secondary imperfective suffixes in BCMS are in fact a combination of two TVs (and a glide, which is an exponent of floating features affiliated with the second TV). The semelfactive suffix *-nV-/-ne-*, on the other hand, is a combination of a diminutive morpheme *-nV* and a TV, ∅/e, as detailed in Štarkl et al. (2024), see also Chapter 4 for further discussion. Finally, the BCMS verbalizing affix *-ova-* (in, e.g., *v(j)er-ov-a-ti* 'believe', related to the noun *v(j)er-a* 'faith'), can again be decomposed into the element *-ov-*, which potentially realizes a nominal (or non-verbal) head and the TV a/je (cf. Kovačević et al. 2024). Similar decomposition can also be proposed for other verbal affixes, which then potentially contain a TV preceded by multifunctional affixes found in other categories as well, as discussed in detail in Chapter 6.

The key theoretical implication of these findings, in summary, is that they call into question the very existence of specialised *verbal* derivational affixes: these affixes are always decomposable into TVs, which function as verbalisers, and/or multifunctional elements commonly found in other grammatical categories.

1.2.3 The status of (Western South) Slavic prefixes

Among Slavic morphemes, prefixes have likely received the most prominent attention from a theoretical perspective. In particular, the division of prefixes into two classes, lexical (LPs) and superlexical prefixes (SLPs) has figured as one of the most prominent questions of Slavic morphosyntax. LPs are usually argued to merge low in the structure, *v*P/VP-internally, while SLPs are merged higher up, in the aspectual domain (e.g. Svenonius 2004b, Ramchand 2004, Romanova 2004, Gehrke 2008, among many others, though see Arsenijević 2007, Žaucer 2009 for unified approaches to LPs and SLPs, and Milosavljević 2023b for a recent overview of different approaches). Multiple prefixation in Slavic refers to the possibility of verbal prefixes stacking over each other, as in *na-do-dati* (on-to-give) 'to add a lot of' from BCMS. Prefix stacking has been used as one of

the main diagnostics for the difference between lexical and superlexical prefixes, in that it is usually claimed that only SLPs can stack (cf. Svenonius 2004b, Milićević 2004, Istratkova 2004, Markova 2011, among many others). This diagnostic has been questioned in the literature, as some lexical-like prefixes can stack too (cf. Žaucer 2009, 2013, Marušič et al. 2025 for Slovenian, Milosavljević 2023b for BCMS, Zinova 2021 for Russian). Chapter 4 presents quantitative data from BCMS and Slovenian regarding the status of stacked prefixes. The quantitative patterns observed with prefix stacking in both languages provide support for the approaches arguing that prefix stacking cannot be taken as a clear-cut diagnostic for distinguishing between LPs and SLPs. That is, without distinguishing between specific uses of individual prefixes, almost all prefixes that function as the first prefix can also function as the second one (counting from the root), i.e., all except the archaic *vaz-* in BCMS and *pod-* in Slovenian. This includes prefixes that are typically considered lexical or that are never found on lists of superlexical prefixes but apparently also occur as the second, stacked prefix: *u-* 'in', *s-* 'from', *o-* 'about', *ob-* 'around', *raz-* 'apart', *uz-*/*vz-* 'up', and in BCMS also *pod-* 'under'.

In addition to LPs and SLPs, the status of the so-called purely perfectivizing prefixes (PPPs) is still debated as well. PPPs are traditionally conceived to be prefixes that turn an imperfective verb into a perfective one without affecting the lexical meaning, i.e., without any other contribution apart from the pure aspectual effect. The status of PPPs is controversial, with some scholars arguing that they are more similar to LPs (e.g. Gehrke 2008, Endresen et al. 2012), and others analyzing them as SLPs (e.g. Ramchand 2008). Endresen et al. (2012) provide quantitative evidence from Russian that PPPs are not actually purely perfectivizing, but their lexical semantics gets bleached due to the overlapping semantics with the verb they combine with. While in both BCMS and Slovenian various prefixes can function as PPPs, Dickey (2012) argues that in Slovenian, unlike in BCMS, there is a grammaticalized prefix that functions as a pure perfectivizer: *s-*/*z-*. The continuous decline in frequency of prefixes reported in Chapter 4 can be taken as quantitative evidence against singling out any of the prefixes in BCMS or Slovenian as a purely grammatical perfectivizing prefix. In Slovenian, *po-* stands out slightly, but not with a difference that would support the status of a grammaticalized perfective marker. Moreover, Dickey actually identifies *s-*/*z-* as the Slovenian partially grammaticalized empty perfectivizer, rather than *po-* which displays the highest frequency.

Finally, the quantitative patterns presented in Chapter 4 show significant interactions between prefixes and other affixes. Namely, the frequency and distribution of prefixes vary depending on the type of theme vowel classes and types

of suffixes. The full evaluation of these interactions and their theoretical significance must await future explorations.

1.2.4 The nature of deverbal affixation

Deverbal derivations are investigated in Chapter 5, which considers possible ways to expand *WeSoSlaV* beyond its original scope and addresses methodological questions about its development. Specifically, it demonstrates how, with a specific extension, the database can be used to investigate deverbal derivation in BCMS and Slovenian. A selected subset of annotated properties, including some that were not included in the main published body of *WeSoSlaV* but in auxiliary databases, is examined for their influence on the verbs' ability to enter various derivational processes. The findings show that most properties have an effect on some suffixes, with certain properties showing consistent effects across suffixes, while others vary. It is likely that unexamined properties would also follow one of these two patterns. Additionally, the chapter highlights systematic differences in how theme vowel classes influence derivation, suggesting that these classes represent non-arbitrary groupings of verbs with shared underlying characteristics.

Further research could expand into investigating interactions between annotated verbal properties in affecting derivations, in addition to the examined main effects, or to explore further deverbal suffixes, or possibly to focus on specific details like base suffix deletion triggered by deverbal derivation.

1.2.5 Multifunctionality in morphology

Multifunctional affixes are those that appear in different structural/categorial environments, such as English *ic*, which appears in both nouns (*a/the comic*) and adjectives (*a/the comic tone*). They pose theoretical challenges to both lexicalist and (mainstream) realizational models of morphology. In the former, all lexical items carry a categorial specification, and it is thus unexpected to have items that appear across categories, and whose meaning is often difficult (or impossible) to determine without the structural context in which they appear (thus evading the traditional notion of a morpheme as a pairing of form and sound). For more recent syntactocentric approaches to morphology, such as Distributed Morphology (Halle & Marantz 1993), the semantic variability may be captured by the notion of allosemy – the semantic interpretation being parametrized on the context (Marantz 2013, Wood & Marantz 2017). However, even if semantically flexible, in mainstream Distributed Morphology, affixes are still taken to strictly

spell out syntactic features, including categorial heads, so categorial ambiguities in affixes are completely unexpected, as noted in Lowenstamm (2014).

Multifunctionality is one of the cornerstones of the grammatical architecture known as the Universal-Spine Hypothesis (Wiltschko 2014). For Wiltschko, multifunctionality is one of the core properties of human grammar, and all types of language items (*units of language* or *UoLs* in her terminology) can be multifunctional: for each UoL, an underspecified semantic core can be determined, and its more specific functions are determined through the interaction with different layers on the syntactic spine. This holds for traditional roots, functional items, affixes. Another approach in which multifunctionality (or syncretism, conceived as a subtype of multifunctionality) is crucial is Nanosyntax (Caha 2009, Starke 2009). In this approach, lexical trees can lexicalize multiple chunks of syntactic structure as long as these chunks share a subpart of syntactic structure, i.e. bigger chunks structurally contain the smaller ones. Both the Universal-Spine Hypothesis and Nanosyntax rely on clearly identifiable semantic or syntactic features, which predicts that all multifunctionality items should have some common core. While this is theoretically preferable, the overview in Chapter 6, along with the general inventory of multifunctional affixes in the database WeSoSlaV, suggests the existence of potentially semantically empty affixes. Their distribution is so versatile that it challenges the viability of identifying any consistent semantic core — though pursuing such identification should undoubtedly remain a priority on the research agenda for future studies.

The plausible semantic emptiness of multifunctional items identified in *WeSoSlaV*, at least at first glance, aligns closely with the research agenda of the so-called morphomic approaches (Aronoff 1994). These approaches posit the existence of (potentially) meaningless affixes that occur across categories, viewing them as key evidence for the existence of purely morphological entities (the so-called morphomes). This, in turn, supports the idea of morphology as an independent module of grammar, separate from syntax, semantics, and phonology. The similarity is only apparent, however, since, as shown in Chapter 6 and papers cited there (which largely build on the data from *WeSoSlaV*), different behavior of multifunctional affixes, specifically prosodic patterns of words they appear in, is sensitive to phasal spellout, and thus to the phonology-syntax interface, contrary to what one would expect if they are purely morphological entities.

Finally, and equally importantly, Chapter 6, along with the entire database *WeSoSlaV* annotated for multifunctional affixes, provides significant quantitative evidence for the existence of multifunctional items. This evidence demonstrates that these items cannot be dismissed as a small set of potentially explainable

exceptions — a conclusion similar to that of De Belder (2011), who identifies 20% of affixes in Dutch as multifunctional (see Borer 2013 for a critical perspective).

1.2.6 Verbal lexical prosody

The most significant finding regarding the lexical prosody of WSS verbs is that theme-vowel classes have different prosodic patterns, which is why theme vowels may be considered the carriers of lexical prosody of verbs in WSS. This means that we have not identified any evidence for either inflectional affixes or roots (in deradical verbalisations) as carriers of lexical prosody. An important complication of the system, discussed also in previous literature (Simonović 2022b, Simonović & Kager 2023, Simonović 2022a), is constituted by verbs derived from other categories, which are in some cases able to preserve the prosodic pattern of the incorporated noun/adjective. In Section 6.4.1.2 a case study is presented where there is variation between incorporated nouns and theme vowels as determiners of the prosody of the verb.

While the discussion of the source of lexical prosody typically concerns lexical stress and tone, our findings indicate that vowel length on verbs in standard BCMS may also be part of the lexical representation of theme vowels (as a floating mora). As we detail in Chapter 3, virtually all long vowels are adjacent to theme vowels and theme vowel classes display significant differences regarding the distribution of vowel quantity.

1.3 A note on examples and glosses

For the purposes of this monograph, we have made several decisions regarding the presentation and glossing of examples. Firstly, when relevant, information about the perfectivity of the verb is indicated with a superscript placed next to the verb. Secondly, prefixes are glossed using the closest English preposition in terms of meaning. Lastly, not all suffixes are glossed in the same way. Derivational suffixes (in fact, all suffixes that are not purely functional) are not translated into corresponding English suffixes; instead, they are glossed as they appear, in italicized text. Functional/grammatical suffixes, on the other hand, are glossed with grammatical category labels, which are listed at the end of this monograph. Example (1) illustrates this approach.

(1) u-dom-ač-i-ti$^{\text{PFV}}$
in-home-*ač*-TV-INF
'to domesticate'

2 Description of the database

Since the monograph centers around the *Annotated Database of the Western South Slavic Verbal System* (henceforth *WeSoSlaV*), we start with a description of the database. In this chapter, we describe the materials included in *WeSoSlaV*, the properties that were annotated, and provide a preview of the results. For each of the annotated properties (marked in a column or a group of columns in *WeSoSlaV*), we present the employed criteria and tests. We furthermore discuss some challenges encountered during the process of annotation, with alternative solutions, problematic cases, and possible prospects for future work.

This chapter is organized according to the verbal properties annotated in the database. §2.1 gives an overview of the material included in the database, §2.2 shows how verbal inflection was annotated, §2.3 does so for verb derivation. Deverbal derivation is considered in §2.4, and verbal aspect in §2.5. Finally, §2.6 summarizes the chapter.

2.1 Materials and infrastructure

WeSoSlaV consists of two sub-bases: the BCMS and the Slovenian sub-database. While the general goal was to make the two sub-databases as comparable as possible, both in terms of the material and the annotated properties, the sub-databases do exhibit some differences due to the specifics of the included varieties. The first difference is in size. The inclusion criterion for both sub-databases was frequency. Specifically, the top 3000 highest-frequency verbs from each consulted corpus were included and annotated in *WeSoSlaV*. The 3000 most common verbs were determined using the Slovenian National Corpus *Gigafida 2.0* (2019) for Slovenian and *srWaC, hrWaC, bsWaC* and *meWaC* (Ljubešić & Klubička 2014) for Serbian, Croatian, Bosnian and Montenegrin, respectively.[1] This means that for BCMS, 3000 most frequent verbs were extracted from each of the mentioned corpora and the final BCMS sub-database contains the union of the verb lists obtained from the four corpora, 5300 verbs in total. Not included in this total are

[1]These five corpora are part of Clarin.si's infrastructure, that enables the use of NoSketch Engine to search and analyze different corpora.

forms that overlap in all 4 varieties (e.g., *čitati* 'to read' is included only once). On the other hand, different shapes that the same verb can have in different varieties were introduced as separate entries. Some typical examples of variants are Ekavian and Ijekavian versions, such as *verovati* and *vjerovati* 'to believe', or versions emerging from using different suffixes to integrate borrowed verbs, e.g. *lajk-a-ti* and *lajk-ova-ti* 'to like (on social media)'.

Items that were included on the list due to mistakes in annotation in the corpus were excluded from our list and replaced by the next verb on the list of most common verbs. One such example from Slovenian is *Hoče*. *Hoče* is indeed the PRS.3SG form of the verb *hoteti* 'to want', but it is also the name of a Slovenian municipality. Since *hoteti* 'to want' was independently on the list, the form *Hoče* was excluded.

The lists of the most frequent verbs obtained from the corpora naturally include homophonous verbs. Since the used corpora are not annotated for meaning, homophonous verbs are counted as one verb. For example, in Slovenian, the verb *brati* can mean either 'to read' or 'to gather, collect'. In such cases, the annotation in *WeSoSlaV* is based on the homonym considered to be more frequent by the annotator. The same procedure was independently applied to derived verbs, so for instance, *prebrati*, which can mean either 'to finish reading' or 'to pick through' was considered independently of *brati*. Finally, there are also prefixed verbs where the ambiguity of the base verb does not survive. For instance, *odbrati* only ever means 'to collect some items from a set, separate', so in these cases the issue did not arise in the first place.

2.2 Verbal inflection

This section describes how the various aspects of verbal inflection were annotated in the first subpart of *WeSoSlaV* (Marušič et al. 2022), which is exclusively focused on verbal inflection. Specifically, the following properties were included: the dictionary form, i.e. the infinitive, its present tense (annotated as the third person singular form of the verb), theme vowels, root allomorphy, prosodic features. These properties are reflected in individual columns or groups of columns in the database.

The first two columns, the dictionary form and the present tense, can be illustrated by the BCMS verb *čitati* 'to read', and its PRS.3SG form *čita* or the Slovenian pair *delati, dela* 'to work, (s)he works'. The annotation of the remaining columns is described in the following subsections.

2.2.1 Theme vowels

For the purposes of the database, a theme vowel (TV) is taken to be the piece of morphology that precedes the inflectional ending and is traditionally considered as the determiner of the conjugation class. Since the theme vowel in Slovenian and BCMS can differ between different tenses, theme vowel classes are represented as pairs of theme vowel exponents: the first part of the pair corresponds to the theme vowel in the infinitive, whereas the second part is the exponent of the theme vowel in the PRS.3SG form. Consider the following examples (same for BCMS and Slovenian):

(1) a. gled-a-ti$^{\text{IPFV}}$, gled-a-∅$^{\text{IPFV}}$
look-TV-INF, look-TV-3SG
'to look, (s)he looks' TV class: a/a

b. za-drž-a-ti$^{\text{PFV}}$, za-drž-i-∅$^{\text{PFV}}$
behind-hold-TV-INF, behind-hold-TV-3SG
'to keep, (s)he keeps' TV class: a/i

The theme vowel classes identified for BCMS are given in Table 2.1.[2] Table2.2 gives the TV classes in Slovenian as annotated in *WeSoSlaV*. To make a clear distinction between TVs and agreement morphology, in Tables 2.1 and 2.2 we give the PRS.1PL form rather than PRS.3SG form, as the latter has a zero person/number exponent.

In the following two subsections we turn to the issues which arose in the process of annotation of TV classes, mostly relying on the annotation of the BCMS sub-database. In 2.2.1.1 we focus on the overall shape of theme vowel allomorphs and the distinction between theme vowels and derivational suffixes, whereas in 2.2.1.2 we address cases of boundaries and overlaps between TV classes.

2.2.1.1 Defining TV classes and singling out derivational affixes

Our initial annotation TV classes (summarized in Table 2.3 for BCMS) was performed in a completely pre-theoretical manner by simply extracting any sequence found between the inflection and the root (or the last derivational suffix) in the infinitive and the present tense forms. This annotation yielded 13 TV classes.

[2]The segments between brackets are present in some of the regional variants but not in others. The Ekavian theme vowel classes e/i and e/e have Ijekavian correspondents je/i and je/ije, whereby *je* and *ije* are exponent of the yat phoneme, so the more precise names would be ě/i and ě/ě.

Table 2.1: Theme vowel classes in BCMS

TV class	INF	PRS.3SG	GLOSS
a/a	pit-a-ti	pit-a	'ask'
	pro-ʊ(j)er-aʊ-a-ti	pro-ʊ(j)er-aʊ-a	'check'
	domin-ir-a-ti	domin-ir-a	'dominate'
i/i	vis-i-ti	vis-i	'hang'
	od-laz-i-ti	od-laz-i	'leave'
a/je	plak-a-ti	platʃ-e < /plak-je/	'cry'
	trep-t-a-ti	trep-tɕ-e < /trep-t-je/	'blink'
	formul-is-a-ti	formul-iʃ-e < /formul-is-je/	'formulate'
	is-pit-iva-ti	is-pit-uje	'question'
	za-braɲ-iva-ti	za-braɲ-uje	'forbid'
	k-oʊa-ti	k-uje	'mint'
	rezult-oʊa-ti	rezult-uje	'result'
	da-ʊa-ti	da-je	'give'
	sa-zna-ʊa-ti	sa-zna-je	'find out'
∅/e	pas-∅-ti	pas-e	'graze'
	bos-∅-ti < /bod-∅-ti/	bod-e	'stab'
	pi-∅-ti	pij-e	'drink'
	u-mr(ij)e-∅-ti	u-mr-e	'die'
u/e	gur-nu-ti	gur-ne	'push'
(j)e/i	gor-(j)e-ti	gor-i	'burn'
	zr-e-ti	zr-i	'ripen'
∅/ne	sta-∅-ti	sta-ne	'stop'
	pas-∅-ti < /pad-∅-ti/	pad-ne	'fall'
a/i	struj-a-ti	struj-i	'flow'
	za-sp-a-ti	za-sp-i	'fall asleep'
a/e	(x)rʊ-a-ti	(x)rʊ-e	'wrestle'
(j)e/(ij)e	sm-(j)e-ti	sm-(ij)e	'be allowed'
	pro-xt-(j)e-ti	pro-xt-(ij)e	'desire'

As is clear from Table 2.3, stem-final vowels, as well as all other material alternating between the two considered forms, were taken to be the exponents of TVs. In this classification, six TV classes had clearly vocalic exponents, which either alternate between the infinitive and the present tense (e/i[3], a/i, a/e) or have

[3]In this discussion we use the Ekavian versions of the TV classes for simplicity, see fn. 2.

Table 2.2: Theme vowel classes in Slovenian

TV class	INF	PRS.3SG	GLOSS
a/a	del-a-ti	del-a	'work'
	pre-po-zna-ʋ-a-ti	pre-po-zna-ʋ-a	'recognize'
	kop-ir-a-ti	kop-ir-a	'copy'
	dozd-ev-a-ti	dozd-ev-a	'seem'
	po-govar-j-a-ti	po-govar-j-a	'talk'
i/i	del-i-ti	del-i	'share'
	v-stop-i-ti	v-stop-i	'enter'
a/je	or-a-ti	or-je	'plough'
	sij-a-ti	si(j)-je	'shine'
	pis-a-ti	piʃ-e < /pis-je/	'write'
	vz-drʒ-ev-a-ti	vz-drʒ-u-je	'abstain'
	pri-kaz-oʋ-a-ti	pri-kaz-u-je	'show'
∅/e	pas-∅-ti	pas-e	'graze'
	pas-∅-ti < pad-∅-ti	pad-e	'fall'
	od-pre-∅-ti	od-pre	'open'
i/e	mi-ni-ti	mi-ne	'pass'
	iz-leg-ni-ti	iz-leg-ne	'stretch out'
e/i	zven-e-ti	zven-i	'sound'
∅/ne	sta-∅-ti	sta-ne	'cost'
	pri-tʃe-∅-ti	pritʃ-ne	'start'
a/i	b-a-ti	boj-i	'be afraid'
	beʒ-a-ti	beʒ-i	'run away'
a/e	iz-zʋa-a-ti	iz-zoʋ-e	'challenge'
	bra-a-ti	ber-e	'read'
e/e	um-e-ti	um-e	'know'

the same vowel in both forms (a/a, i/i, e/e). The a/je class is almost fully vocalic, as it has a vocalic exponent in the infinitive, but its present-tense TV has an additional palatalizing element realized on the preceding consonant. On the other hand, the ∅/e class does not involve any consonants but it has a zero exponent of the theme vowel in the infinitive. The remaining five TV classes from this initial classification involve stable consonantal material. Two have a [n] element (∅/ne and nu/ne) and three have elements that can be analyzed as glides (va/je, ova/uje and iva/uje). The two classes that have a [n] element had a different treatment

Table 2.3: Theme vowel classes: an initial version

TV class	INF	PRS.3SG	GLOSS
a/a	pit-a-ti	pit-a	‘ask’
	pro-ʊ(j)er-aʊ-a-ti	pro-ʊ(j)er-aʊ-a	‘check’
	domin-ir-a-ti	domin-ir-a	‘dominate’
i/i	vis-i-ti	vis-i	‘hang’
	od-laz-i-ti	od-laz-i	‘leave’
a/je	plak-a-ti	platʃ-e < /plak-je/	‘cry’
	trep-t-a-ti	trep-tɕ-e < /trep-t-je/	‘blink’
	formul-is-a-ti	formul-iʃ-e < /formul-is-je/	‘formulate’
iva/uje	is-pit-iva-ti	is-pit-uje	‘question’
	za-braɲ-iva-ti	za-braɲ-uje	‘forbid’
Ø/e	pas-Ø-ti	pas-e	‘graze’
	bos-Ø-ti < /bod-Ø-ti/	bod-e	‘stab’
	pi-Ø-ti	pij-e	‘drink’
	do-n(ij)e-Ø-ti	do-nes-e	‘bring’
	u-mr(ij)e-Ø-ti	u-mr-e	‘die’
nu/ne	gur-nu-ti	gur-ne	‘push’
ova/uje	k-oʊa-ti	k-uje	‘mint’
	ʊ(j)er-oʊa-ti	ʊ(j)er-uje	‘believe’
	rezult-oʊa-ti	rezult-uje	‘result’
(j)e/i	gor-(j)e-ti	gor-i	‘burn’
Ø/ne	sta-Ø-ti	sta-ne	‘stop’
	pas-Ø-ti < /pad-Ø-ti/	pad-ne	‘fall’
a/i	struj-a-ti	struj-i	‘flow’
	za-sp-a-ti	za-sp-i	‘fall asleep’
va/je	da-ʊa-ti	da-je	‘give’
	sa-zna-ʊa-ti	sa-zna-je	‘find out’
a/e	(x)rʊ-a-ti	(x)rʊ-e	‘wrestle’
(j)e/(ij)e	sm-(j)e-ti	sm-(ij)e	‘be allowed’
	pro-xt-(j)e-ti	pro-xt-(ij)e	‘desire’

in the final version of the classification. The nu/ne class was eventually reanalyzed as u/e, despite the fact that the exponent pair u/e always surfaces after [n],[4]

[4]A step further in the analysis of this TV class is made by Štarkl et al. (2024), who argue that *-n*

whereas the ∅/ne class survived in the final version of the classification. The remaining three TV classes, iva/uje, ova/uje and va/je were eventually included in the a/je class, reducing the number of TV classes to 10.

The rationale behind initially categorizing the TV classes according to surface-oriented criteria was to avoid considering the function of the relevant formatives. This approach stems from the insight that a clear-cut distinction between theme vowels and verbal (derivational) suffixes is not straightforward. Consider for example secondary imperfectivisation. In BCMS, one secondary imperfectivizer is the sequence *ava*, arguably a derivational suffix. On the other hand, secondary imperfectivisation can be (at least *prima facie*) achieved by changing only the TV, e.g., in *na-pas-a-ti*$^{\text{IPFV}}$ 'graze' from *na-pas-∅-ti*$^{\text{PFV}}$ 'graze'. The same is true of morphemes used for loanword integration. On the one hand, there are sequences with this function which are never considered to be single TVs, e.g., *ir-a*, as in *šut-ir-a-ti* 'kick' (based on the English *shoot*) and *džog-ir-a-ti* 'jog', but the same function can be performed by a single theme vowel, e.g., *šut-a-ti* 'kick' and *print-a-ti* 'print'.

While the function of formatives is not really useful in establishing theme vowel classes, there is an important surface generalization that holds of verbal derivational suffixes in BCMS and Slovenian: they all end in a very limited set of vowels, which all also appear as theme vowels in simplex verbs (Quaglia et al. 2022). This means that all verbal derivational affixes can be analyzed as containing a theme vowel. This assumption was taken into account from the get-go, so that derivational affixes that could be analyzed as containing an independently motivated theme vowel without any further assumptions were decomposed already in the first version. For instance, the TV class a/je includes verbs with the suffix *is-a*/*iš-e* and the a/a class includes, among others, verbs with the derivational affix *ir-a* and the secondary imperfective forms with *av-a*. However, the inclusion of the classes iva/uje, ova/uje and va/je in the a/je class required a more detailed consideration of phonological evidence, which was performed in Simonović et al. (2023).

A parallel process took place in the annotation in the Slovenian sub-database. Examples of morphologically complex sequences between the root and the inflection from Slovenian include (together with a final analysis):

is a diminutive suffix, whereas the TV proper is actually ∅/e, and hence the entire nu/ne class (and its Slovenian cognate ni/ne) should be subsumed under the ∅/e class. In this analysis, *u* in BCMS and *i* in Slovenian are floating vowels on the *-n* that help optimize the syllable structure. However, given that this analysis hinges upon theoretical assumptions regarding phonology, morphology, syntax and semantics of nu/ne verbs, we decided to keep u/e as a more descriptive label in order to stick with the decision that the annotation should be as pre-theoretical as possible.

- -irati/-iramo (e.g., in *organizirati, organiziramo* ‘to organize, we organize’); analyzed as: -ir-a-ti/-ir-a-mo; TV class: a/a
- -ja/-ja (e.g., in *pogovarjati, pogovarjamo* ‘to talk, we talk’); analyzed as: -j-a-ti/-j-a-mo; TV class: a/a
- -avati/-avamo (as in *prepoznavati, prepoznavamo* ‘to recognize, we recognize’); analyzed as: -av-a-ti, -av-a-mo, TV class: a/a
- -evati/-evamo (as in *dozdevati, dozdevamo* ‘to seem, we seem’); analyzed as: -ev-a-ti/-ev-a-mo; TV class: a/ a
- -ovati (or the phonologically predictable allomorph -evati)/-ujemo (e.g., in *oblikovati, oblikujemo* ‘to design, we design’, *razmnoževati*, *razmnožujemo* ‘to duplicate, we duplicate); analyzed as: -ov-a-ti/-u-je-mo; TV class: a/je
- -niti/-nemo (e.g., in *kihniti, kihnemo* ‘to sneeze, we sneeze’); analyzed as: -n-i-ti/-n-e-mo; TV class: i/e

2.2.1.2 Boundaries and overlaps between TV classes

As described above, separate theme-vowel classes were assumed in cases where the inclusion of verbs into existing classes would require phonological processes that cannot be independently justified. However, we at the same time tried to avoid assuming TV classes restricted to particular phonological environments.

Consider, for example, verbs such as *pi-∅-ti*/*pi-je-mo* ‘to drink / we drink’ and *ču-∅-ti*/*ču-je-mo* ‘to hear / we hear’. These forms could, in principle, be taken to justify the introduction of a separate TV class ∅/*je*, alongside the well-established class ∅/*e*. As is known from the established *a*/*je* class, the theme vowel *je* triggers palatalisation of the preceding consonant, as in *pis-a-ti*/*piš-e-mo* (from the underlying /*pis-je-mo*/) ‘to write / we write’. If verbs like *pi-∅-ti*/*pi-je-mo* were analysed as belonging to a distinct ∅/*je* class, we would therefore expect to encounter verbs such as *pas-∅-ti*/*paš-e-mo*, where the s–ʃ alternation would result from palatalisation. However, no such verbs are attested—an undesirable situation in a language in which the vast majority of roots are consonant-final. Moreover, analysing verbs like *pi-∅-ti*/*pi-je-mo* and *ču-∅-ti*/*ču-je-mo* as members of a separate ∅/*je* class would imply that the class ∅/*e* is restricted exclusively to verbs with consonant-final bases, since there are no verbs of the type *pi-∅-ti*/*pi-e-mo* or *ču-∅-ti*/*ču-e-mo*, i.e. without the additional glide. To avoid such an undesirable division, we include verbs like *pi-∅-ti*/*pi-je-mo* ‘to drink / we drink’ in the ∅/*e*

class and assume that the glide arises through a productive process of hiatus repair. For the full analysis, see Simonović et al. (2023); for a detailed discussion of hiatus resolution in Slovenian, see Simonović & Mišmaš (2023a).

An apparently comparable case arises in a small class of BCMS verbs that justify the TV class a/e (in addition to the well-established class a/je), in which the present-tense forms lack the expected palatalization/iotation of the root-final consonant. Our sub-database comprises 4 such roots (attested in 19 verbs): *greb-a-ti*/*greb-e-mo* 'to scratch/we scratch', *(h)rv-a-ti*/*(h)rv-e-mo* 'to wrestle/we wrestle', *jeb-a-ti*/*jeb-e-mo* 'to fuck/we fuck' and *zv-a-ti*/*zov-e-mo* 'to call/we call'. In deciding how to classify these four roots, the first question is whether there are uncontroversial a/je verbs with labial-final bases, which could constitute minimal pairs with the verbs just listed. Our sub-database contains two such verbs: *kap-a-ti*/*kap-lje-mo* 'to drip/we drip' and *hram-a-ti*/*hram-lje-mo* 'to limp/we limp'. The sub-database, however, contains no verbs in *-bati* or *-vati* which would belong to the a/je class, so we cannot find perfect minimal pairs for all the a/e verbs. But there exists a significant number of verbs that primarily fall into the a/a class that also have attested forms in the a/je class, e.g., *gib-a-ti*/*gib-a-mo* (but also *gibljemo*) 'to move/we move' and *pri-ziv-a-ti*/*pri-ziv-a-mo* (*prizivljemo* is also attested) 'to invoke/we invoke'. Given that the above-mentioned four roots never have palatalized PRS forms (i.e., the hypothetical forms **grebljemo*, **(h)rvljemo*, **jebljemo* and **zovljemo* are not attested), we take the existence of verbs like *gibati* as an argument for keeping the a/e class. Note, however, that this TV class is extremely small and restricted to roots ending in *v* and *b*, which is quite undesirable given our criteria. And finally, one further disadvantage of having both a/e and a/je lies in the fact that many verbs become in principle ambiguous between the two classes.

The issue of TV class ambiguity arises to different extents in all proposed classifications of West South Slavic verbs. In our initial system described in Table 2.3 for BCMS, there were many cases where the exponents of one TV class contained those of another class. For instance, it was not only the case that the exponents of the combination *iva/uje* contained those of the theme vowel combination *a/je*. Rather, *a/je* also contained the TV pair *a/e* and this in turn contained the TV pair *∅/e*.[5] Such a containment relation leads to theme-vowel indeterminacy (Simonović & Mišmaš 2023b): some verbs are analytically ambiguous and can be assigned to different TV classes. The TV classes *a/je* and *a/e* can again be used to show this issue. Consider, for example, a verb like *pis-a-ti*/*piš-e-mo* (from the underlying /*pis-je-mo*/) 'to write/we write'. This verb can be unequivocally

[5]We only discuss BCMS in this section, but see Simonović & Mišmaš (2023b) for Slovenian.

assigned to the class a/je as only this class motivates the s~ʃ alternation. There are, however, verbs that cannot unequivocally be assigned to one of the two classes because the consonant preceding the TV cannot be overtly palatalized. For instance, BCMS verbs like *postajati/postajemo* 'to become/we become' and *orati/oremo* 'to plough/we plough' can be assigned to either of the two classes. That is, the surface forms [postaj-e-mo] and [or-e-mo] would result from both underlying /postaj-e-mo/ and /postaj-je-mo/ and /or-e-mo/ and /or-je-mo/, respectively.

The situation becomes even more pressing in cases of verbs which display root allomorphy. For instance, the verb *slati/šaljemo* 'to send/we send', can be analyzed either as belonging to the TV class *∅/e* (implying the segmentation *sla-∅-ti/šalj-e-mo*), belonging to the TV class *a/e* (implying the segmentation *sl-a-ti/šalj-e-mo*) or belonging to the TV class *a/je* (implying the segmentation *sl-a-ti/šalj-e-mo*, the latter deriving from the underlying /*šal-je-mo*/).

While some ambiguous cases can only be resolved by arbitrary choices, in many cases prosodic generalisations which hold of unambiguous cases help us decide between the different analyses. As it turns out, all three TV classes that can be argued to host the verb *slati/šaljemo* (a/je, a/e, and ∅/e) display such generalizations. (In order to show this, we will use IPA in the remainder of this paragraph.) For instance, all verbs in the ∅/e class which display unpredictable root allomorphy have a long root-final vowel in the infinitive, e.g., uzeː-∅-ti/uzm-eː-mo 'to take/we take' or don(ij)eː-∅/-ti/dones-eː-mo 'to bring~we bring'. On the other hand, in verbs that have the theme vowel *je* in the present tense, the high tone falls on the syllable preceding it, e.g., ˈtes-á-ti/ˈtéʃ-eː-mo 'to carve/we carve'. Finally, all unambiguous members of the TV class a/e, on the other hand, have the high tone on the theme vowel, e.g., ˈgreb-á-ti/ˈgreb-éː-mo 'to scratch/we scratch' (potentially also justifying terming this class á/é). Taking these prosodic generalizations into consideration when analyzing the verb slati 'to send', its prosodic pattern ˈsláti/ˈʃáɾeːmo only fits with the unambiguous cases of the a/je class, which is why we can assign this verb to the a/je class.

2.2.2 Root allomorphy

For the purposes of *WeSoSlaV*, root allomorphy is defined as any unpredictable difference in the exponence of the root morpheme (i.e. the material preceding the TV) in the non-finite and finite form of the verb. In annotating root allomorphy, we disregarded defective verbs, defined as verbs for which we cannot determine

the TV class.[6] The verbs annotated as defective have this information ('defective') repeated in the two columns pertaining to root allomorphy.

Our definition of root allomorphy excludes instances of predictable phonologically conditioned allomorphy (such as *pisati, pišemo* 'to write, we write' in both BCMS and Slovenian). Further examples of phonologically conditioned allomorphy include:

- *peći, pečemo* in BCMC and *peči, pečemo* in Slovenian, 'to bake, we bake', from /pek-∅-ti/, /pek-e-mo/ (where velar palatalisation is productive in this position in both languages and /kt/⟶[tɕ] and /kt/⟶[tʃ] are regular in this position in BCMS and Slovenian, respectively)
- *piti, pijemo* 'to drink, we drink' from /pi-∅-ti/, /pi-e-mo/ (where *j*-epenthesis is productive in this position)
- *obuti, obujemo* 'to put on shoes, we put on shoes' from /obu-∅-ti/, /obu-e-mo/ (where *j*-epenthesis is productive in this position)
- *gristi, grizemo* 'to bite, we bite' from /griz-∅-ti/, /griz-e-mo/ (where regressive voicing assimilation is productive)
- *krasti, krademo* 'to steal, we steal' from /krad-∅-ti/, /krad-e-mo/ (where /tt/⟶[st] is exceptionless in this environment)
- *kupovati, kupujemo* 'to buy, we buy' (since *ova-uje* is not counted as a part of the root)

We illustrate the cases of unpredictable root allomorphy in Table 2.4 for BCMS and Table 2.5 for Slovenian. In *WeSoSlaV*, we annotate if the verb has root allomorphy with a 1 in the ROOT ALLOMORPHY column (0 if the verb has no root allomorphy). Additionally, in the column ROOT ALLOMORPHS (LIST), the root allomorphs are listed for verbs that exhibit root allomorphy.

2.2.3 Prosodic properties

Both Slovenian and BCMS display considerable variation in the implementation of the standard prosody, varying from pitch-accent systems with distinctive

[6] Verbs are marked as defective in the THEME VOWEL column. For BCMS, the verbs *sam*, *jesam*, *biti*, all 'to be', *nisam* 'not to be', hteti/htjeti 'to want' are marked as defective. For Slovenian, the verbs *biti* 'to be', *imeti* 'to have' and *iti* 'to go' are marked as defective.

Table 2.4: Examples of root allomorphy in BCMS

Infinitive	PRS.3SG	GLOSS	Root allomorphy	Root allomorphs (list)
početi	počne	'begin'	1	če, čn
žvakati	žvaće	'chew'	1	žvak, žvać
razneti	raznese	'blow up'	1	ne, nes
doći	dođe	'come'	1	ć, đ

Table 2.5: Examples of root allomorphy in Slovenian

Infinitive	PRS.3SG	GLOSS	Root allomorphy	Root allomorphs (list)
početi	počne	'do'	1	če, čn
klati	kolje	'slaughter'	1	kl, kolj
obiti	obide	'go-around'	1	i, id
gnati	žene	'goad'	1	gn, žen

vowel length to stress systems with no vowel-length distinction. However, these systems have different formal status and are described to different extents.

Standard Slovenian grammars, e.g., Toporišič (2004), recognize two variants which share stress position and vowel length distinction: one with distinctive pitch accent and one without. Vowel length is largely predictable in the verbal system and, as convincingly argued in Jurgec (2022), phonetic studies have failed to confirm its reality in actual speech. For this reason, in the columns pertaining to prosody, only stress is marked. The annotation reflects the intuition of the annotators (native speakers of Slovenian) and the prominence marked in SSKJ (2014).

BCMS standard grammars all describe a system with pitch accent and distinctive vowel length, even though in big parts of Serbia and Croatia a variant is spoken with no distinctive tone or vowel length (Simonović & Kager 2023). Simonović & Kager (2023) show that in the pure stress standard BCMS, stress is virtually fully predictable in the verbal domain (except in denominal verbs). For this reason, in the columns pertaining to prosody for BCMS the pitch-accent system is described. In this pitch-accent system the position of stress is predictable from the position of High tone (see Simonović 2022b and references therein), so that in the BCMS sub-database only the position of the High tone and the vowel length are marked. The annotation reflects the intuition of the annotators (na-

tive speakers of BCMS), taking into account the prosodic specifications marked in the standard dictionaries (most prominently Nikolić 2000 and HJP 2006).

2.2.3.1 Prosodic prominence

Prosodic prominence (associated with High tone in BCMS and stress in Slovenian) is marked across eight columns in both sub-databases. Four columns mark the prosodic prominence of the INF, and four of the PRS.1PL form.

The annotation was performed in the following way. The relevant form (INF or PRS.1PL) was segmented into four parts. Starting from the end of the word form, these parts are: 1) inflectional ending, 2) TV, 3) the syllable preceding the TV and 4) all syllables before the syllable preceding the TV. The prominence is marked as 1 in the column carrying the number of the relevant part of the word. In other words, prominence on position 1 means the verb form has prosodic prominence on the inflectional ending. Prominence on position 2 means prominence on the theme vowel. Prominence on position 3 means that stress (Slovenian) or High tone (BCMS) falls on the syllable preceding the TV. Finally, prominence on position 4 means that the relevant form has prosodic prominence on a syllable preceding the pre-TV syllable.

The segmentation and the marking of prosody is shown in Tables 2.6-2.7 for verb *pogledati*, *pogledamo* [poˈgledati, poˈgledamo] 'to look, we look' in Slovenian. Note that the actual database only includes information in the topmost and the bottommost row.

Table 2.6: Prosodic prominence: *pogledati* 'to look'

all preceding syllables	pre-TV syllable	TV	INF
4	3	2	1
po	gled	a	ti
0	1	0	0

While in Slovenian determining the stressed syllable is straightforward, in BCMS determining the syllable carrying the H is rendered more difficult by the traditional terminology which employs *rising* and *falling accents*, lumping together stress and High tone. The instructions for identifying the syllable carrying the H based on the traditional BCMS representations are summarized below.

1. If there is a falling accent (*kȕća* 'house', *prâvda* 'justice'), the stressed syllable is prominent, i.e carries a H ([kútɕa], [práːʋda]). Examples from the

Table 2.7: Prosodic prominence: *pogledamo* ‘we look’

all preceding syllables	pre-TV syllable	TV	INF
4	3	2	1
po	gled	a	mo
0	1	0	0

verbal domain: *pȁdati* is [pádati] ‘to fall’, *pȁdāmo* is [páda:mo] ‘we fall’, *mȍlīmo* is [móli:mo] ‘we ask’, *kûmīmo* is [kú:mi:mo] ‘we beg’, *kârtati* is [ká:rtati] ‘to play cards’, *kârtāmo* is [ká:rta:mo] ‘we play cards’.

2. If there is a rising accent (*rúka* ‘arm’, *nòga* ‘leg’), the syllable following the syllable with accent mark carries a H ([ru:ká], [nogá]). Examples from the verbal domain: *gráditi* is [gra:díti] ‘to build’, *lòmiti* is [lomíti] ‘to break’, *lòmīmo* is [lomí:mo] ‘we break’, *nàpadati* is [napádati] ‘to attack’, *nàpadāmo* is [napáda:mo] ‘we attack’, *mòliti* is [molí:ti] ‘to ask’, *zàmolimo* is [zamóli:mo] ‘we ask’.

In 2.8 we provide illustrative examples of prominence placement from both languages. We use regular spelling for the segmental content, but mark prominence using IPA (ˈ*ta* for stress, *tá* for High tone).

Table 2.8: Prosodic prominence: examples

			INF				-	PRS			
Verb	Gloss	Lang.	1	2	3	4	-	1	2	3	4
iˈgrati, iˈgramo	‘play’	Slovenian	0	1	0	0	-	0	1	0	0
voˈziti, ˈvozimo	‘drive’	Slovenian	0	1	0	0	-	0	0	1	0
ˈgledati, ˈgledamo	‘watch’	Slovenian	0	0	1	0	-	0	0	1	0
ˈmalicati, ˈmalicamo	‘snack’	Slovenian	0	0	0	1	-	0	0	0	1
bojáti, bojímo	‘fear’	BCMS	0	1	0	0	-	0	1	0	0
molíti, mólimo	‘ask’	BCMS	0	1	0	0	-	0	0	1	0
uláziti, ulázimo	‘enter’	BCMS	0	0	1	0	-	0	0	1	0
organízovati, organízujemo	‘organize’	BCMS	0	0	0	1	-	0	0	0	1

2.2.3.2 Vowel length

Since vowel length was not annotated for Slovenian, this section is only relevant for BCMS. Just like with prominence, vowel length was annotated across eight columns – four for the INF, and four for the PRS. To annotate vowel length, each relevant form was segmented into the same four parts as for the annotation of prominence. Consequently, all parts which contain a long vowel were marked with a 1. This is shown in the schema below for *poigravati, poigravamo* [poigraːvati, poigraːvaːmo] 'to play, we play' (Tables 2.9-2.10) and *zavisiti, zavisimo* [zaːvisiti, zaːvisiːmo] 'to depend, we depend' (Tables 2.11-2.12). Note that only the top and bottom line can be found in *WeSoSlaV*.

Table 2.9: Vowel length: *poigravati* 'to play'

all preceding syllables	pre-TV syllable	TV	INF
4	3	2	1
poi	graːv	a	ti
0	1	0	0

Table 2.10: Vowel length: *poigravamo* 'we play'

all preceding syllables	pre-TV syllable	TV	INF
4	3	2	1
poi	graːv	aː	mo
0	1	1	0

Table 2.11: Vowel length: *zavisiti* 'to depend'

all preceding syllables	pre-TV syllable	TV	INF
4	3	2	1
zaː	vis	i	ti
1	0	0	0

As with High tone, the traditional prosodic representations were transformed into those relevant for our purposes. In this case, the instructions are much more

Table 2.12: Vowel length: *zavisimo* ‘we depend’

all preceding syllables	pre-TV syllable	TV	INF
4	3	2	1
za:	vis	i:	mo
1	0	1	0

straightforward, as any vowel carrying one of the three diacritics in Table 2.13 is long.

Table 2.13: Diacritics

Diacritic	Example
ˆ	*mâjka* is *ma:jka* ‘mother’
´	*rúka* is *ru:ka* ‘hand’
ˋ	*nàpadāmo* is *napad:amo* ‘we attack’

The instructions for marking vowel length with specific examples from the verbal domain are summarized in (2). These instructions also mention the positions which always have a long vowel and the ones that never have one.

(2) Infinitive length: instructions
   a. Vowel length marked on 1: Enter 1 if the infinitive ending has a long vowel, which was never the case.
   b. Vowel length marked on 2: Enter 1 if the infinitive theme vowel is long, which is never the case.
   c. Vowel length marked on 3: Enter 1 if the final syllable of the base is long, e.g., *poigrávati* – *poigra:v-a-ti* ‘to play’.
   d. Vowel length marked on 4: Enter 1 if a non-final syllable of the base is long, e.g., *závisiti* – *za:vis-i-ti* ‘to depend’.

The instructions for marking the vowel length in the PRS are summarized in (3).

(3) PRS vowel length: instructions
   a. Vowel length marked on 1: Enter 1 if the PRS.1PL ending has a long vowel, which is never the case.

b. Vowel length marked on 2: Enter 1 if PRS theme vowel is long, e.g., *ǐgrāmo – igr-a:-mo* 'we play', *závisīmo – za:vis-i:-mo* 'we depend on', *poìgrāvāmo – poigra:v-APRS-mo* 'we play'. This is always the case.
c. Vowel length marked on 3: Enter 1 if the final syllable of the base is long, *poìgrāvāmo – poigra:v-a:-mo* 'we play'.
d. Vowel length marked on 4: Enter 1 if a non-final syllable of the base is long, e.g., *závisīmo – za:vis-i:-mo* 'we depend'.

2.2.3.3 Prosodic variants

For verbs that have only a single prosodic shape prosody is marked using 1's. In cases where multiple prominence patterns are attested, multiple numbers are used. There is a difference between the Slovenian and the BCMS sub-databases when it comes to the use of numbers higher than 1.

In Slovenian, the typical case of variation in prosodic prominence is the verb *motiti, motimo* 'to disturb, we disturb'. This verb allows two stress patterns in the INF: [ˈmotiti] and [moˈtiti], but only one in the in the PRS.1PL: [ˈmotimo]. This means that a single form, INF, can be considered the locus of variation. For this reason, only the INF form has more than one prominence marked. For this form, 1 is entered in the column for the stress position perceived as less marked and a 2 in the column for the stress position perceived as more marked. The same principle was applied in the very rare cases where both INF and PRS have multiple prosodic shapes, e.g., *oglasiti, oglasimo* 'to answer (the phone), we answer (the phone)' [oˈglasiti/oglaˈsiti, oˈglasimo/oglaˈsimo]. In such cases as well, for each form the less marked prosodic shape is marked with a 1, whereas the more marked shape was marked with a 2. This is illustrated in 2.14 for both these verbs.

Table 2.14: Prosodic prominence for the verb *motiti* 'to disturb' and *oglasiti* 'to answer (the phone)' in Slovenian

Verb	1	2	3	4	-	1	2	3	4
ˈmotiti/ moˈtiti, ˈmotimo	0	2	1	0	-	0	0	1	0
oˈglasiti/oglaˈsiti, oˈglasimo/oglaˈsimo	0	1	2	0	-	0	1	2	0

In BCMS, the typical cases of variation in prosodic prominence involve cases of two parallel paradigms, which often differ in vowel length and have no overlapping members. An example is the verb *kužiti, kužimo* 'to contaminate, we contaminate'. Its two possible prosodic paradigms are [kúʒiti, kúʒiːmo] and [kuːʒ́iti,

kúːʒiːmo][7]. In order to capture the existence of two possible paradigms, in the BCMS sub-database, we used numbers in cells to describe entire paradigms over the 16 columns that describe prosody. This means that all 1's describe the least marked prosodic pattern (in this case [kúʒiti, kúʒiːmo]), all 2's the next least marked one [kuːʒíti, kúːʒiːmo] and so forth. As a consequence, some cells have values such as 12 or 123, which only means that multiple parallel paradigms have High tone or vowel length in the same position. The example annotations for the discussed verb are in 2.15 for High tone and 2.16 for vowel length.

Table 2.15: Prosodic prominence for the verb *kužiti* [kúʒiti, kúʒiːmo] and [kuːʒíti, kúːʒiːmo] 'to contaminate'

Verb	1	2	3	4	-	1	2	3	4
kužiti, kužimo	0	2	1	0	-	0	0	12	0

Table 2.16: Vowel length for the verb *kužiti* [kúʒiti, kúʒiːmo] and [kuːʒíti, kúːʒiːmo] 'to contaminate'

Verb	1	2	3	4	-	1	2	3	4
kužiti, kužimo	0	0	2	0	-	0	12	2	0

2.3 Verbal derivation

This section provides a description of the second part of *WeSoSlaV* – the one devoted to the verbal derivation (Milosavljević et al. 2023). The section introduces the annotation of the following (groups of) columns: roots (§2.3.1), prefixes (§2.3.2), compounds (§2.3.3), suffixes (§2.3.4), and the morphological complexity of the verb (§2.3.5).

2.3.1 Roots

Roots are listed in the column Root, which can be used as a rough indicator of which verbs belong to the same word family. While we do not claim that the annotated form is the exact lexical content of the root of the verb, we did

[7]The traditional representations would be *kȕžiti, kȕžīmo* and *kúžiti, kûžīmo*.

attempt to capture what we believe is the common core of all the related verbs. The annotated form should rather be taken as a cell that can help the users group together verbs that are related by some common (phonological and semantic) core. In this sense, the chosen form could in principle be replaced by a random number/index (see §2.3.1.1 for a discussion on how our annotation relates to current theorizing on roots).

The material in the cell Root refers to the verb without prefixes, verbal suffixes, TVs and inflectional material. A simple example is provided in Table 2.17. Here, the first verb is simplex, i.e. it doesn't contain a prefix or a suffix, so the root corresponds to the part of the verb without a TV and the inflectional ending. The second verb is prefixed, and we get the root by removing the prefix *na-* in addition to the TV and the inflectional ending. The remaining two verbs in the table contain a secondary imperfectivizing suffix in addition to other pieces of morphology. Again, the root is what remains once we take away all these morphological layers.

Table 2.17: Roots, simple examples from BCMS and Slovenian

Language	Verb	Root	Translation
BCMS/Slovenian	pis-a-ti$^{\text{IPFV}}$	pis	'to write'
BCMS/Slovenian	na-pis-a-ti$^{\text{PFV}}$	pis	'to write'
BCMS	pre-pis-iva-ti$^{\text{IPFV}}$	pis	'to copy'
Slovenian	pre-pis-ova-ti$^{\text{IPFV}}$	pis	'to copy'

As evident from Table 2.17, the meaning of verbs was taken as a motivating factor to determine that verbs with homophonous material belong in the same word family.

Another important ingredient for detecting the same root shared by multiple verbs comes from alternation patterns. Specifically, if multiple verbs have the same alternation pattern in their paradigm and/or within the aspectual pair, we treat them as having the same root. As discussed in more detail in §2.3.1.1, the alternation found among two forms treated as the same root must be regular, i.e. attested in multiple roots. This is illustrated in Table 2.18:[8] the first four verbs (from Slovenian) share the common core *l-g*, whereas the remaining four verbs (from BCMS) are annotated as *b-r*. Crucially, the same alternations between perfective variants and their imperfective counterparts (descriptively, the change

[8]We will explain the role of the hyphen in our annotation later in this section.

from *o* to *a* in the Slovenian examples, i.e. from ∅ to *i* in the BCMS examples) are also attested in many other verbs, i.e. their roots (for BCMS, see Simonović et al. 2023).

Table 2.18: Roots with alternations

Language	Verb (infinitive)	Root	Translation
Slovenian	razložiti^{PFV}	l-g	'to explain', 'to put down'
Slovenian	razlagati^{IPFV}	l-g	'to explain', 'to put down'
Slovenian	položiti^{PFV}	l-g	'to lay down'
Slovenian	polagati^{IPFV}	l-g	'to lay down'
BCMS	odabrati^{PFV}	b-r	'to choose'
BCMS	odabirati^{IPFV}	b-r	'to choose'
BCMS	sabrati^{PFV}	b-r	'to sum'
BCMS	sabirati^{IPFV}	b-r	'to sum'

If there are two partially homophonous roots and only one of them displays an alternation, or they participate in different alternation patterns, they are considered as two different roots, which is reflected by a different consonant in our annotation, as in Table 2.19. For instance, while the verbs *primaknuti* and *primicati*, on the one hand, and *umočiti* and *umakati* on the other, possibly share some of their meaning and thus perhaps their roots overlap, we annotate them as separate roots (*m-k* vs. *m-č*) because they display different alternations: the vowel change *a* to *i* in the first pair, and *o* to *a* in the second pair. Again, these alternations are also found in several other roots.

If the verb displays root allomorphy, we choose one of the forms of the root. While we gravitated towards what plausibly is the underlying form, this is not necessarily always so.[9] Crucially, the same form of the root is then used with all the verbs within the same word family, as illustrated in (2.20).

In addition to these general instructions, there are further specific annotation peculiarities that need to be explained away.

Numbers were used to differentiate homophonous roots in verbs that, according to the annotators' intuitions, could not be connected through common meanings. This is illustrated in Table 2.21.

[9]This is because the notion of an "underlying form" is itself a theoretical construct, and reasonable disagreement exists as to how such forms should be defined.

Table 2.19: Roots: Alternation as a factor in disambiguation

Language	Verb (infinitive)	Root	Translation
Slovenian	dotakniti$^{\text{PFV}}$	t-k	'to touch'
Slovenian	dotikati$^{\text{IPFV}}$	t-k	'to touch'
Slovenian	točiti$^{\text{IPFV}}$	t-č	'to pour'
Slovenian	natakati$^{\text{IPFV}}$	t-č	'to pour'
BCMS	primaknuti$^{\text{PFV}}$	m-k	'to get closer'
BCMS	primicati$^{\text{IPFV}}$	m-k	'to get closer'
BCMS	umočiti$^{\text{PFV}}$	m-č	'to dip'
BCMS	umakati$^{\text{IPFV}}$	m-č	'to dip'

Table 2.20: Roots: Roots with predictable allomorphy

Language	Verb (infinitive)	Root	Translation
Slovenian	počiti$^{\text{IPFV}}$	pok	'to burst'
Slovenian	pokati$^{\text{IPFV}}$	pok	'to burst'
Slovenian	napočiti$^{\text{PFV}}$	pok	'to arrive'
BCMS	izbeći$^{\text{PFV}}$	beg	'to avoid'
BCMS	izbegnuti$^{\text{PFV}}$	beg	'to avoid'
BCMS	izbegavati$^{\text{IPFV}}$	beg	'to avoid'
BCMS	bežati$^{\text{IPFV}}$	beg	'to run away'

Hyphens were used in the database for roots that display apophony/ablaut (i.e., that do not have stable vowels), typically between the imperfective and perfective forms of the verb. This is illustrated in Table 2.18 above.

A **plus** sign is used if the 'root' appears with an affix that also occurs in other categories (and is not thus a verbal affix). For instance, for the Slovenian verbs *tolmačiti* 'to interpret' and *krvaveti* 'to bleed', roots are annotated as *tolm+ač* (as *-ač* is a common nominalizer) and *krv+av* (*-av* is a common adjectivizer), we return to this in 6. Similarly, for the BCMS verbs *krvariti* 'to bleed' and *besniti* 'to rage', the corresponding roots are *krv+ar* and *bes+n*. Here, the affix *-ar* typically derives nouns, whereas *-n* is often found in adjectives. The plus symbol is also used in compounds (see §2.3.3 for the annotation of compounds). For instance, for the verbs *telovaditi* 'to exercise' (Slovenian) and *obelodaniti* 'to disclose' (BCMS), roots are annotated as *tel+o+vad* and *bel+o+dan*, respectively.

Table 2.21: Roots: the role of numbers in annotation

Language	Verb (infinitive)	Root	Translation
Slovenian	prati^{IPFV}	p-r	'to wash'
Slovenian	sprati^{PFV}	p-r	'to wash away'
Slovenian	spirati^{IPFV}	p-r	'to wash away'
Slovenian	odpreti^{PFV}	p-r1	'to open'
Slovenian	odpirati^{IPFV}	p-r1	'to open'
BCMS	biti^{IPFV}	bi	'to be'
BCMS	zbivati^{IPFV}	bi	'to happen'
BCMS	ubiti^{PFV}	bi2	'to kill'
BCMS	ubijati^{IPFV}	bi2	'to kill'

Table 2.22: Roots: Plus in annotation

Language	Verb (infinitive)	Root	Translation
Slovenian	tolmačiti^{IPFV}	tolm+ač	'to interpret'
Slovenian	krvaveti^{IPFV}	krv+av	'to bleed'
Slovenian	telovaditi^{IPFV}	tel+o+vad	'to exercise'
BCMS	krvariti^{IPFV}	krv+ar	'to bleed'
BCMS	besniti^{PFV}	bes+n	'to rage'
BCMS	obelodaniti^{IPFV}	bel+o+dan	'to disclose'

The **colon** was used in the BCMS sub-database to mark cases where the annotator had the intuition that the verb is prefixed, but the root reconstructed on this analysis does not appear in other words. e.g. *u:ze* 'take', where *ze* does not appear in any other verb. The colon was used in both the root column (as *u:ze*) and the columns listing prefixes (as *u:*). In the Slovenian sub-database, the colon was used when the base potentially includes a prefix, but, according to the annotator's intuition, the verb itself is not derived by prefixation (these instances are also marked as potentially prefixed; see §2.3.2). The relevant example is *odmevati* 'to echo', presumably derived from the noun *odmev* 'echo', as **mevati* does not appear as a separate verb, whereas *od-* is a well-attested preposition/prefix. This verb is marked as having the root *od:mev*.

2.3.1.1 Issues with the annotation: How to identify a root?

In this section, we highlight some problems, doubts and open questions emerging from our annotation of roots. In general, the annotation procedure in *WeSoSlaV* is intended to be as theory-neutral as possible. However, no attempt to maintain a purely descriptive standpoint is theoretically innocent. This is also true for roots. The nature of roots as units in morphological modeling has recently gained abundant attention in the literature, yet their status is still a matter of ongoing debate (see, e.g., Alexiadou & Lohndal 2017, Panagiotidis 2020, Embick 2021 for recent overviews). Traditionally, roots are assumed to be Saussurean signs, i.e. their nature is determined by pairing of form and meaning (e.g., in the lexicalist theories such as Rappaport Hovav & Levin 2010, Rappaport Hovav 2014, 2017, 2021). In some works, it is argued that roots are individuated phonologically, but are devoid of syntax and semantics (e.g. within the Exo-Skeletal approach of Borer 2005, 2009, 2014). There are also approaches, especially those in Distributed Morphology, arguing that roots are individuated in the syntax, as independent indices on the root node that again serves to pair sound and meaning (cf. Acquaviva 2008, Harley 2014, Panagiotidis 2020). While we remain agnostic about the nature of roots, below we comment on how our ultimate annotation relates to the existing approaches in identifying (common/same) roots. To do so, we first elaborate on our annotation procedure illustrated above.

As stated in the previous section, the most general guideline in identifying common roots was to unify verbs whose roots share both form and meaning. This procedure may seem intuitive and can be straightforwardly applied in typical examples, such as BCMS or Slovenian *pisati* 'to write' and *zapisati* 'to write down', which clearly share the root *pis*, with fully transparent form and meaning. However, many examples pose real challenges for the annotation. In 2.3.1, we described some of the dilemmas and how we handled them, e.g. the use of a hyphen to indicate that roots display apophony. This is already one example that shows that our annotation could not rely only on the surface forms of roots, and is thus not purely descriptive: it hinges on what we analyze as cases of regular apophony. We take regular apophony to be the same type of phonological change that is detected for several roots (usually at least five). For example, the change *o>a* is observed regularly in BCMS aspectual pairs such as *sk**o**č-i-ti/skak-**a**-ti* 'to jump', *po-m**o**g-ti (>pomoći)/pom**a**gati* 'to help', *s-tv**o**r-i-ti/s-tv**a**r-a-ti* 'create', etc. Here, the first verb within the pair is always perfective and the second one systematically imperfective. The *o>a* change can also be observed in Slovenian perfective-imperfective pairs, e.g., *pre-v**o**z-i-ti/pre-v**a**ž-a-ti* 'to drive', *ob-n**o**v-i-ti/ob-n-**a**vlj-a-ti* 'to renovate'. This means that verbs are annotated as

having the same root if these roots share either the same surface form or the same common formal core that undergoes regular changes *and* the same core meaning. If some of these criteria is not fulfilled, roots are marked differently.

However, there are also many borderline cases in which it is not clear whether roots share the same form and/or meaning. Some verbs share a similar form and have very similar meanings, but they still could not be annotated as having the same root according to the criteria specified above. For instance, the BCMS verbs *žmiriti* and *žmuriti* can be both roughly translated as 'to squint' or 'to narrow one's eyes', yet they differ in the roots' vowel (*i* and *u*, respectively) that do not display a predictable regular change. Consequently, the two verbs are annotated with different roots: *žmir* and *žmur*, respectively.

Further, some verbs in the BCMS sub-database include the root *sp*, e.g., *sp-a-ti* 'sleep' and *za-sp-a-ti* 'fall asleep', which is also part of the verbs *spav-a-ti* 'sleep', *u-spav-a-ti* 'fall asleep' and *pre-spav-a-ti* 'oversleep'. But since the latter group of verbs also include *av* as part of the base, the two groups of verbs are annotated as *sp* and *spav* for their respective roots – despite the shared part of the form and the same underlying meaning. Alternatively, we could annotate *spavati* as *sp+av*, treating *av* as some kind of (multifunctional) affix (see Chapter 6 for an in-depth discussion). However, such multifunctional affixes usually appear (semi-)regularly across various roots with a similar function, which is not the case with *av* in this case. Specifically, while BCMS employs a secondary imperfective suffix *ava*, the case under discussion is the only example where the hypothesized affix *av* does not have an imperfetivizing role in the verbal domain, as both *spati* and *spavati* are imperfective. We therefore treat *sp* and *spav* as separate roots in BCMS. Note, though, that even if *av* were treated as a multifunctional element, the two roots would still be annotated differently: as *sp* and *sp+av*, respectively.[10] The annotation in the Slovenian sub-database is slightly different with respect to multifunctional affixes. Here, in similar pairs, such as *gost-i-ti* 'to host', *gost-ov-a-ti* 'to guest', the suffix *-ov* was not explicitly annotated in the Root column, but it was annotated as a verbal suffix. The suffix *-ov* was annotated in the Root column if the verbal base independently included it, e.g., *raz-polov-i-ti* 'to halve' related to *polovica* 'half' and *pol* 'half'.

Perhaps even more radically, the BCMS verbs *sjati*, *sijati* and *sjajiti* are annotated as having three different roots (*sj*, *sij* and *sjaj*, respectively), although they share the common sequence *sj* and all have very similar meaning roughly translatable as 'to shine'. A similar example in Slovenian is the triplet *zvoniti* 'to ring'

[10] Note that the in Slovenian, *-av* is treated as a multifunctional affix; see also Simonović (2020), Simonović & Mišmaš (2023a).

and *zveneti* 'to sound' and the secondary imperfective *odzvanjati* 'to resonate', which were annotated as having different roots. That is, despite these similarities in both meaning and form, there is no regular change of the phonological form that would enable us to treat them as having the same root.

Additionally, some verbs share the same form and the same core meaning in the infinitive, but differ in the present tense. BCMS verbs *ići* 'go' and *mimoići* 'go past' are a case in point. Both verbs have the final sequence *-ići* and behave exactly the same in all the forms traditionally derived from the infinitival base. However, the PRS.3SG form of the first verb is *i̲d̲e*, and of the second one *mimoi̲d̲e*. Since this difference cannot be ascribed to any productive process in the contemporary BCMS phonology or morphology, the two verbs are annotated as having different roots. Alternatively, we could hypothesize that verbs like *mimoići* have the TV *je* in the present tense, so that the underlying representation of the PRS.3SG form would be /mimoidje/, which would give [mimoidʑe] by productive phonological rules. This, however, is not plausible, as the TV combination ∅/je is not independently attested.

The same procedure was applied to verbs with the same meaning but slightly different forms in regional language variants. For instance, the verb 'to prolong' takes the form *dužiti* in the Eastern BCMS (roughly, Serbian), and the form *duljiti* in the Western BCMS (roughly, Croatian). Based on our guidelines, the respective roots/bases are 'dug' and 'dul+j', respectively.

The emphasis on phonological identity for the BCSM sub-database can also be observed in suppletive forms. For instance, the verbs *otići* and *odlaziti* form an aspectual pair where both verbs mean 'leave' (and have the same syntactic properties, such as argument structure) and only differ in that the first verb is perfective and the second one is imperfective. Yet, since they have different forms (*id* v. *laz*), they are annotated as different roots. Somewhat differently, in the Slovenian sub-database the suppletive verb *iti, gremo* 'to go, we go' (otherwise annotated as defective), was annotated as having the root *i*. Since the same root can also be observed in the infinitive of verbs like *oditi, odidemo* 'to leave, we leave', both verbs were annotated as having the root *i*.

The discussion so far points out one limitation of our annotation procedure: *WeSoSlaV* does not contain information on formal and semantic relations between roots discussed above, which certainly have much in common formally and/or semantically. Consequently, they are treated as different roots to the same extent as roots with completely different forms and totally unrelated meanings.

On the opposite end of the scale, it was difficult to determine whether roots share a common semantic component so that they should be grouped together. For instance, is it the case that BCMS verbs *na-prav-i-ti* 'to make' and *iz-prav-i-ti*

'to correct' contain the same root *prav* which preserves the same meaning? Same for the Slovenian *po-prav-i-ti* 'to fix' and *na-prav-i-ti* 'to make'. There is no established formal procedure to answer this and similar questions, so the ultimate decision had to depend on the annotators' intuitions – with the agreement of at least two annotators. In this particular case, both verbs are marked for the same root (*prav*).

One further interesting puzzle for the annotation of roots comes from examples that apparently consist just of a prefix and a suffix. For instance, if the BCMS verb *prenuti* 'to startle' consists of parts that can be plausibly analyzed as a prefix *pre-* 'across' and the semelfactive suffix *-nu*. The question, then, arises, what counts as a root. One possible solution is that the root is also *n* and that one *n* gets deleted, so we annotated this root as *n*. Note, however, that this is an arbitrary choice for the purpose of annotation rather than the result of a deep analysis of such patterns, which must be left for future research. Additionally, since the same verb does not appear with other prefixes, we eventually annotated the root as *pre:n*. On the other hand, a slightly different arbitrary decision was made for the Slovenian verb *vz-e-ti, vzam-e-mo* 'to take, we take', which was annotated as a part of the e/e TV class and *vz-* as a prefix (note that the verb is perfective), which could again mean that the verb does not have a phonologically overt root. Here the root was simply annotated as *vz*, but the verb was also annotated as having root allomorphy (the allomorphs being *vz* and *vzam*). Again, we leave a precise analysis of this verb (and its prefixed pairs) for future research.[11]

Coming back to different theoretical approaches to roots introduced above, our descriptive approach is most closely related to the traditional view of pairing form and meaning, as we tended to group verbs that share the same form and the same core meaning. However, the examples discussed in this section indicate that phonological identity has precedence over the shared meaning when it

[11] We also leave to future work if our decision with respect to a group of forms related to the verb *vzeti* 'to take' was the most sound one. That is, *vzeti* can form two different type of imperfective counterparts, *vzemati*^{IPFV}, *vzemamo*^{IPFV} 'to take, we take' and the *jemati*^{IPFV}, *jemljemo*^{IPFV} 'to take, we take'. The first were annotated as having the root *vzem*, the second *jem*, hence no relation to the perfective verb is made in the Slovenian sub-database. But in addition, there is another class of verbs that seem to have the root allomorph *jem* in the finite form, i.e., *za-j-e-ti*^{PFV}, *za-jam-e-mo*^{PFV} 'to scoop up, we scoop up'. These were annotated as having the root *j1* and as having root allomorphy (*j, jam*). And as the number next to the root implies, there are other types of *j*-verbs, exemplified with the pairs *pre-j-e-ti*^{PFV}, *pre-jm-e-mo*^{PFV} 'to receive., we receive.' and *pri-j-e-ti*^{PFV}, *pri-(i)m-e-mo*^{PFV} 'to hold, we hold', annotated as *j2* (with allomorphs *j, jm*) and *j3* (with allomorphs *j, im*), respectively. Again, the annotation implies that there is no relation between these verbs, while their meaning (all are related to the act to holding) and, in fact, etymology, see Snoj (2009), implies differently.

comes to the individuation of roots: verbs that share meaning but differ formally (e.g. suppletive forms, but also forms that share meaning and part of the form) are treated as separate roots despite their common meaning. We conclude this section by emphasizing, once again, that this decision was made for the purposes of the annotation, and is not intended to opt for any of the competing accounts on the morpho-phonological, semantic or syntactic nature of roots.

2.3.2 Prefixes

Prefixes are marked in six columns: PREFIXED_VERB, PREFIXES (LIST), 1, 2 and 3, and POTENTIALLY PREFIXED. The column PREFIXED_VERB shows whether the verb has at least one prefix (1) or not (0). In the column PREFIXES (LIST), we show all the prefixes that the verb contains. Where there is more than one prefix, the prefixes are entered in their linear order separated by a "+". If the verb is not prefixed, a 0 is entered. This is illustrated in Table 2.23.

Table 2.23: Prefixes – Examples of lists

Language	Verb	Translation	List of prefixes
BCMS	pisati	'to write'	0
BCMS	pročitati	'to read'	pro
BCMS	otpočeti	'to start'	od+po
BCMS	sporazumeti	'to agree'	s+po+raz
Slovenian	brati	'to read'	0
Slovenian	pročitati	'to read	pro
Slovenian	razočarati	'to disappoint'	raz+o
Slovenian	prerazporediti	'to redistribute'	pre+raz+po

The next three columns (1, 2, 3) show specific prefixes, as well as the position of each individual prefix with respect to the root: 1 refers to the position closest to the root, 2 to the second (middle) one, and 3 to the prefix that is most distant from the root. This means that the order of prefixes in these three columns is reversed compared to their linear order in the verb itself (and in the column **Prefixes (list)**). If there is no prefix in the respective position, we annotate this with 0. This is illustrated in Tables 2.24 and 2.25 for BCMS and Slovenian, respectively. Importantly, this annotation only marks the linear position of the prefix – we are making no claims about the syntactic position of each prefix (whether it is lexical, intermediate, super-lexical, etc., for more on the topic, see, a.o., Milićević 2004, Svenonius 2004b, Žaucer 2009, 2013).

Table 2.24: Prefixes – Examples of BCMS prefixes in three positions

Verb	Translation	1	2	3
pisati	'to write'	0	0	0
pročitati	'to read'	pro	0	0
otpočeti	'to start'	po	od	0
sporazumeti	'to come to an agreement'	raz	po	s

Table 2.25: Prefixes – Examples of Slovenian prefixes in three positions

Verb	Translation	1	2	3
brati	'to read'	0	0	0
nabrati	'to gather'	na	0	0
razočarati	'to disappoint'	o	raz	0
prerazporediti	'to redistribute'	po	raz	pre

The general guideline that we followed in determining if the verb is prefixed or not is 'if it can be taken to be a prefix, then it is a prefix', where the starting point were existing lists of prefixes in BCMS and Slovenian (Babić 2002, Klajn 2002 for BCMS, Toporišič 2004 for Slovenian). To this end, the first step was to determine if a verb that seems to be prefixed has an unprefixed pair. For example, the Slovenian verb *prebrati* 'to finish reading' has the pair *brati* 'to read', and its BCMS verbs cognate *pročitati* 'to finish reading' also has the unprefixed counterpart *čitati* 'to read'. Consequently, *pre-* and *pro-* are annotated as prefixes. If the seemingly prefixed verb does not have an unprefixed pair, we checked if a verb with the same root but a different prefix exists. If it does, the verb was taken to be prefixed and annotated as such. Examples of this kind include all the verbs with the form *četi* in both BCMS and Slovenian, e.g. in Slovenian *po-četi* 'to do', *na-četi*, *za-četi* 'start' etc. Here many verbs with different prefixes are attested, but *četi* itself is not attested as an unprefixed verb in either of the two languages.

If there were no corresponding unprefixed verbs or verbs with the same root and other prefixes, the respective verbs were still marked as prefixed if their root is independently attested in other categories (such as nouns, adjectives, compounds). Examples of this kind include e.g. *uz-ne-miriti* 'upset' in BCMS, where *nemiriti* is not attested on its own or with other prefixes, but the noun *nemir* 'unrest, lit. not-peace' exists.

Finally, if the verb cannot be annotated as prefixed using the criteria described above, but there is a strong intuition that the verb contains a prefix, this is marked in column POTENTIALLY_PREFIXED_VERBS with 1 (and 0 otherwise). For instance, in BCMS, the verb *za-taškati* 'to cover up' seems to include the prefix *za-*, which retains the same meaning (roughly 'behind') as in other verbs with *za-*, e.g. *za-kopati* 'to burry', related to *kopati* 'to dig'. However, the base *tašk* is not attested independently or with other prefixes (including categories other than verbs). A similar example from Slovenian is the already mentioned *odmevati* 'to echo', which seems to contain the prefix *od-*, retaining the same meaning ('from') as in other verbs with this prefix, e.g. *od-rezati* 'to cut off', related to *rezati* 'to cut'. And yet, *mevati* is not attested neither on its own nor with any other prefix.

As is clear from this section, we do not distinguish between different homophonous prefixes, by separating e.g. the distributive prefix *po-* from the delimitative prefix *po-*. There is an ongoing debate in the literature on whether such uses should be treated as instances of polysemy or homonymy (see e.g. Milosavljević 2022 for a recent overview of the debate). The general principle 'what you see is what you get' was followed, i.e. the precedence was given to the phonological identity (just as in the case of roots; see §2.3.1.1). One of the most radical examples in this regard comes from the annotation of the Slovenian prefix *s/z/se-* (or, pretheoretically, prefixes *s*, *z*, and *se*). Specifically, in the Slovenian sub-database, these three units in verbs like *s-pisati* 'to write up', *z-gledati* 'to appear' and *se-stati* (*s[ə]-stati*) 'to meet' were annotated as separate items, although they can be plausibly analyzed as phonologically conditioned variants of a single prefix, see for example Toporišič (2004: 223). Despite this, they were annotated as different prefixes in *WeSoSlaV*, but we note that for the purposes of Chapter 4, they were counted as a single item.

2.3.2.1 Issues with the annotation

The greatest difficulty for the annotation of prefixes was to establish a boundary between a (potential) prefix and the respective root it combines with. As described in the previous subsection, we handled these cases by marking such verbs as potentially prefixed and by listing the relevant prefixes with a colon. Here we discuss some additional difficulties in the annotation of prefixes.

Firstly, in some cases, it is unclear whether the item in question is a single prefix or if it could be further separated. One such example is Slovenian *izpod-*, where prefixes *iz-* and *pod-* also exist (as do prepositions *iz*, *pod* and *izpod*), see Žaucer (2002) for a discussion on such examples. In *WeSoSlaV*, we decompose these prefixes. However, a potential problem for this decision comes from verbs

like *iz-pod-biti* 'to rebut', where the version with the root-adjacent prefix alone (*pod-biti*) is not attested. This situation, however, resembles instances with 'simplex' prefixes in situations in which the unprefixed verb is not attested, as with the root *čet* discussed in the previous section (*po-četi*, *za-četi*, etc., but **četi*).

Another dilemma in the annotation process emerged from verbs with loan prefixes, such as *re-* in *reciklirati* 'to recycle' (both languages) or *de-* in the BCMS *destabilizovati* 'to destabilize'. The final decision was not to annotate such prefixes as prefixes. The rationale behind the decision is that they behave differently from native prefixes in a systematic way: while 'true', native prefixes retain their perfectivizing function when combining with new verbs, including borrowed ones (e.g., *s-bindžovati* in BCMS or *z-bindžati* in Slovenian, both with the meaning 'to binge') must be perfective, verbs with loan prefixes are systematically biaspectual and hence pass the test for imperfectivity (e.g., above mentioned *reciklirati* or *destabilizovati*). This in turn suggests that they are borrowed as an integral part of the base/root, and do not have functions characteristic of Slavic prefixes. Consequently, verbs like *stabilizovati* 'to stabilize' and *destabilizovati* 'to destabilize' are marked as having different roots in BCMS (*stabil+iz* and *destabil+iz*, respectively). We are aware that such a decision leaves these verbs 'unlinked' in our database and therefore does not enable a possible analysis of the borrowed prefixes themselves. Still, we leave their annotation as a task for future work.

2.3.3 Compounds

One column (the column Compound) in the database is devoted to compounds. If the base is a compound, this is marked with 1, and if it is not, with 0. The relevant criteria here is whether the base consists of more than one lexical root. The relevant example is *blag-o-slov-i-ti* 'to bless' in both BCMS and Slovenian, which consists of two roots: *blag* 'gentle' and *slov* 'word' (the morpheme -o- functions as a linking vowel); see additional examples in Table 2.22 above. As noted in section 2.3.1, parts of the compounds are also annotated in the column Root.

Apart from such typical examples with two lexical roots, the following items are also considered compounds in *WeSoSlaV* (and thus marked with 1): (i) examples with the negation particle, e.g., BCMS *ne-stati* 'to disappear' (lit. not+stop) or *ni-sam* 'I am not' (*ni* – negation particle, *sam* – auxiliary); Slovenian *vz-ne-miriti* 'to upset' from *ne-mir* 'unrest' (lit. not+peace); (ii) examples with prepositions that do not function as prefixes, e.g. *bez* 'without' in BCMS *o-bez-hrabr-i-ti* 'to discourage' and in Slovenian *zoper-staviti* 'to oppose' (lit. against+put).

2.3.4 Suffixes

WeSoSlaV comprises multiple columns devoted to individual suffixes appearing in the verbs. We have identified three broad classes of suffixes:

1. Verbal suffixes. These include:
   - secondary imperfectivizers consisting of more than just theme vowels;
   - verbalizers used (also) in loanword integration which consist of more than just theme vowels;
   - other suffixes with a clear semantic contribution in the verbal domain (typically the aspectual semantics).
2. Multifunctional suffixes.
3. Suffix-like items and other 'borderline' morphemes.

The column VERBAL SUFFIX (Y/N) reflects a summary of all individual verbal suffixes columns – verbs that have no verbal suffix are marked with a 0, whereas verbs that have a verbal suffix, get a 1.

In the subsequent sections, we turn to the description of all three groups of suffixes introduced above.

2.3.4.1 Verbal suffixes

For the purposes of the annotation in *WeSoSlaV*, we differentiate between theme vowels (see §2.3.4.1) and verbal suffixes. The latter are taken to be suffixes that have one of the following functions: verbalizing, aspectual and/or modifying function (e.g., diminutive).

Verbal suffixes are annotated in two ways. As already noted, in column VERBAL SUFFIX (Y/N), it is simply shown if a verb has a verbal suffix (1) or not (0). It is then specified which verbal suffix the verb has in separate columns devoted to each suffix, where 1 indicates that the given verb has that suffix. This means that we assumed a list of specific verbal suffixes for each individual sub-database. The lists of verbal suffixes annotated in this way are provided in Tables 2.26 and 2.27 for BCMS and Slovenian, respectively. Since suffixes end in TVs, and some suffixes actually combine with multiple TVs (e.g. *-t* in both languages), we list the labels for each suffix as combinations of the suffix proper and a TV (separated by the plus symbol).

Table 2.26: Verbal suffixes in BCMS

Suffix	Verb	Translation
av+a	iz-beg-av-a-ti$^{\text{IPFV}}$	'to avoid'
ek+a	prd-ek-a-ti$^{\text{IPFV}}$	'to fart'
et+a	lup-et-a-ti$^{\text{IPFV}}$	'to fool around'
ir+a	aktiv-ir-a-ti$^{\text{IPFV}}$	'to activate'
is+a	afirm-is-a-ti$^{\text{IPFV}}$	'to affirm'
iv+a	do-ček-iv-a-ti$^{\text{IPFV}}$	'to welcome'
k+a	čač-k-a-ti$^{\text{IPFV}}$	'to poke around'
	pre-ćut-k-iv-a-ti$^{\text{IPFV}}$	'keep quiet repeatedly'
n+0	trep-nu-ti, trep-n-e-m$^{\text{PFV}}$	'to blink'
	sta-∅-ti, sta-n-e-m$^{\text{PFV}}$	'to stop'
	na-dah-nj-iv-a-ti, na-dah-nj-uj-e-m$^{\text{IPFV}}$	'to inspire'
ov+a	ver-ov-a-ti$^{\text{IPFV}}$	'to believe'
t+a/a	za-huk-t-a-ti, za-huk-t-a-m$^{\text{PFV}}$	'to heat up'
t+a/je	tre-p-t-a-ti, trep-ć-e-m$^{\text{IPFV}}$ < trep-t-je-m	'to blink'
t+e/i	drh-t-je-ti, drh-t-i-m$^{\text{IPFV}}$	'to tremble'
uc+a	pij-uc-k-a-ti$^{\text{IPFV}}$	'to sip'

2.3.4.2 Multifunctional suffixes

Suffixes that do not have a verbalizing, aspectual or modifying function are annotated as multifunctional as they (i) potentially originate from a different category and (ii) have a vague meaning. These suffixes are annotated in a single column called MULTIFUNCTIONAL SUFFIXES. We mark if the verb has such a suffix with 1 (if not, 0). A discussion and exhaustive lists of multifunctional suffixes will be provided in Chapter 6. For completeness, we illustrate a subset of these suffixes for both BCMS and Slovenian in Tables 2.28 and 2.29, respectively. Recall also that multifunctional suffixes are also given in the ROOT column (separated by the plus symbol from the root), and that the lists of these suffixes are different for BCMS and Slovenian.

2.3.4.3 Suffix-like items

If the verb apparently has the form (prefix+)root+TV, but there are reasons to assume a more complex structure, this is indicated with a 1 in column SUFFIX_LIKE.

Table 2.27: Verbal suffixes in Slovenian

Suffix	Verb	Translation
av+a	o-dobr-av-a-ti^{IPFV}	'to approve'
ev+a	pre-hit-ev-a-ti^{IPFV}	'to overhaul'
et+a	lesk-et-a-ti^{IPFV}	'to glitter'
ik+a	svet-l-ik-a-ti^{IPFV}	'to gleam'
ir+a	identific-ir-a-ti^{IPFV}	'to identify'
k+a	tip-k-a-ti^{IPFV}	'to type'
lj+a	pri-cur-lj-a-ti^{PFV}	'to trickle out'
n+i/e	kih-n-i-ti^{PFV}	'to sneeze'
ov+a	vz-dih-ov-a-ti^{IPFV}	'to sigh'
	dež-ev-a-ti^{IPFV}	'to rain'
t+a	tep-t-a-ti^{IPFV}	'to trample'
t+e	iz-puh-t-e-ti^{IPFV}	'to evaporate'
v+a	pri-dob-i-v-a-ti^{IPFV}	'to extract'

Table 2.28: Multifunctional suffixes in BCMS *WeSoSlaV*

Suffix	Verb	Translation
ać	o-dom-ać-i-ti^{PFV}	'to domesticate'
ač	raz-tum-ač-i-ti^{PFV}	'to clarify'
ar	krv-ar-i-ti^{IPFV}	'to glitter'
ev	o-duš-ev-i-ti^{PFV}	'to enthuse'
iz	simbol-iz-ova-ti^{IPFV}	'to symbolize'

All types of examples that are annotated as suffix-like in the BCMS sub-database are listed in the tables below (for the rationale behind the assumed segmentation, see Simonović et al. 2023[12]). For Slovenian, this column is intended for secondary imperfective verbs in which we do not necessarily have phonologically overt material (which is annotated in the verbal suffixes columns). Two indicators of this additional structure are the ablaut (the root exhibits a vowel change) and a *-j-*. Note that while *-v-* has the same distribution in the two languages, for Slovenian,

[12]For BCMS, we mark certain morphemes with a mora (μ). These are the morphemes that arguably give rise to vowel lengthening in the root.

Table 2.29: Multifunctional suffixes in Slovenian *WeSoSlaV*

Suffix	Verb	translation
oš	pust-oš-i-ti$^{\text{IPFV}}$	'to ravage'
n	bes-n-e-ti$^{\text{IPFV}}$	'to rage'
š	z-manj-š-a-ti$^{\text{PFV}}$	'to reduce'
ov	raz-pol-ov-i-ti$^{\text{PFV}}$	'to halve'
išč	sklad-išč-i-ti$^{\text{IPFV}}$	'to store'

we annotated it as a verbal suffix (on par with imperfectivizing suffixes *ava, ova, eva*). The Slovenian and the BCMS sub-databases are thus different in this respect, as some of the arguments in Simonović et al. (2023) for the non-suffixal nature of *v* in BCMS do not extend to Slovenian.

2.3.5 Morphological complexity of the verb

The only column explicitly encoding the morphological complexity of the verb in both BCMS and Slovenian sub-bases is the column SIMPLEX_VERB. Verbs that consist of only a root, TV, and an inflectional ending (traditional simplex or simple verbs), are marked with 1, and all other verbs are marked with 0. In other words, all the verbs that contain any prefixes or suffixes, including potential prefixes and suffix-like elements, are considered derived, and receive the value 0. Since all these affixes were illustrated in the previous sections, in Table 2.32, we provide several examples of simplex verbs as annotated in *WeSoSlaV*.

Of course, once equipped with all the information on derivational morphology presented in this section, the user can easily derive other derivational categories of verbs, such as 'unprefixed verbs' (which would include verbs that don't contain a prefix, but may potentially contain suffixes), 'suffixed verbs', as a category that contains a suffix irrespective of other morphemes (e.g. includes both prefixed and unprefixed verbs), etc. Much of this derived information has been exploited in Chapter 4, where we present the properties of verb derivation that rely on the columns described in the present section.

2.4 Deverbal derivation

Two columns are devoted to deverbal derivation in both sub-databases. One focuses on the *-nje/-tje/-će* derivations (*-nje/-tje* in Slovenian and *-nje/-će* in BCMS),

Table 2.30: Suffix-like items in BCMS *WeSoSlaV*

Example	Assumed segmentation	Translation
padati$^{\text{IPFV}}$	pad+∅+a+ti	'to fall'
skretati$^{\text{IPFV}}$	skret+∅+a+ti	'to diverge'
davati$^{\text{IPFV}}$	da+∅+a+ti	'to give'
duvati$^{\text{IPFV}}$	du+∅+a+ti	'to blow'
snabdevati$^{\text{IPFV}}$	snabd+e+∅+a+ti	'to supply'
primati$^{\text{IPFV}}$	prim+i+μa+ti	'to receive'
stizati$^{\text{IPFV}}$	stig+∅+[+sibilant]a+ti	'to arrive'
zamišljati$^{\text{IPFV}}$	zamisl+i+μa+ti	'to imagine'
skrivati$^{\text{IPFV}}$	skri+∅+μa+ti	'to hide'
upijati$^{\text{IPFV}}$	up+i+μa+ti	'to absorb'
voziti$^{\text{IPFV}}$	vez+∅+o-i+ti	'to drive'
odmicati$^{\text{IPFV}}$	odm-k+i[+sibilant, -velar]-a+ti	'to move away'
uzdisati$^{\text{IPFV}}$	uz+d-h+i[+sibilant, -velar]-a+ti	'to sigh'
ubirati$^{\text{IPFV}}$	ub-r+a+i-a+ti	'to harvest'
obasipati$^{\text{IPFV}}$	obas-p+∅+i-a+ti	'to shower'
uplitati$^{\text{IPFV}}$	uplet+∅+i-a+ti	'to plait'
umirati$^{\text{IPFV}}$	umr+∅+i-a+ti	'to die'
oduzimati$^{\text{IPFV}}$	oduz-m+∅+i-a+ti	'to take away'
sklapati$^{\text{IPFV}}$	sklop+i+μa-a+ti	'to assemble'
počinjati$^{\text{IPFV}}$	[Supp]	'to begin'
odlaziti$^{\text{IPFV}}$	[Supp]	'to leave'

while the other focuses on *-lec/-lac* derivations (*-lec* in Slovenian, *-lac* in BCMS). These two types of nominalizations were chosen as the first takes as its base the passive or N/T-participle of the verb and the second takes the past or L-participle (though see Simonović et al. 2024 for a different view of the nature of the respective bases).

The annotation was conducted as follows:

- **Step 1:** A human annotator who is a native speaker of the language and a linguist applied the derivational operation to each verb, and then judged whether the derived noun is acceptable. If it was acceptable, the derived form was entered, if unacceptable cardinal 0. In some cases, the operation had to be applied in a special way, for instance by changing the theme

Table 2.31: Suffix-like items in Slovenian *WeSoSlaV*

Example	Assumed segmentation	Translation
padati$^{\text{IPFV}}$	pad+∅+a+ti	'to fall'
dosegati$^{\text{IPFV}}$	doseg+∅+a+ti	'to reach'
ubirati$^{\text{IPFV}}$	u+br+i+a+ti	'to choose'
umirati$^{\text{IPFV}}$	u+mr+i+a+ti	'to die'
sprejemati$^{\text{IPFV}}$	s+pre+jem+∅+a+ti	'to accept'
zamišljati$^{\text{IPFV}}$	za+misl+i+a+ti	'to imagine'
poglabljati$^{\text{IPFV}}$	po+glob+a+i+a+ti	'to deepen'
prevajati$^{\text{IPFV}}$	pre+ved+a+i+a+ti	'to translate'
ogrožati$^{\text{IPFV}}$	o+groz+i+a+ti	'to threaten'
jemati$^{\text{IPFV}}$	[Supp]	'to take'
prihajati$^{\text{IPFV}}$	[Supp]	'to arrive'

Table 2.32: Simplex verbs in *WeSoSlaV*

Language	Verb		Translation
BCMS	čit-a-ti$^{\text{IPFV}}$	[read-TV-INF]	'to read'
BCMS	rad-i-ti$^{\text{IPFV}}$	[work-TV-INF]	'to work'
Slovenian	b-a-ti$^{\text{IPFV}}$	[afraid-TV-INF]	'to be afraid'
Slovenian	del-a-ti$^{\text{IPFV}}$	[work-TV-INF]	'to work'

vowel of the verb (e.g. replacing the TV ∅/e in the BCMS verb *vesti*, *veze* with the vowel *i* to derive *vezilac*, instead of the blocked **vezlac*). In cases where the annotator was not absolutely certain in their judgment, they first searched the respective corpus for the derived form (the Slovenian national corpus *Gigafida 2.0* (2019), the Croatian web corpus CLASSLA-web.hr (Ljubešić et al. 2024a) or the Serbian web corpus CLASSLA-web.sr (Ljubešić et al. 2024b)), and if examples were found with the intended interpretation, the noun was entered as acceptable. For BCMS, three additional steps were undertaken:

- **Step 2:** Another human annotator repeated the same procedure.
- **Step 3:** The two annotations were compared, differences were discussed, comments entered.

- **Step 4:** The first human annotator went through the results of the comparison, and made final decisions for all verbs remaining unresolved after Step 3.

2.5 Verbal aspect

The only column related to the verbal aspect that was entered in the final version of *WeSoSlaV* is the column IMPERFECTIVE ASPECT in which it is marked if a verb passes the test for imperfectivity (1) or not (marked with 0). We used the following test as the main test for BCMS: A verb is imperfective if it can be used in the present tense to answer the question of what is happening right now, as illustrated in (4). This test also returns traditional biaspectual verbs as imperfective ones, as in (5).

(4) a. Šta se trenutno dešava?
what REFL right_now happens
'What is going on right now?'

b. Jovan otvara^{IPFV} / *otvori^{PFV} prozor.
Jovan opens / opens window
'Jovan is opening the window.'

(5) a. Šta se trenutno dešava?
what REFL right_now happens
'What is going on right now?'

b. Jovan downloduje film.
Jovan downloads movie.
'Jovan is downloading a movie.'

For BCMS, the test with phasal verbs was used as an additional test in cases of doubt, while for Slovenian, it was used as the only test. According to this test, if a verb can be used as a complement of a phasal verb, it is imperfective; otherwise, it is perfective, as in (6) and (7) from BCMS, and (8) from Slovenian. This test too delivers biaspectual verbs as imperfective ones, as shown by (7) and (9).

(6) Jovan je počeo da peva^{IPFV} / *od-peva^{PFV}
Jovan AUX begin.TV.PTCP.M COMP sing.PRS.3SG from-sing.PRS.3SG
pesmu.
song.ACC.SG
'Jovan began to sing a song.'

(7) Jovan je počeo da downloduje^{IPFV} film.
Jovan AUX begin.TV.PTCP.M COMP download.PRS.3SG movie.ACC.SG
'Jovan began to download a movie.'

(8) Maja je začela teči^{IPFV} / *od-teči^{PFV} maraton.
Maja AUX begin.TV.PTCP.F run from-run marathon.ACC.SG
'Maja started to run a marathon.'

(9) Maja je začela improvizirati govor.
Maja AUX begin.TV.PTCP.F improvise speech.ACC.SG
'Maja started improvise a speech.'

The fact that the two imperfectivity tests introduced above unify imperfective and biaspectual verbs can be seen as a drawback from the perspective of the majority of traditional approaches. There are, however, recent analyses that treat imperfective and biaspectal verbs as a natural class of imperfectives (e.g., Arsenijević 2018; see Milosavljević 2023b for a recent approach and a critical overview). This is in accordance with the problem of finding a positive test for perfectivity: most imperfective verbs pass all the available perfectivity tests (see Milosavljević 2023b:§4.2 for an overview). This has led some researchers to propose that imperfective verbs are unspecified for grammatical aspect and thus compatible with both imperfective and perfective viewpoints (cf. Łazorczyk 2010, Arsenijević 2018, 2023, Milosavljević 2023b,a). On this view, the lack of a positive test for perfectivity follows naturally, as all verbs can be used under the perfective viewpoint. For the purposes of *WeSoSlaV*, it suffices to assume that the imperfectivity tests described above detect verbs that pass the test as imperfective and those that do not, the latter clearly being perfective. The nature of the group that passes the imperfectivity tests (only imperfective, biaspectual, unspecified, etc.), however, is to be subjected to further research.

2.6 Summary

The chapter has introduced the principles of (and instructions for) the annotation of the verbal properties in *WeSoSlaV*, a database of two Western South Slavic languages: BCMS and Slovenian. Specifically, after introducing the basic infrastructure beyond compiling the two sub-databases (BCMS and Slovenian), the remainder of the chapter presented the properties of verbal inflection, verbal derivation, deverbal derivation and verbal aspect as annotated in *WeSoSlaV*.

3 The domain of verbal inflection

3.1 Introduction

This chapter focuses on the information relevant for the domain of verbal inflection included in the database. The issue of identifying the forms that belong to the verbal inflection (or the verbal paradigm) is by no means uncontroversial. For example, all descriptive works on WSS (e.g., Barić et al. 1997, Klajn 2002, Šekli 2010, Toporišič 2000) include under the rubric of verbal inflection forms like the infinitive, the present tense, the imperative, the L-participle and the passive participle. And yet, the latter form is typically described as intermediate between verbal and adjectival inflection, because WSS passive participles also decline for case, number and gender, but also because it has been argued to include derivational affixes used to derive adjectives from other categories (Simonović & Arsenijević 2020). Some descriptions also include, e.g. the deverbal noun in *-je* as part of the verbal inflection, e.g., Toporišič (2004: 345), see also Klajn (1998). While typically implicit, five criteria appear to play a role in the inclusion of a form in the list of verbal inflections: realizing the verb-specific category of person, not realizing case/number/gender, applicability to all verbs, strict semantic compositionality (i.e. unavailability to lexicalization) and not involving clearly derivational material.

Independently of the specific assumptions about the expansion of the inflectional paradigm, there is a consensus that a relatively small number of forms is sufficient to capture the lexical phonological content of the verbal lexeme. Since the database is structured to contain the relevant features of specific verbs, our main task in setting up the part of the database related to verbal inflection involved identifying all the phonological material that needs to be included in the lexical representation of WSS verbs in order to predict the realization of their inflection paradigms, as well as determining which of these pieces are inserted in which specific contexts. To illustrate this with an example, all WSS PRS.1PL forms end in *-mo*, making *-mo* an inflectional ending that does not constitute a part of the lexical representation of any verb. This morpheme is therefore not to be found in the database. However, the piece of morphology that precedes *-mo*,

traditionally referred to as the theme vowel of the present stem, is far less predictable. For instance, the Slovenian verbs *pisati* 'to write' and *kisati* 'to pickle' have the PRS.1PL forms *pis-je-mo* [piʃemo] and *kis-a-mo*, respectively. The present tense theme vowel, therefore, is not predictable and must be included in (or be predictable from) the lexical representation. Relatedly, prosody varies greatly between the two verbal stems, which is why prosodic information about both stems is arguably both lexical and inflection-related (it has even been argued that the lexical prosody of WSS verbs can be analyzed as specified on the theme vowel only, Simonović 2022a,b).

The database (Marušič et al. 2022) follows a long tradition of descriptive work in treating two verbal forms — the infinitive and the present tense — as sufficiently informative to determine a verb's conjugation pattern, that is, as the principal parts (Stump & Finkel 2013) of the verbal paradigm. The rationale is that these two forms contain the two realisations of the theme vowel attested across the verbal paradigm. This approach is consistent with a tradition going back at least to Leskien (1871) and Miklosich (1874), in which Slavic verbs are discussed as involving two stems. Throughout this chapter, we therefore focus on these two stems, without taking a stance on the precise set of forms constituting the verbal inflectional paradigm in BCMS and Slovenian (but see Caha & Scheer 2008 for the argument that in Czech the L-participle is a better predictor of the underlying representation than the infinitive). The two forms are annotated for each verb: the infinitive in the column called *Dict.Form* and the present tense (in the third person singular, as its morphologically least marked form) in the column *3P.SG present tense*. The theme vowel was determined for each form and the combination of the two theme vowels was annotated in the column *Theme Vowels*. If the comparison of the infinitive and the present tense forms showed differences in the root which could not be reduced to predictable interactions with the theme vowels, this was marked in the column *Root allomorphy (y/n)*. For verbs which display root allomorphy, the specific allomorphs are annotated in the column *Root allomorphs (list)*. A final piece of lexical information annotated for each verb was prosody. The prosodic pattern was annotated both for the infinitive and for the PRS. For Slovenian, only the prosodic prominence (realized as stress) was annotated. For BCMS, on the other hand, both the prosodic prominence (realized as High tone), and vowel length are annotated. For more details on the specific choices, see Chapter 2.2.3.

The remainder of this chapter is organized as follows. In 3.2 we discuss the segmental content of the theme vowels and its interaction with inflectional material. In 3.3 we consider the prosodic features of the theme vowel classes in an

attempt to answer the question where the prosodic specification of verbs comes from. Finally, 3.4 concludes the chapter.

3.2 Theme-vowel classes in WSS and interaction with further pieces of morphology

We identified ten theme vowel classes (plus a small class of defective verbs) in WSS. Below we provide theme vowel classes arranged by size in Marušič et al. (2022). Tables 3.1 and 3.2 contain the relevant figures for the two languages.

Table 3.1: Theme vowel classes by size in BCMS

TV class	N (%)	INF	PRS.1PL	GLOSS
a/a	1702 (32.1%)	pit-a-ti	pit-a-mo	'ask'
i/i	1601 (30.2%)	vis-i-ti	vis-i-mo	'hang'
a/je	1029 (19.4%)	plak-a-ti	plač-e-mo < /plak-je-mo/	'cry'
∅/e	298 (5.6%)	pas-∅-ti	pas-e-mo	'graze'
u/e	258 (4.9%)	gurn-u-ti	gurn-e-mo	'push'
e/i	184 (3.5%)	gor-(j)e-ti	gor-i-mo	'burn'
∅/ne	124 (2.3%)	sta-∅-ti	sta-ne-mo	'stop'
a/i	62 (1.2%)	struj-a-ti	struj-i-mo	'flow'
a/e	19 (0.4%)	hrv-a-ti	hrv-e-mo	'wrestle'
e/e	17 (0.3%)	razum-(j)e-ti	razum-(ij)e-mo	'understand'
defective	6 (0.1%)	ht(j)e-ti	hoće-mo	'want'

The overall picture is quite comparable for the two languages: a/a and i/i are by far largest, followed by a/je and ∅/e, whereas all the other classes have less than 5% of all verbs.

The overviews presented so far do not take into consideration that multiple verbs have the same roots and verbal suffixes, which are typically always followed by the same theme vowel. For instance, BCMS verb *pit-a-ti* 'to ask' and its prefixed versions *upit-a-ti* and *zapit-a-ti* 'to ask' count as three different verbs of the a/a class, as do *imit-ir-a-ti* 'to imitate', *irit-ir-a-ti* 'to irritate' and *organiz-ir-a-ti* 'to organize', all containing the derivational suffix *-ir-*, which is always followed by the theme vowel a/a. It is therefore relevant to consider the sizes of theme vowel classes in simplex verbs, i.e. verbs that only display a root, a theme vowel and an inflectional ending. Simplex verbs are annotated in the column *Simplex_verbs* in Milosavljević et al. (2023). If only simplex verbs are taken

Table 3.2: Theme vowel classes by size in Slovenian

TV class	N (%)	INF	PRS.1PL	GLOSS
a/a	1044 (34.8%)	del-a-ti	del-a-mo	'work'
i/i	863 (28.8%)	del-i-ti	del-i-mo	'divide'
a/je	378 (12.6%)	ples-a-ti	pleš-e-mo < /ples-je-mo/	'dance'
∅/e	284 (9.5%)	pas-∅-ti	pas-e-mo	'graze'
i/e	142 (4.7%)	str-n-i-ti	str-n-e-mo	'condense'
e/i	128 (4.3%)	gor-e-ti	gor-i-mo	'burn'
e/e	47 (1.6%)	razum-e-ti	razum-e-mo	'understand'
a/e	46 (1.5%)	br-a-ti	ber-e-mo	'read'
a/i	36 (1.2%)	sp-a-ti	sp-i-mo	'sleep'
∅/ne	27 (0.9%)	osta-∅-ti	osta-ne-mo	'stay'
defective	3 (0.1%)	ime-ti	ima-mo	'have'

into consideration, the BCMS sample shrinks to 679 verbs, whereas the Slovenian sample shrinks to 449 verbs. The tables 3.3 and 3.4 show the relevant figures.

Table 3.3: TV classes by size in BCMS simplex verbs

TV class	N (%)	INF	PRS.1PL	GLOSS
i/i	324 (47.7%)	vis-i-ti	vis-i-mo	'hang'
a/a	187 (27.5%)	pit-a-ti	pit-a-mo	'ask'
a/je	50 (7.4%)	plak-a-ti	plač-e-mo < /plak-je-mo/	'cry'
e/i	50 (7.4%)	gor-(j)e-ti	gor-i-mo	'burn'
∅/e	36 (5.3%)	pas-∅-ti	pas-e-mo	'graze'
a/i	19 (2.8%)	struj-a-ti	struj-i-mo	'flow'
a/e	5 (0.7%)	hrv-a-ti	hrv-e-mo	'wrestle'
e/e	3 (0.4%)	sm-(j)e-ti	sm-(ij)e-mo	'be permitted'
defective	5 (0.7%)	ht(j)e-ti	hoće-mo	'want'
u/e	0 (0%)			
∅/ne	0 (0%)			

Again, the two languages show a very similar relation between the counts based on all verbs and those based on simplex verbs only. Among the two largest classes, i/i is considerably larger than a/a if simplex verbs are considered. Among

Table 3.4: TV classes by size in Slovenian simplex verbs

TV class	N (%)	INF	PRS.1PL	GLOSS
i/i	183 (40.8%)	del-i-ti	del-i-mo	'divide'
a/a	127 (28.3%)	del-a-ti	del-a-mo	'work'
e/i	41 (9.1%)	gor-e-ti	gor-i-mo	'burn'
∅/e	40 (8.9%)	pas-∅-ti	pas-e-mo	'graze'
a/je	23 (5.1%)	ples-a-ti	pleš-e-mo < /ples-je-mo/	'dance'
a/i	15 (3.3%)	sp-a-ti	sp-i-mo	'sleep'
e/e	10 (2.2%)	sm-e-ti	sm-e-mo	'be permitted'
a/e	6 (1.3%)	br-a-ti	ber-e-mo	'read'
defective	3 (0.7%)	ime-ti	ima-mo	'have'
∅/ne	1 (0.2%)	de-∅-ti	de-ne-mo	'say'
i/e	0 (0%)			

the smaller classes, a/je drops considerably, whereas the class characterized by the sequence u/e (in Slovenian i/e) disappears altogether.[1] On the other hand, the class e/i now ranks considerably above the 5% mark in both languages.

Having presented a first impression of the quantitative relations between the theme vowel classes in WSS, we can turn to the role that these classes play in WSS inflectional morphology. As mentioned at the beginning of this chapter, information extracted from the infinitive and present tense forms is generally sufficient to predict all the other forms in the paradigm. This of course does not mean that all pieces of morphology that get added to the theme vowel will always have the same phonological content. In other words, we may have to allow some allomorphic conditioning by theme vowels of further pieces of morphology. However, it is not clear whether there is any case in which allomorphic rules need to reference specific conjugation classes (i.e. theme vowel combinations) rather than the phonological content of the theme vowels. In order to illustrate this issue, in Tables 3.5 and 3.6 we provide two additional forms which can plausibly be derived taking the present tense theme vowel as a starting point to which further pieces of morphology are added: PRES.3PL and IMP.1PL.

In Tables 3.5 and 3.6 the theme vowel classes are arranged by the phonological content of the present tense theme vowel, so that it is easier to compare classes

[1]The disappearance of BCMS u/e and Slovenian i/e in the simplex count is due to the fact that this is the only theme vowel considered a derivational morpheme across the board in *WeSoSlaV*. For arguments supporting this analytical choice, see Štarkl et al. (2024).

Table 3.5: Three forms with the present tense theme vowel in BCMS

TV class	PRS.1PL	IMP.1PL	PRES.3PL	GLOSS
a/a	pit-a-mo	pit-aj-mo	pit-a-ju	'ask'
i/i	vis-i-mo	vis-i-mo	vis-e	'hang'
e/i	gor-i-mo	gor-i-mo	gor-e	'burn'
a/i	struj-i-mo	struj-i-mo	struj-e	'flow'
a/je	plač-e-mo	plač-i-mo	plač-u	'cry'
∅/ne	sta-ne-mo	sta-ni-mo	sta-nu	'stop'
∅/e	pas-e-mo	pas-i-mo	pas-u	'graze'
u/e	gurn-e-mo	gurn-i-mo	gurn-u	'push'
a/e	hrv-e-mo	hrv-i-mo	hrv-u	'wrestle'
e/e	razum-(ij)e-mo	razum-i-mo %razum-ej-mo	razum-e-ju/ razum-i-ju	'understand'

Table 3.6: Three forms with the present tense theme vowel in Slovenian

TV class	PRS.1PL	IMP.1PL	PRES.3PL	GLOSS
a/a	del-a-mo	del-aj-mo	del-a-jo	'work'
i/i	del-i-mo	del-i-mo	del-i-jo	'divide'
e/i	gor-i-mo	gor-i-mo	gor-i-jo	'burn'
a/i	sp-i-mo	sp-i-mo	sp-i-jo	'sleep'
a/je	pleš-e-mo	pleš-i-mo	pleš-e-jo	'dance'
i/e	str-ne-mo	str-ni-mo	str-ne-jo	'condense'
∅/ne	osta-ne-mo	osta-ni-mo	osta-ne-jo	'stay'
∅/e	pas-e-mo	pas-i-mo	pas-e-jo	'graze'
e/e	razum-e-mo	razum-i-mo	razum-e-jo	'understand'
a/e	ber-e-mo	ber-i-mo	ber-e-jo	'read'

which have the same theme vowel in a given form. If we first focus on the imperative forms in the two languages, it is clear that there is no single segmental exponent that can be added to the comparable present tense form in order to obtain the imperative form. Rather, there seems to be a more abstract piece of phonological material that serves as the exponent of the imperative morpheme, presumably a floating segment/element which can get realized on some vowels but not others. Specifically, this element can get realized on the theme vowel *i*

without changing it in any way and on a theme vowel (ending in) *e*, whereby this *e* gets raised to an *i*. However, this element cannot get realized on the theme vowel *a*, which is presumably why it is realized as a separate consonant. Crucially, for the purposes of our discussion, all *a*'s, *e*'s, and *i*'s display the same behavior independently of the theme vowel class they are associated with. For this reason, the mechanisms that regulate the exact phonological shape of the imperative forms need not refer to conjugation classes but can rather operate on the phonological material presented by the theme vowel.

Turning now to the PRES.3PL, the first observation is that the two languages differ significantly. In Slovenian, the exponent of 3PL is as predictable as that of 1PL and, as such, does not introduce any additional allomorphy. In BCMS, on the other hand, the situation is more complex. As in the previous cases, the theme vowel *a* is the only one that remains unchanged, and next to it, there is the 3PL exponent *ju*. All three theme vowels that otherwise surface as *i* get lowered to *e*. Finally, five out of six theme vowel classes that otherwise display an *e* in the present tense theme vowel have this *e* replaced by a *u*. Only the theme vowel of the e/e class surfaces as *e* in Ekavian and *i* in Ijekavian, to which a *ju* is added.

It goes beyond the scope of this chapter to propose a full analysis of the exponence pattern of PRES.3PL (but see Stojković & Simonović forthcoming for a proposal in terms of a common exponent for PRES.3PL involving highly abstract floating material). What is crucial for our purposes is that, even in this relatively complex exponence pattern, the theme vowel exponents *a*, *e*, and *i* generally behave alike. Tellingly, the odd one out is the theme vowel of the e/e class, which diachronically derives from the yat vowel and independently requires a special representation in Ijekavian, where it still alternates between [ije], [e], and [i].

Before turning to prosodic information, a remark is in order concerning aspects of the inflectional morphology which cannot be directly read off the sub-databases, but which can be additionally annotated using the sub-databases as a starting point. The sub-databases do not contain any information on the PASS.PTCP form because this form is by and large predictable from the theme vowel class. Yet there are areas of relative unpredictability. For instance, WSS verbs of the ∅/e class generally take the suffix *-en* in this form (e.g. *griz-∅-en* 'bitten' from *gris-∅-ti* 'bite'). However, BCMS verbs of the same theme-vowel class ∅/e whose root surfaces with a final [i] behave uniformly except in the PASS.PTCP form, where, preceding the ending *-en*, we in some cases observe [j] (first two verbs in Table 3.7) and in others [ʊ] (last two verbs). The same hiatus-resolving consonant then surfaces in the secondary imperfectivisation derived from these verbs. Pairs like [u-pi-j̲en] but [u-mi-ʊ̲en] indicate that the distinction is lexical, i.e. that the first two verbs have a different lexical representation from the second

two. Simonović et al. (2023) argue for the representations /pi^{j}/ and /bi^{j}/ versus /mi/ and /li/, where the superscripts stand for floating consonants.

Table 3.7: v and j insertion in the same context

PFV.INF	PFV.PRS.1PL	PFV.PASS.PTCP	IPFV.INF	IPFV.PRS.1PL	GLOSS
u-pi-∅-ti	u-pi-je-mo	u-pi-jen	u-pi-ja-ti	u-pi-ja-mo	'absorb'
za-bi-∅-ti	za-bi-je-mo	za-bi-jen	za-bi-ja-ti	za-bi-ja-mo	'stab'
u-mi-∅-ti	u-mi-je-mo	u-mi-ven	u-mi-va-ti	u-mi-va-mo	'wash'
za-li-∅-ti	za-li-je-mo	za-li-ven	za-li-va-ti	za-li-va-mo	'water'

In Slovenian, for the class discussed above there is no lexical distinction and all verbs of the theme-vowel class ∅/e whose root surfaces with a final [i] take the same PASS.PTCP ending (*-t*, e.g., in *u-mi-∅-t* 'washed' from *u-mi-∅-ti* 'to wash'). There are, however, other areas of relative unpredictability. A case in point is the palatalisation of dental obstruents in passive participles of the i/i class (see, e.g., Toporišič 2004). The *i/i* class features triplets like *ponuditi* 'offer', *začuditi* 'bewilder', *prisoditi* 'attribute', whose passive participles are *ponujen/ponuden*, *začuden* and *prisojen*, respectivelly ([j] being derived from /dj/). In each of these cases, the sub-databases can be easily complemented by additional annotation for quantitative insight on which a formal analysis can be based.

3.3 Prosody of WSS verbs

In traditional descriptions of WSS verbs, prosody is discussed separately for separate theme vowel classes (e.g., Šekli 2010, Klaić 2013). This practice is taken over by formal approaches (e.g., Langston 1997, Zec 2018), which typically focus on a limited number of frequent theme vowel classes. In this section, we consider WSS verbal prosody based on the *WeSoSlaV* sub-databases, where, as discussed above, prosody is described separately for the two verbal stems. This will put us in a position to quantitatively assess the assumption that theme vowels constitute the locus of lexical verbal prosody (Simonović 2022b,a).

Two main positions for prosodic prominence are identified in the annotated forms in WSS: the prominent syllable is either the theme vowel or the syllable preceding the theme vowel (the latter position will be referred to as *penultimate stem vowel*). Among the BCMS verbs in the sub-database, 4952 (93.4%) exhibit

their preferential prosodic pattern[2] for both the infinitive and the present tense within this two-syllable window. The same holds true for 2956 (98.5%) of the verbs in the Slovenian sample.

Determining the lexical prosody is complicated by the non-availability of a vowel in the theme vowel position in verbs of the classes ∅/e and ∅/ne. Therefore, we will disregard these classes in further discussion in this chapter.

The standard varieties of BCMS and Slovenian described in the *WeSoSlaV* databases manifest distinct prosodic systems: the targeted BCMS standard variety is a pitch-accent system, whereas the targeted Slovenian variety is a stress system. Hence, we will discuss the results for the two languages separately. Before delving into the separate discussions, we highlight one important commonality between the two systems: the three prosodic types encountered in the majority of WSS verbs. Table 3.8 illustrates the three types, where PSV indicates prominence on the penultimate stem vowel and TV prominence on the theme vowel.

Table 3.8: The three major prosodic types in WSS verbs

Language	Type	inf	prs.1pl	Gloss
Slovenian	PSV/PSV	ˈčist-i-ti	ˈčist-i-mo	‘clean’
BCMS	PSV/PSV	číst-i-ti	číst-iː-mo	‘clean’
Slovenian	TV/TV	loˈv-i-ti	loˈv-i-mo	‘chase’
BCMS	TV/TV	lov-í-ti	lov-íː-mo	‘chase’
Slovenian	TV/PSV	voˈz-i-ti	ˈvoz-i-mo	‘drive’
BCMS	TV/PSV	voz-í-ti	vóz-iː-mo	‘drive’

What is prominently lacking from the typology in Table 3.8 is a fourth type, wherein prosodic prominence would be on the penultimate stem vowel in the infinitive but on the theme vowel in the present tense. The way the three-way distinction is typically captured in terms of the distinction between prominence linked to a specific morpheme, floating prominence mark and a lack of prominence. For instance, Simonović (2022b), working on i/i verbs in BCMS, argues that all verbal lexical prosody is located on the theme vowel. The three BCMS verbs in Table 3.8 illustrate the three possible lexical constellations: verbs with the PSV/PSV pattern take a theme vowel with no lexical H(igh tone) /i/, those with the TV/TV pattern have a theme vowel linked to a H /í/, whereas verbs

[2]Since the sub-databases register multiple prosodic patterns for certain verbs (see 2.2.3), in this chapter we only consider the prosodic pattern annotated as most unmarked for each verb.

with the TV/PSV pattern have a theme vowel affiliated with a floating H /i, H/. Here, the two stems must differ in some respect so as to favour different attachment patterns for floating H's. In the Classical Neo-Štokavian variety discussed in Simonović (2022b), such a difference can be located in phonology: present-tense theme vowels are always long — or, more precisely, lengthened by the present-tense morpheme — and Simonović argues that lengthened vowels are dispreferred attachment sites for floating tone.

The overall distribution among the three patterns in the two languages is presented in Table 3.9. Note that these (and all subsequent) counts exclude verbs that do not adhere to one of the three prosodic patterns (e.g., by having their prominent syllable further towards the left edge of the word), as well as all verbs belonging to the theme vowel classes ∅/e and ∅/ne.

Table 3.9: The distribution of three major prosodic types in WSS verbs

Language	PSV/PSV N (%)	TV/TV N (%)	TV/PSV N (%)
Slovenian	1642 (62.7%)	594 (22.6 %)	387 (14.7 %)
BCMS	1206 (25.9%)	596 (12.8 %)	2852 (61.3 %)

The numbers in Table 3.9 reveal a distinct pattern in the two languages. Slovenian generally favors the PSV/PSV type, while BCMS shows a preference for the alternating TV/PSV type. However, this apparent difference underscores the fact that the two systems are not directly comparable. Unlike Slovenian, in BCMS, the prosodic patterns are significantly influenced by the lexical vowel length of the penultimate stem syllable (i.e. the syllable preceding the theme vowel). Table 3.10 presents a comparison between BCMS verbs with a long penultimate stem vowel (e.g. *p*[iː]*sati*, *p*[iː]*šemo* 'write') and those with a short penultimate stem vowel (e.g. *br*[i]*sati*, *br*[i]*šemo* 'delete'). For the purpose of these counts, we have excluded verbs annotated as displaying root allomorphy.

As can be seen from the counts in Table 3.10, the dominant prosodic type in verbs with a short penultimate stem vowel is PSV/PSV, just like in Slovenian. However, in the more numerous verbs with a long penultimate stem vowel, there is an overwhelming majority of the TV/PSV type.

In the subsequent sections, we will enhance the presented picture by examining the distribution of prosodic types within theme vowel classes in both languages. As previewed above, this discussion holds significance for the analysis of the verbal system, as it constitutes a crucial step in determining whether lexical prosody of WSS verbs is specified on the theme vowels.

Table 3.10: The distribution of three major prosodic types in BCMS verbs by the length of the penultimate stem vowel

Penultimate stem vowel	total N	PSV/PSV N (%)	TV/TV N (%)	TV/PSV N (%)
short	2060	1186 (57.6%)	430 (20.9 %)	444 (21.5 %)
long	2229	20 (0.9%)	151 (6.8 %)	2058 (92.3 %)

3.3.1 Prosody of Slovenian verbs

The counts provided for Slovenian thus far did not take into account the existence of repeating roots and suffixes, which typically exhibit the same prosodic patterns (e.g. ˈ*delati* 'work', *pre-*ˈ*delati* 'process' and *do-*ˈ*delati* 'elaborate' all have the same root and the same prosodic pattern). Therefore, it is beneficial to also examine the counts based on verbs annotated as simplex. Table 3.11 presents a comparison between prosodic types for all verbs and only simplex verbs.

Table 3.11: The distribution of three major prosodic types in Slovenian verbs Simplex verbs vs Simplex + Complex verbs

	Total N	PSV/PSV N (%)	TV/TV N (%)	TV/PSV N (%)
Simplex + Complex	2627	1642 (62.7%)	594 (22.6 %)	387 (14.7 %)
Simplex	393	223 (56.7%)	144 (36.6%)	26 (6.6%)

The picture in Table 3.11 shows that the ranking of the three prosodic patterns remains the same, although among the simplex verbs, both PSV/PSV and TV/PSV are somewhat smaller than in the overall count.

Tables 3.12 and 3.13 illustrate the distribution of the three prosodic patterns within the specific theme vowel classes for all verbs and just for simplex verbs, respectively. Recall that the theme vowel class i/e does not have any simplex members.

In Tables 3.12 and 3.13 the theme vowel classes are arranged by the percentage of the generally most common PSV/PSV pattern. Just considering this column already shows considerable quantitative differences between the theme vowel classes, although almost all classes have at least some members in this class. Potentially more relevant for a formal account are gaps in relatively big classes, which indicate that some theme vowels are incompatible with some of the lexical prosodic patterns. Perhaps the most telling is the lack of verbs in the classes

Table 3.12: The distribution of three major prosodic types in Slovenian verbs by TV class (Simplex + Complex)

TV class	Total N	PSV/PSV N (%)	TV/TV N (%)	TV/PSV N (%)
a/a	1042	967 (92.8%)	74 (7.1%)	1 (0.1%)
i/e	143	127 (88.8%)	/	16 (11.2%)
i/i	880	470 (53.4%)	351 (39.9%)	59 (6.7%)
a/je	376	73 (19.4%)	/	303 (80.6%)
a/i	36	4 (12.1%)	29 (87.9%)	/
e/e	10	1 (10%)	7 (70%)	2 (20%)
e/i	128	4 (3.1%)	124 (96.9%)	/
a/e	15	/	9 (60%)	6 (40%)

Table 3.13: The distribution of three major prosodic types in Slovenian simplex verbs by TV class

TV class	Total N	PSV/PSV N (%)	TV/TV N (%)	TV/PSV N (%)
a/a	126	111 (88.1%)	15 (11.9%)	/
i/i	185	98 (53%)	73 (39.5%)	14 (7.5%)
a/je	22	11 (50%)	/	11 (50%)
e/e	3	1 (33.3%)	2 (66.7%)	/
a/i	13	1 (7.7%)	12 (92.3%)	/
e/i	41	1 (2.4%)	40 (97.6%)	/
a/e	3	/	2 (66.7%)	1 (33.3%)
i/e	0	/	/	/

a/je and i/e which belong to the TV/TV pattern. The generalization seems to be that the two present tense theme vowels that are unstressable are exactly those that have a consonantal component.

We calculated Pearson's Chi-squared test for the crossing of the the variable of prosodic types and theme vowel classes, which suggests a different behavior of theme classes regarding the prominence patterns: χ^2 = 212.97, df = 12, p-value < 2.2e-16.

3.3.2 Prosody of BCMS verbs

In this section we examine the prosody of BCMS verbs, which, as demonstrated earlier, significantly depends on the quantity of the penultimate stem vowel.[3] A first observation concerns the distribution of long penultimate stem vowels across theme vowel classes. As shown in Table 3.14, where only simplex verbs without root allomorphy are considered, long penultimate stem vowels are not equally distributed across theme vowel classes, with the classes that have an [i]-exponent of the present-tense theme vowel displaying the highest average numbers of member verbs involving long vowels.

Table 3.14: Distribution of long penultimate stem vowels in simplex BCMS verbs without root allomorphy

TV Class	N	Short PSV V	Long PSV V	% Long PSV V
a/i	19	8	11	57.89
i/i	324	157	167	47.72
e/i	50	30	20	40
a/je	50	35	15	30
a/a	187	126	61	27.54
a/e	5	5	0	0
e/e	3	3	0	0
Grand Total	679	405	274	40.35

Turning now to the prosodic prominence, Table 3.15 displays the distribution of the three prosodic patterns both for all verbs and just simplex verbs, categorized based on the length of the penultimate stem vowel.

The data in Table 3.15 indicate that verbs with a long PSV rarely maintain prominence on this vowel (i.e., they seldom belong to the PSV/PSV pattern) and by far most of them belong to the mobile TV/PSV pattern. Another noteworthy observation across vowel-length classes, which also holds of Slovenian, is that the TV/TV class exhibits a significantly higher percentage in counts based on simplex verbs compared to those based on all verbs.

Tables 3.16 and 3.17 show the distribution of the prosodic patterns by theme vowel class. Since previous discussion showed that short and long vowels in the

[3]Long vowels are not restricted to a single syllable and vowel length can arguably influence prominence in other positions as well. However, there is only one verb that has a long vowel on the syllable preceding the penultimate stem syllable: the verb *taložiti* 'sediment', which is why we excluded this verb from the dataset as an outlier.

Table 3.15: The distribution of three major prosodic types in BCMS verbs by the length of the penultimate stem vowel: Simplex vs. all verbs

	PSV	total N	PSV/PSV N (%)	TV/TV N (%)	TV/PSV N (%)
Simplex+Complex	short	2060	1186 (57.6%)	430 (20.9 %)	444 (21.5 %)
Simplex	short	356	173 (48.6%)	155 (43.5 %)	28 (7.9 %)
Simplex+Complex	long	2229	20 (0.9%)	151 (6.8 %)	2058 (92.3 %)
Simplex	long	273	10 (3.6%)	69 (25.3 %)	194 (71.1 %)

penultimate syllable of the stem constitute totally different phonological conditions, Table 3.16 presents the counts for verbs with a short vowel in the penultimate syllable of the stem, while Table 3.17 presents verbs with a long vowel in the penultimate syllable of the stem. The theme vowel classes are ordered by the percentage of verbs adhering to the most common prosodic type in the considered sample. For Table 3.16 this is PSV/PSV, whereas for Table 3.17 this is TV/PSV.

Tables 3.16 and 3.17 show that the BCMS theme vowel classes, just like Slovenian ones, behave quite distinctly from each other in terms of prosody. The gap observed in Slovenian, whereby present tense theme vowels with a consonantal component (classes a/je and u/e[4] in BCMS) cannot carry prominence, can also be observed in BCMS (with a single exception in the a/je class). A final important observation is the highly distinct behavior between the two largest classes i/i and a/a. While the former shows a three-way distinction in both tables, the latter allows only two patterns in each subsample: in the class of verbs with a short vowel in the penultimate syllable of the stem, there are no verbs belonging to the TV/PSV pattern, whereas in the class of verbs with a long vowel in the penultimate syllable of the stem there are no verbs belonging to the TV/TV pattern.

We conducted Pearson's chi-squared tests on the cross-tabulations of prosodic types and theme vowel classes. The results suggests a different behavior of theme classes regarding both the prominence patterns: χ^2 = 522.77, df = 20, p-value < 2.2e-16, and the quantity of the penultimate vowel of the stem: χ^2 = 29.465, df = 5, p-value = 1.879e-05.

Finally, we have also crossed the two prosodic variables, prominence and vowel length, as shown in Table 3.18. When both the infinitive and the present tense

[4]The BCMS theme u/e and the Slovenian i/e occur only in the context of a preceding segment n, often considered one morpheme: nu/ne, i.e. ni/ne. See Štarkl et al. (2024) for arguments that *n* is a separate morpheme in both Slovenian and BCMS.

Table 3.16: The distribution of three major prosodic types in BCMS verbs with short penultimate stem vowels by TV class: Simplex vs. all verbs

	TV Class	total N	PSV/PSV N (%)	TV/TV N(%)	TV/PSV N (%)
Simplex+Complex	u/e	173	138 (79.8%)	/	35 (20.2%)
Simplex	u/e	/	/	/	/
Simplex+Complex	a/je	331	262 (60.4%)	/	68 (39.6 %)
Simplex	a/je	28	18 (64.3%)	/	10 (35.7 %)
Simplex+Complex	a/a	548	331 (60.4%)	217 (39.6 %)	/
Simplex	a/a	126	67 (53.2%)	59 (46.8 %)	/
Simplex+Complex	i/i	828	429 (51.8%)	67 (8.1%)	332 (40.1%)
Simplex	i/i	157	84 (53.5%)	58 (36.9 %)	15 (9.5%)
Simplex+Complex	e/i	117	21 (17.9%)	89 (76.1%)	7 (6%)
Simplex	e/i	30	4 (13.3%)	24 (80 %)	2 (6.7%)
Simplex+Complex	a/i	38	5 (13.2%)	33 (86.8%)	/
Simplex	a/i	8	/	8 (100 %)	/
Simplex+Complex	a/e	8	/	8 (100%)	/
Simplex	a/e	4	/	4 (100 %)	/
Simplex+Complex	e/e	17	/	16 (94.1%)	1 (5.9%)
Simplex	e/e	3	/	2 (66.7%)	1 (33.3%)

show penultimate stem vowel prominence (PSV/PSV), the tendency is not to have long syllables (or vice versa, when the penultimate stem vowel syllable is short, it tends to carry prominence in both forms). Somewhat weaker, but in the same direction, is the tendency when both the infinitive and the present tense have prominence on the theme vowel (TV/TV prominence). Perhaps the strongest generalization is that the length of the penultimate vowel of the stem goes with a prominence mismatch between the infinitive and the present tense. The difference in distributions has been tested by Pearson's Chi-squared test, yielding the following results: χ^2 = 293.4, df = 4, p-value < 2.2e-16.

3.4 Conclusions

In this chapter, we presented the information pertaining to verbal inflection annotated in *WeSoSlaV* – in particular the theme vowel classes and the prosodic features in the infinitive and the present tense. Our considerations of the prosodic

Table 3.17: The distribution of three major prosodic types in BCMS verbs with long penultimate stem vowels by TV class: Simplex vs. all verbs

	TV Class	total N	PSV/PSV N (%)	TV/TV N(%)	TV/PSV N (%)
Simplex+Complex	u/e	82	/	/	82 (100%)
Simplex	u/e	/	/	/	/
Simplex+Complex	a/je	131	/	1 (0.8%)	130 (99.2%)
Simplex	a/je	14	/	/	14 (100 %)
Simplex+Complex	a/a	1154	13 (1.1%)	/	1141 (98.2%)
Simplex	a/a	61	8 (13.1%)	/	53 (86.9%)
Simplex+Complex	i/i	773	7 (0.9%)	62 (8%)	704 (91.1%)
Simplex	i/i	167	2 (1.2%)	38 (22.7%)	127 (76%)
Simplex+Complex	a/i	22	/	22 (100%)	/
Simplex	a/i	11	/	11 (100%)	/
Simplex+Complex	e/i	67	/	66 (98.5%)	1 (1.5%)
Simplex	e/i	20	0	20 (100%)	/
Simplex+Complex	a/e	/	/	/	/
Simplex	a/e	/	/	/	/
Simplex+Complex	e/e	/	/	/	/
Simplex	e/e	/	/	/	/

Table 3.18: The crossing of the prominence pattern and penultimate stem vowel length in BCMS simplex verbs

	no PSV length	PSV length
TV/TV prominence	149	69
TV/PSV prominence	34	195
PSV/PSV prominence	185	10

features in both Slovenian and BCMS confirmed the adequacy of approaching the prosody of different theme vowel classes separately, as there are not only different tendencies in different theme vowel classes, but also multiple cases of total incompatibility of certain theme vowel classes and prosodic patterns. One entirely novel finding (at least to our knowledge) is that BCMS theme vowel classes behave differently not only with respect to prominence, but also with respect to the length of the penultimate vowel of the vowel. This means that

both prominence and vowel length should be considered either as present in the underlying representation of theme vowels or as conditioned by their lexical representation.

While the results presented here are promising, more fine-grained analyses that will hopefully follow will bring further insight. Although our analysis was able to show overall statistically significant differences between theme vowel classes, a more detailed analysis might reveal further commonalities between theme vowel classes and suggest a simplification of the system. In this vein, it is worth investigating whether smaller theme classes can be modeled as underlying subclasses of the larger ones, where the different phonological realizations are related to special prosody. Some additional support for this view is presented in Chapter 5, where similar patterns are observed regarding the derivation of new items from verbs of various theme classes.

4 Properties of verb derivation

4.1 Introduction

In this chapter, we present the verbal derivational system in BCMS and Slovenian as annotated in *WeSoSlaV*. The chapter is organized in two main sections. Section 4.2 introduces examples of basic descriptive statistics in the two languages and how they are or can be employed to draw theoretical conclusions. Section 4.3 then zooms into the properties of WSS verbal prefixes as documented in *WeSoSlaV* and their theoretical potential. We conclude the chapter in Section 4.4.

The *WeSoSlaV* database annotates a large number of properties, subgroups of which can easily be used as smaller databases. For the purposes of this chapter we focus on the properties that cover the domain of verbal derivation thereby effectively creating a derivational sub-database of *WeSoSlaV*. The derivational sub-database contains information about the following properties: whether the verb has morphemes other than the root/base and the theme vowel and which suffixes these are, whether the verb has a prefix/a second prefix/a third prefix and which prefixes these are, and whether the verb contains a compound in the base. A detailed description of various subcategories of these types of morphemes can be found in Chapter 2. Our discussion of verbal derivation will also extend to its interaction with the properties of verbal inflection, mainly with theme vowels (see Chapter 3).

4.2 Basic statistics and how it can be informative

Table 4.1 provides an overview statistics of the distribution of verbs in BCMS across derivational categories annotated in the derivational subbase of *WeSoSlaV* in the column *Class*, i.e., Simplex verbs, Unsuffixed verbs, etc. While the annotation of these categories is described in detail in Chapter 2, Section 2.2, we here illustrate each category, with a brief explanation. Note also that some of the categories displayed in the table are not directly represented in *WeSoSlaV*, but are based on the combinations of the existing ones.

(1) Derivational categories (*Class*)

a. **Simplex verbs**: *čit-a-ti* [ROOT-TV-INF] 'read' – verbs consisting of just a root and an inflectional ending, i.e. without additional affixes such as prefixes or suffixes.
b. **Unsuffixed verbs**: *čit-a-ti* [ROOT-TV-INF] 'read', *pro-čit-a-ti* [PREF-ROOT-TV-INF] 'read' – verbs that do not contain any type of suffix or suffix-like element, but can contain a prefix.
c. **Prefixed verbs**: *pro-čit-a-ti* [PREF-ROOT-TV-INF] 'read', *u-čit-av-a-ti* [PREF-ROOT-SUFF-TV-INF] 'read in' – all verbs that contain a prefix, irrespective of the presence of other affixes.
d. **Prefixed unsuffixed verbs** – *pro-čit-a-ti* [PREF-ROOT-TV-INF] 'read' – verbs that contain a prefix, but not suffixes or suffix-like items, i.e. the intersection of **Unsuffixed verbs** and **Prefixed verbs**.
e. **Potentially prefixed verbs** – *za:tašk-a-ti* [PREF?-ROOT-TV-INF] – verbs that contain a prefix-like element whose status is not clear (i.e. it is potentially part of the base).
f. **Unprefixed verbs**: *čit-a-ti* [ROOT-TV-INF] 'read', *domin-ir-a-ti* [ROOT-SUFF-TV-INF] 'dominate' – all verbs without a prefix or a potential prefix, irrespective of whether they are simple or contain a suffix or suffix-like item.
g. **Verbs with a verbal suffix**: *u-čit-av-a-ti* [PREF-ROOT-SUFF-TV-INF] 'read in', *domin-ir-a-ti* [ROOT-SUFF-TV-INF] 'dominate' – all verbs that contain a verbal suffix, irrespective of all other affixes.
h. **Verbs with suffix-like items**: *zamišljati* < *za-misl-i-a-ti* 'imagine' – typically verbs that function as secondary imperfectives, but do not have a fully segmental suffix (yet they plausibly involve an underlying suffix realized as mora, etc.).
i. **Verbs with suffix-like items (without verbal or multifunctional affixes)** – the same as the previous one, but excluding verbs that additionally have typical verbal affixes or multifunctional affixes.
j. **Verbs with multifunctional affixes**: *krv-ar-i-ti* [ROOT-MULTI_AFF-TV-INF] 'bleed', *o-dom-**ać**-iv-a-ti* [PREF-ROOT–MULTI_AFF-SUFF-TV-INF] 'domesticate' – verbs that contain a suffix that is not clearly verbal, but appears readily across categories.
k. **Verbs with multifunctional affixes (without suffixed or suffixed-like elements)** – the same as the previous one, but without verbal suffixes or suffix-like elements.

l. **Compounds**: *blag-o-slov-i-ti* [ROOT-LINKING_VOWEL-ROOT-TV-INF] 'bless' – verbs that contain more than one lexical root.

The first two columns, with the values *All* and *%All*, specify the row and relative frequencies of each class, respectively. For instance, there are 2694 unsuffixed verbs in the BCMS sub-database, which is 50.83% of all 5,300 verbs in the sub-database. The column *Impf* specifies how many verbs in each class are imperfective, with two subsequent columns showing the relative shares for the class (*%Class*) and all verbs (*%All*). For example, looking at the third row, there are 702 imperfective verbs among unsuffixed verbs[1], which accounts for 26.06% of unsuffixed verbs (702/2694) and 13.25% of all the verbs in the sub-database (702/5300). The same holds for the column *Perf*, which specifies the number of perfective verbs, and the two subsequent columns (*%Class* and *%All*): e.g. among simplex verbs, there are 28 perfective verbs, which constitutes 4.12% of the simplex verbs and 0.53% of all the verbs in the sub-database.

Table 4.1: Counts and share of various types of verbs and verbal elements in the BCMS sub-database.

Class	All	%All	Impf	%Class	%All	Perf	%Class	%All
Simplex verbs	679	12.81	651	95.88	12.28	28	4.12	0.53
Unsuffixed verbs	2694	50.83	702	26.06	13.25	1992	73.94	37.58
Prefixed verbs	3659	69.04	1237	33.81	23.34	2422	66.19	45.70
Prefixed unsuffixed Vs	1986	37.47	38	1.91	0.72	1948	98.09	36.75
Potentially prefixed Vs	38	0.72	19	50.00	0.36	19	50.00	0.36
Unprefixed verbs	1603	30.24	1476	92.08	27.85	127	7.92	2.40
Vs with a verbal suffix	1753	33.08	1290	73.59	24.34	463	26.41	8.74
Vs with suffix-like items	966	18.23	935	96.79	17.64	31	3.32	0.58
Vs with suffix-like items (without verbal or multifunctional affixes)	718	13.55	687	95.68	12.96	31	4.32	0.58
Vs with multifunct. aff.	235	4.43	151	64.26	2.85	84	35.74	1.58
Vs with multifunct. aff. (without suffixed or suffix-like items)	132	2.49	50	37.88	0.94	82	62.12	1.55
Compounds	76	1.43	37	48.68	0.70	39	51.32	0.74

[1]Recall from Chapter 2 Section 2.5 that these are verbs that pass the tests with the present tense and phasal verbs, i.e. both traditional pure imperfective verbs and biaspectual ones.

The table shows, for instance, that simplex verbs account for 12.81% of the 5300 verbs in the sub-database, that roughly half of the verbs do not contain a suffix or a suffix-like element (Unsuffixed verbs, 50.83%), or that more than two-thirds of the verbs are prefixed (69.04%). It further makes it possible to quantitatively test some claims from the literature. For example, it is often claimed in the literature that simplex perfective verbs (such as *rešiti* 'solve') are a small closed class (see Milosavljević 2023b for a recent overview). *WeSoSlaV* shows that this is indeed the case, as such verbs account for 4.12% of simplex verbs and 0.53% of all verbs. Arsenijević (2023) uses this quantitative support to claim that these verbs are listed as bare *v*Ps with perfective/telic semantics. Further, it is usually claimed that prefixed verbs, in the absence of additional imperfectivizing suffixes or suffix-like elements, should be perfective, with some possible exceptions. Table 4.1 reveals that there are 38 such exceptions, which is 1.91% of all prefixed verbs without an additional suffix. Some relevant examples include the verbs *izgledati* 'look, appear' and *za-visiti* 'depend'. Although this class is indeed very small, it calls for further research on why there are any such verbs in the first place.

The situation is fairly similar in Slovenian, as shown by Table 4.2. Compared with 12.81% in BCMS, simplex verbs account for 14.97% of the 3,000 verbs in the Slovenian sub-database; compared with 50.83% for BCMS, unsuffixed verbs account for 63.17% of all verbs in the Slovenian sub-database; compared with the 69.04% share of prefixed verbs in BCMS, prefixed verbs have a 75.33% share in the Slovenian sub-database; compared with the simplex perfectives representing 4.12% of simplex verbs and 0.53% of all verbs in BCMS, simplex perfectives represent 7.35% of simplex verbs and 1.10% of all verbs in Slovenian.

Although the BCMS and Slovenian shares tend to be relatively similar, some results may start to look different when put in the context of specific claims in the literature. One such case is prefixed unsuffixed imperfective verbs: compared with the 1.81% share of all prefixed unsuffixed imperfectives in BCMS, this share, while still relatively small, goes up to 7.07% in Slovenian. The assumption in the literature, however, is usually the same for BCMS and Slovenian, i.e., that in the absence of additional imperfectivizing suffixes or suffix-like elements, prefixed verbs are perfective, and deviations from this can be treated as exceptions. But unlike the 1.91% share from BCMS, the 7.07% share (101 of 1,429) may no longer support treating such verbs simply as exceptions. Note, though, that Dickey (2003) (but see also Marušič & Žaucer 2006: 132) actually does argue that the above generalization may be weaker in Slovenian.

Table 4.2: Counts and share of various types of verbs and verbal elements in the Slovenian sub-database.

Class	All	%All	Impf	%Class	%All	Perf	%Class	%All
Simplex verbs	449	14.97	415	92.43	13.83	33	7.35	1.10
Unsuffixed verbs	1895	63.17	526	27.76	17.53	1368	72.19	45.60
Prefixed verbs	2260	75.33	715	31.64	23.83	1545	68.36	51.50
Prefixed unsuffixed Vs	1429	47.63	101	7.07	3.37	1392	93.11	46.40
Potentially prefixed Vs	39	1.30	24	61.54	0.80	15	38.46	0.50
Unprefixed verbs	740	24.67	674	91.08	22.47	65	8.78	2.17
Vs with a verbal suffix	638	21.27	462	72.41	15.40	176	27.59	5.87
Vs with suffix-like items	367	12.23	366	99.73	12.20	1	0.27	0.03
Vs with suffix-like items (without verbal or multifunctional affixes)	358	11.93	357	99.72	11.90	1	0.28	0.03
Vs with multifunct. aff.	153	5.10	87	56.86	2.90	66	43.14	2.20
Vs with multifunct. aff. (without suffixed or suffix-like items)	104	3.47	39	37.50	1.30	65	62.50	2.17
Compounds	27	0.90	9	33.33	0.30	18	66.67	0.60

Tables 4.3 to 4.6 represent an example of the possibility of looking for an interaction between derivation and inflection. Specifically, they show how different derivational classes are distributed across theme vowel classes in BCMS. The following derivational categories are included: all verbs (*All*), unsuffixed verbs (*-S*), simplex verbs (*Simplex*), prefixed unsuffixed verbs (*P-S*), verbs with a verbal suffix (*VS*), verbs with suffix-like elements, without verbal or multifunctional items (*S-like**), verbs with multifunctional affixes but without verbal or suffix-like affixes (*Multi**), and compounded verbs (*Comp*).

From Table 4.3, it can be observed that the TV a/a is the most frequent one overall, but the i/i theme vowel is substantially more frequent than a/a in simplex and prefixed verbs without a suffix. By contrast, as can be seen from Table 4.4, verbs with verbal suffixes or suffix-like elements favor the TV a/a alongside the TV a/je, whereas i/i is not attested among verbs with a verbal suffix. i/i is a dominant TV also in verbs with multifunctional affixes and compounds. Formal analyses arising from this and more specific quantitative data are offered in Arsenijević & Milosavljević (2021), Milosavljević & Arsenijević (2022), Simonović et al. (2023), Simonović & Mišmaš (2023a), Arsenijević et al. (2025). Further, TVs u/e and ∅/ne

are not attested in derivational categories other than verbs with a derivational suffix (they are annotated in the sub-database as having the suffix *-n*). This has led Štarkl et al. (2024) to propose an analysis according to which these two theme classes actually contain the TV ∅/e and the diminutive suffix *-n*.

Table 4.3: TVs across different derivational categories in BCMS (1)

TV	All	%All	-S	%-S	Simplex	%Simplex	P-S	%P-S
a/a	1702	32.11	524	19.45	187	27.54	331	16.66
i/i	1601	30.21	1430	53.08	324	47.72	1093	55.01
a/je	1029	19.42	171	6.35	50	7.36	121	6.09
∅/e	298	5.62	298	11.06	36	5.30	256	12.88
u/e	258	4.87	0	0	0	0	0	0
e/i	184	3.47	176	6.53	50	7.36	124	6.24
∅/ne	124	2.34	0	0	0	0	0	0
a/i	62	1.17	53	1.97	19	2.80	34	1.71
a/e	19	0.36	19	0.71	5	0.74	14	0.70
e/e	17	0.32	17	0.63	3	0.44	14	0.70
def.	6	0.11	6	0.22	5	0.74	0	0
Total	5300	100	2694	100	679	100	1987	100

Table 4.4: TVs across different derivational categories in BCMS (2)

TV	VS	%VS	S-like*	%S-like	Multi*	%Multi*	Comp	%Comp
a/a	703	40.10	441	61.42	34	25.76	13	16.88
i/i	0	0	81	11.28	90	68.18	44	57.14
a/je	661	37.71	192	26.74	5	3.79	17	22.08
∅/e	0	0	0	0	0	0	0	0
u/e	258	14.72	0	0	0	0	0	0
e/i	5	0.29	0	0	3	2.27	0	0
∅/ne	124	7.07	0	0	0	0	2	2.60
a/i	2	0.11	4	0.56	0	0	0	0
a/e	0	0	0	0	0	0	0	0
e/e	0	0	0	0	0	0	0	0
def.	0	0	0	0	0	0	1	1.30
Total	1753	100	718	100	132	100	77	100

For Slovenian, Table 4.5 shows that just like in BCMS, the TV a/a is the most frequent one overall and i/i is substantially more frequent than a/a in simplex and prefixed verbs without a suffix. Also in the same spirit as in BCMS, Table 4.6 shows that Slovenian verbs with verbal suffixes or suffix-like elements favor the TV a/a alongside the TV a/je, whereas i/i is essentially unattested (1 occurrence, 0.16%) among verbs with a verbal suffix. Also quite like in BCMS, i/i is, conversely, the dominant TV in compound verbs, but much unlike in BCMS, the dominance of a/a over i/i, as well as a bigger share of a/je, extends also to verbs with multifunctional affixes.[2]

Further, the Slovenian traditional TV ni/ne comes very close to the BCMS TVs u/e and ∅/ne in terms of their being unattested in derivational categories other than derivationally suffixed verbs, which may, in the spirit of Štarkl et al. (2024), perhaps suggest a reanalysis of this theme class as a composite of a TV and a diminutive *-n*.

Table 4.5: TVs across different derivational categories in Slovenian (1)

TV	All	%All	-S	%-S	Simplex	%Simplex	P-S	%P-S
a/a	1044	34.80	419	22.11	127	28.29	284	19.85
i/i	863	28.77	801	42.27	183	40.76	612	42.77
a/je	378	12.60	111	5.86	23	5.12	88	6.15
∅/e	285	9.50	285	15.04	40	8.91	244	17.05
e/i	128	4.27	121	6.39	41	9.13	80	5.59
e/e	47	1.57	47	2.48	10	2.23	36	2.52
a/e	46	1.53	46	2.43	6	1.34	40	2.80
a/i	36	1.20	35	1.85	15	3.34	21	1.47
∅/ne	27	0.90	27	1.42	1	0.22	26	1.82
def.	3	0.10	3	0.16	3	0.67	0	0
Total	3000		1895	100	449	100	1431	100

4.3 Quantitative insights into verbal prefixes

In this section, we outline the properties of prefixes as annotated in *WeSoSlaV*. The section is organized into three main parts. In Section 4.3.1, we present the fre-

[2]The observed situation with multifunctional affixes calls for further investigation, as these affixes appear to form a highly heterogeneous category, lacking clear-cut tests and potentially subject to slightly different criteria depending on the language. See Chapter 6 for a detailed discussion.

Table 4.6: TVs across different derivational categories in Slovenian (2)

TV	VS	%VS	S-like*	%S-like*	Multi*	%Multi*	Comp	%Comp
a/a	226	35.42	360	98.09	63	41.18	5	18.52
i/i	1	0.16	2	0.54	59	38.56	21	77.78
a/je	266	41.69	5	1.36	25	16.34	1	3.70
∅/e	0	0	0	0	0	0	0	0
e/i	3	0.47	0	0	4	2.61	0	0
e/e	0	0	0	0	0	0	0	0
a/e	0	0	0	0	0	0	0	0
a/i	1	0.16	0	0	0	0	0	0
∅/ne	0	0	0	0	0	0	0	0
def.	0	0	0	0	0	0	0	0
Total	638	100	367	100	153	100	27	100

quency of root-adjacent prefixes, Section 4.3.2 focuses on prefix stacking, while the interaction of prefixes with other affixes, i.e. theme vowels and suffixes is a topic of Section 4.3.3.

4.3.1 Frequency of root-adjacent prefixes

Table 4.7 illustrates the distribution of root-adjacent prefixes in all prefixed verbs and in unsuffixed verbs in BCMS (the distributions across other categories will be provided in the subsequent section).[3] It shows that the prefixes *po-*, *u-* and *za-* are the most frequent prefixes both when all prefixed verbs are included and when only unsuffixed prefixed verbs are taken into account. The bottom three prefixes from Table 4.7, *vaz-*, *mimo-* and *suprot-*, are not productive in contemporary BCMS. Slovenian is again similar: as seen in Table 4.8, the most frequent prefixes both overall as well as in the subset of unsuffixed verbs are *po-* and *za-*. The prefix *u-* scores a little lower in frequency in Slovenian than in BCMS, though it should be kept in mind that the BCMS *u-* is actually the counterpart of both Slovenian *u-* and *v-*, and the combined count of these two prefixes would end up in third place overall, just behind *po-* and *za-*. Similarly, *s-/z-/se-* score higher than *s-* in BCMS, but a combined count for *s-* and *sa-* in BCMS comes close to the

[3] Prefixed verbs whose root-adjacent prefixes are potential prefixes, i.e. marked with a colon, are excluded from the calculation.

relative frequency of Slovenian *s-/z-/se-*.[4] A noticeable feature particularly of the BCMS table, and with the exception of *po-* also the Slovenian table, is the continuous, steady cline in frequency from the most to the least frequent prefixes without any big gaps.

Can this data be used to shed light on any theoretical claims from the previous literature? Dickey (2012) argues that in Slovenian, unlike in BCMS, there is a grammaticalized prefix that functions as a pure perfectivizer: *s-/z-*. The continuous decline in frequency of prefixes can be taken as quantitative evidence against singling out any of the prefixes in BCMS or Slovenian as a purely grammatical perfectivizing prefix. In Slovenian, *po-* sticks out slightly, but surely not with a percentage that would suggest the status of a grammaticalized perfective marker, and Dickey actually identifies *s-/z-* as the Slovenian partially grammaticalized empty perfectivizer, not *po-*.

4.3.2 Prefix stacking

Multiple prefixation in Slavic refers to the possibility of verbal prefixes stacking on each other, as in *na-do-dati* (on-to-give) 'add a lot of' from BCMS. In generative approaches, prefix stacking has been used as one of the main diagnostics for the difference between lexical and superlexical prefixes, in that it is usually claimed that only superlexical prefixes can be stacked (cf. Svenonius 2004b, Milićević 2004, Istratkova 2004, Markova 2011, among many others). This diagnostic has been questioned in the literature, though, as some lexical-like prefixes can be stacked too (cf. Žaucer 2009, 2013, Marušič et al. 2025 for Slovenian, Milosavljević 2023b for BCMS, Zinova 2021 for Russian). In this section, we present quantitative data from BCMS and Slovenian and discuss how it can be informative regarding the status of stacked prefixes. We start with an overview of basic quantitative patterns.

In the BCMS sub-database, out of 3635 prefixed verbs, 271 are attested with two prefixes (7.46% of all prefixed verbs) and 10 with three prefixes (0.28% of all prefixed verbs); and very similarly in the Slovenian sub-database, out of 2261 prefixed verbs, 178 are attested with two prefixes (7.87% of all prefixed verbs) and 6 with three prefixes (0.27% of all prefixed verbs). Among the 271 BCMS verbs with

[4] Prefixes *sa-* and *s-*, and to some extent *su-* as well, can be plausibly analyzed as a single prefix with three allomorphic realizations in at least some of their uses. However, for the purposes of the database, we counted the three items as separate prefixes because there are cases where these items give rise to different meanings when combined with the same verb (e.g. *sa-plesti* 'stumble' vs. *s-plesti* 'plait') and/or have a different distribution that is not phonologically conditioned, e.g. *su-delovati* (∗sa-delovati) 'participate', or *sa-stati* (∗su-stati) 'meet', etc.

Table 4.7: Distribution of prefixes in BCMS

Prefix	All P-Verbs	%All P-Verbs	Unsuffixed P-Verbs	%Unsuffixed P-Verbs
po	437	12.02	239	12.12
u	424	11.66	229	11.61
za	350	9.63	215	10.90
o	296	8.14	149	7.56
iz	280	7.70	149	7.56
na	245	6.74	145	7.35
s	240	6.60	127	6.44
pro	206	5.67	121	6.14
od	199	5.47	104	5.27
pre	199	5.47	108	5.48
raz	172	4.73	94	4.77
pri	149	4.10	80	4.06
do	116	3.19	62	3.14
ob	78	2.15	38	1.93
sa	73	2.01	41	2.08
pod	65	1.79	25	1.27
uz	42	1.16	18	0.91
nad	20	0.55	10	0.51
su	18	0.50	4	0.20
pred	17	0.47	10	0.51
vaz	5	0.14	2	0.10
mimo	2	0.06	1	0.05
suprot	2	0.06	1	0.05
Total	3635	100.00	1972	100.00

two prefixes, there are 77 unique combinations of two prefixes, and among the 178 Slovenian verbs with two prefixes, there are 68 unique combinations of two prefixes. For both BCMS and Slovenian, this is 18% of all theoretically possible combinations, assuming an idealized scenario where each of the 21 prefixes can combine with any other prefix in any of the two directions (ignoring only the possibility of a prefix combining with itself, i.e., of a root hosting two instances of the same prefix).[5]

[5]The BCMS sub-database actually includes 23 prefixes, but we have considered the prefixes *s-*, *sa-*, *su-* as one prefix here (see fn. 4). The BCMS number would be even higher if the combinations

Table 4.8: Distribution of prefixes in Slovenian

Prefix	All P-Verbs	%All P-Verbs	Unsuffixed P-Verbs	%Unsuffixed P-Verbs
po	350	15.48	224	15.65
za	244	10.79	181	12.65
s/z/se	207	9.16	133	9.29
pre	195	8.62	123	8.60
o	171	7.56	107	7.48
u	149	6.59	77	5.38
iz	144	6.37	92	6.43
pri	139	6.15	92	6.43
na	132	5.84	84	5.87
od	118	5.22	78	5.45
raz	103	4.56	61	4.26
ob	75	3.32	45	3.14
v	71	3.14	43	3.00
do	66	2.92	41	2.87
vz	33	1.46	16	1.12
pod	25	1.11	15	1.05
pro	16	0.71	9	0.63
pred	10	0.44	6	0.42
so	7	0.31	2	0.14
nad	6	0.27	2	0.14
Total	2261	100,00	1431	100,00

There are two basic possibilities when it comes to the verbs with multiple prefixes. The first one is that a prefixed verb is imperfectivized by a secondary imperfective suffix (or suffix-like element), and then the stacked prefix comes on top to make such a verb perfective again. This case is illustrated by BCMS *iz-pre-pis-iva-ti* 'to rewrite all'; here the verb *pisati* 'to write' is perfectivized by the prefix *pre-*, the entire verb is imperfectivized by the suffix *iva*, and, finally, the verb is made perfective again by the prefix *iz-*. The second case includes verbs that consist of a root, a verbalizer, and multiple prefixes on top of that, without an intermediate imperfectivization stage (unless there is one, but triggered

of unproductive prefixes such as *mimo-*, *suprot-* or *vaz-* were not taken into account. In more exhaustive lists, for instance, such as the one provided by Stojanović (2016) based on different traditional dictionaries of Serbo-Croatian, the number of attested combinations reaches 46%.

by a null phonological suffix; for the discussion of this option, see Žaucer 2009, Milosavljević 2023b). This case is illustrated by BCMS *na-do-dati*, where two prefixes (*na-* 'on' and *do-* 'to') are stacked on top of the verb *dati* 'to give', with the resulting meaning 'to add a lot of'. This derivational history has not been directly recorded in the *WeSoSlaV* database. There are, however, some indirect clues to track different options. For instance, of all 271 stacked examples in the BCMS sub-database, 125 are unsuffixed (that is, do not include verbal suffixes or suffix-like items, such as single vowels with imperfectivizing effect; recall 2.3.4.3), and 150 do not include suffixes other than ones traditionally classified as verbalizers (e.g. *ova*, *ira*, *isa*). These 150 verbs therefore include prefixes that are directly stacked on one another, without secondary imperfectivization. Since in both options the first prefix is closer to the root morphologically (and presumably also syntactically), we provide summary statistics for all verbs with two prefixes together. Tables 4.9 and 4.11 summarize the distribution of all prefixes in BCMS and Slovenian *WeSoSlaV* ("1.prefix" is the one closest to the verb root, "3.prefix" the one furthest from the verb root), while Tables 4.10 and 4.12 illustrate the ten most frequent combinations of two stacked prefixes (out of the total of 77 and 68 different combinations for BCMS and Slovenian, respectively).

How can the presented data on stacked prefixes be theoretically informative? Tables 4.9 and 4.11 show that, without distinguishing between specific uses of individual prefixes, almost all prefixes that occur as the first prefix can also be found as the second one, i.e., all except the archaic *vaz-* in BCMS and *pod-* in Slovenian. This includes prefixes that are typically considered lexical or that are absent from lists of superlexical prefixes but apparently also occur as the second, i.e., stacked prefix: *u-* 'in', *s-* 'from', *o-* 'about', *ob-* 'around', *raz-* 'apart', *uz-*/*vz-* 'up', and in BCMS also *pod-* 'under'. Assuming that one of the typical properties of lexical prefixation is, as explained above, that there can normally only be one lexical prefix per verb, the fact that essentially all prefixes occur both as the first and as the second prefix in a pair of two prefixes suggests that no prefix can be claimed to be exclusively lexical.

Working with a slightly larger database of Slovenian prefixed verbs, Marušič et al. (2025) build a similar argument. They conclude that at least in its simplest understanding, the starting assumption (whereby there can be only one lexical prefix per verb) is probably incorrect, since quite some uses of stacked prefixes otherwise exhibit properties of typical lexical prefixes. Furthermore, a look at the aspectual status of verbs with stacked prefixes in *WeSoSLaV* shows that in BCMS, for example, as many as 114 out of 271 (42%) are imperfective, suggesting that 114 such verbs have presumably undergone secondary imperfectivization after taking the second prefix, which additionally speaks against their superlexical-like

Table 4.9: Distribution of stacked prefixes in BCMS

Prefix	1.prefix	%(/271)	2.prefix	%(/271)	3.prefix	%(/10)
po	82	30.26	20	7.38	1	10
u	33	12.18	35	12.92	0	0
s	28	10.33	23	8.49	3	30
o	24	8.86	17	6.27	1	10
do	23	8.49	4	1.48	0	0
na	14	5.17	16	5.90	0	0
iz	12	4.43	23	8.49	0	0
raz	9	3.32	20	7.38	0	0
pro	8	2.95	13	4.80	0	0
pre	6	2.21	25	9.23	5	50
za	6	2.21	26	9.59	0	0
pri	5	1.85	11	4.06	0	0
ob	4	1.48	7	2.58	0	0
sa	4	1.48	2	0.74	0	0
uz	4	1.48	4	1.48	0	0
pod	3	1.11	12	4.43	0	0
od	2	0.74	8	2.95	0	0
su	2	0.74	1	0.37	0	0
vaz	2	0.74	0	0.00	0	0
mimo	0	0.00	0	0.00	0	0
nad	0	0.00	0	0.00	0	0
pred	0	0.00	4	1.48	0	0
suprot	0	0.00	0	0.00	0	0
Total	271	100.00	271	100.00	10	100

Table 4.10: Most frequent combination of the first and second prefix in BCMS (in linear order)

Combination	Count
na+do	14
u+po	13
raz+po	11
za+po	10
za+o	9
pro+iz	9
iz+po	8
s+po	8
u+s	8
o+po	7

status. In the spirit of Marušič et al. (2025), as well as a number of earlier sources such as Tatevosov (2008), Markova (2011) or Žaucer (2013), the stacking situation revealed by a search of *WeSoSLaV* suggests that a simple split into lexical and superlexical prefixes cannot do justice to the complexity of Slavic prefixation.

4.3.3 Prefixes and other affixes

In this section, we discuss quantitative patterns of the interaction between prefixes and other affixes and their theoretical consequences. This includes theme vowels (Section 4.3.3.1) and verbal suffixes (Section 4.3.3.2). Section 4.3.3.3 then discusses affixes at the intersection of the previous two types – verbalizers, i.e. morphemes that (are usually argued to) combine with a base to derive a verb. They include theme vowels and a subset of verbal suffixes.

4.3.3.1 Prefixes and theme vowels

While the interaction between theme vowels with simple verbs and suffixes has received abundant attention in recent years (see e.g. Svenonius 2004a, Gribanova 2015, Quaglia et al. 2022, Matushansky 2024, Arsenijević et al. 2025), including the works based on *WeSoSlaV* (e.g. Arsenijević & Milosavljević 2021, Kovačević et al. 2024, Milosavljević & Arsenijević 2022, Simonović & Mišmaš 2023a), there has not been much focus on the interplay between prefixes and theme vowels. Arsenijević & Milosavljević (2021) note in passing that in BCMS, when pairs of

Table 4.11: Distribution of stacked prefixes in Slovenian

Prefix	1.prefix	%(/178)	2.prefix	%(/178)	3.prefix	%(/6)
po	53	29.78	26	14.61	0	0
za	6	3.37	13	7.30	0	0
s/z/se	10	5.62	33	18.54	5	83.33
pre	14	7.87	17	9.55	1	16.67
o	10	5.62	12	6.74	0	0
u	9	5.06	7	3.93	0	0
iz	7	3.93	16	8.99	0	0
pri	7	3.93	14	7.87	0	0
na	6	3.37	11	6.18	0	0
od	0	0	5	2.81	0	0
raz	4	2.25	11	6.18	0	0
ob	5	2.81	1	0.56	0	0
v	12	6.74	1	0.56	0	0
do	7	3.93	2	1.12	0	0
vz	14	7.87	4	2.25	0	0
pod	10	5.62	0	0	0	0
pro	4	2.25	2	1.12	0	0
pred	0	0	1	0.56	0	0
so	0	0	2	1.12	0	0
nad	0	0	0	0	0	0
Total	178	100	178	100	6	100

verbs that share the same root but take a different TV (i/i vs. a/a) are prefixed, the prefixes with the TV i/i tend to receive a lexical/spatial interpretation, whereas those with the TV a/a are usually superlexical/quantificational. In a similar spirit, Milosavljević et al. (2025), based on a quantitative study of the prefixation of verbs in BCMS that share the root but differ in TV (i/i vs. *ova/uje*), found that *i*-verbs are significantly more prone to be prefixed than their *ova*-counterparts (e.g., while the verb *mir-i-ti* 'to reconcile' is attested with 10 different prefixes, *mir-ov-a-ti* 'to rest' is attested with only one in their corpus study). These two studies have taken the given results to indicate that the prefix-verb combinatorics also depends on the TV class, i.e. that both the distribution of prefixes and their interpretation may be affected by the TV class. Arsenijević & Milosavljević (2021) assume that prefixes are merged higher than TVs and that their (lexical vs. superlexical) interpretation is constrained by the material inside the *v*P (scalar vs.

Table 4.12: Most frequent combination of the first and second prefix in Slovenian (in linear order)

Combination	Count
s+po	13
s+pre	12
iz+po	7
na+v	6
o+po	5
s+pod	5
po+s	5
po+na	4
pre+po	4
pri+po	4
raz+po	4
u+po	4
za+u	4
po+vz	4

non-scalar nature of i/i vs. a/a), and Milosavljević et al. (2025) even go a step further, and take the differences in the distribution and interpretation of prefixes relative to a TV class as evidence that prefixes are generated higher than TVs. While we are not aware of clear cases in which the TV class changes depending on the prefix, thus potentially indicating the opposite direction of influence (i.e. that prefixes influence the choice and/or interpretation of the TV class), this possibility should be explored in more detail in future work. Independently of the correct answer about the derivation history (prefixes > TVs or TVs > prefixes), the fact that there are significant correlations between the two classes of morphemes indicates, at the very least, that TVs should not be treated as purely ornamental morphemes. If they were, one would not expect that the distribution or interpretation of prefixes varies in any significant way across TV classes.

While the above-mentioned studies focused on prefixed minimal pairs of verbs that share the root but differ in a TV, *WeSoSlaV* makes it possible to explore the interaction between the two types of morphemes on more general quantitative grounds – by comparing the shares of TVs across simple and prefixed verbs and looking for significant correlations. Since the lexical vs. superlexical prefix-use distinction, or more generally different uses of individual prefixes, are not annotated in *WeSoSlaV*, predictions revolving around the interpretation of prefixes

depending on the TV class cannot be directly tested. However, a qualitative analysis of this type stands out as the next plausible step based on the quantitative data that suggest significant correlations regarding the representation of different TVs in simple vs. prefixed verbs.

Tables 4.13 and 4.14 summarize the number of simplex vs. prefixed verbs in BCMS and Slovenian across different TV classes. As various suffixes could also be analyzed as sequences of theme vowels (cf. Quaglia et al. 2022, Simonović et al. 2023, Simonović & Mišmaš 2023a, Arsenijević et al. 2025), including verbs with suffixes into this table could skew the comparison.

Table 4.13: Simple vs. prefixed verbs across TV classes in BCMS

TV	Simple	%Simple	%All	Prefixed	%Prefixed	%All
∅/e	36	5.30	0.68	256	12.88	4.83
a/a	187	27.54	3.53	331	16.66	6.25
a/e	5	0.74	0.09	14	0.70	0.26
a/i	19	2.80	0.36	34	1.71	0.64
a/je	50	7.36	0.94	121	6.09	2.28
e/e	3	0.44	0.06	14	0.70	0.26
e/i	50	7.36	0.94	124	6.24	2.34
i/i	324	47.72	6.11	1093	55.01	20.62
Total	679	100	12.81	1987	100	37.49

Table 4.14: Simple vs. prefixed verbs across TV classes in Slovenian

TV	Simple	%Simple	%All	Prefixed	%Prefixed	%All
a/a	135	29.28	4.50	284	19.85	9.47
i/i	189	41.00	6.30	612	42.77	20.40
a/je	23	4.99	0.77	88	6.15	2.93
∅/e	41	8.89	1.37	244	17.05	8.13
e/i	41	8.89	1.37	80	5.59	2.67
e/e	11	2.39	0.37	36	2.52	1.20
a/e	6	1.30	0.20	40	2.80	1.33
a/i	14	3.04	0.47	21	1.47	0.70
∅/ne	1	0.22	0.03	26	1.82	0.87
Total	461	100	15.37	1431	100	47.70

Looking at the two largest TV classes in BCMS, i/i and a/a, it can be observed from Table 4.13 that the share of i/i increases in prefixed verbs compared to simple ones (47.72% simple Vs vs. 55.01% prefixed Vs), while at the same time, the share of a/a verbs decreases (27.54% simple Vs vs. 16.66% prefixed Vs), indicating that the TV i/i is more prone to be found with prefixed verbs in comparison to the TV a/a. This is in line with the observation that i/i significantly correlates with perfective, telic or scalar verbal bases in comparison to the TV a/a also in non-prefixed verbs (Arsenijević & Milosavljević 2021, Milosavljević & Arsenijević 2022). For instance, of the 28 perfective simplex verbs, 26 (93%) are with the theme vowel i/i, 1 (3.5%) with the theme vowel a/a[6] and 1 (3.5%) with the theme vowel ∅/e. The situation is roughly similar in Slovenian (Table 4.14): here, too, the share of the TV a/a significantly decreases in prefixed verbs compared to simplex ones (29.28% simple Vs vs. 19.85% prefixed Vs), and the share of i/i is slightly higher in prefixed verbs than in simplex verbs, albeit by a relatively smaller margin than in BCMS (41% simple Vs vs. 42.77% prefixed Vs). The data from Slovenian, for which the correlation of i/i with perfective, telic or scalar verbal bases compared to a/a and non-prefixed verbs has not been investigated, thus invite a cross-linguistic verification of the BCMS finding of Arsenijević & Milosavljević (2021) and Milosavljević & Arsenijević (2022). We leave the investigation of the correlations between theme vowels and perfectivity, telicity, and scalarity in Slovenian for future research.

Another significant correlation between TV classes and prefixes in both BCMS and Slovenian obtains for the class ∅/e. In both sub-databases, there are significantly more prefixed verbs with this TV in comparison to simplex verbs: in BCMS, it accounts for 12.88% of prefixed verbs and 5.3% of simplex verbs, whereas in Slovenian, the corresponding class comprises 17.05% of prefixed verbs compared to 8.89% of simplex verbs. If we reanalyze the TVs u/e in BCMS (traditional nu/ne) and i/e in Slovenian (traditional ni/ne) as ∅/e that combine with the diminutive suffix *-n* (with Štarkl et al. 2024) and decompose the TV ∅/ne into ∅/e and the suffix *-n*, it turns out that the TV ∅/e strongly correlates with perfectivity/telicity in both languages. That is, not only does it favor prefixed verbs, it is also the only theme vowel that combines with the suffix *-n*, which in contemporary BCMS and Slovenian derives only perfective verbs (Štarkl et al. 2024).

Given these differences in the shares of prefixed verbs across TV classes, it is interesting to check whether the distribution of specific prefixes differs across

[6] As discussed in Milosavljević & Arsenijević (2022), the only such verb is *d-a-ti* 'to give', *d-a-mo* 'we give', which is also attested with the more archaic root *dad*. This might indicate that *a* in this verb is actually part of the root, which, in turn, would mean that the TV a/a is not attested with simple perfective verbs.

TV classes. Tables 4.15 and 4.16 provide the distributions of all annotated prefixes across the three TV classes that are most frequent among prefixed verbs. Indeed, in both languages, there is some variation between the top prefixes depending on the TV class. In BCMS, for instance, prefixes *u-* and *po-* are the most frequent ones in all classes, and also when all verbs with all verbalizers are taken into account (i.e. all except secondary imperfectivizing and other suffixes except for verbalizing ones). However, the difference is in that the prefix *po-* is slightly more frequent than *u-* overall, i.e. when all the verbalizers are taken into account, but the suffix *u-* is more frequent in the three observed TV classes. It remains a task for future qualitative research to determine whether these prefixes tend to receive different interpretations when combined with different TV classes (e.g. lexical vs. superlexical). Just like in BCMS, in Slovenian too, two prefixes, *po-* and *za-*, are the most frequent prefixes in all TV classes. In this language as well, the distribution of the top two prefixes varies a bit across TV classes. While the prefix *po-* is almost twice as frequent as the prefix *za-* in the TV class i/i, the two prefixes are very closely distributed across the class a/a (16.55% vs. 14.08% in favor of *po-*) and equally distributed in the class ∅/e (13.11% each). Again, qualitative analysis in future work will be necessary to fully evaluate the correlations between specific prefixes (and their different interpretations) and TV classes.

4.3.3.2 Prefixes and suffixes

We can now move to the discussion of the interaction between prefixes and suffixes (for BCMS, we limit our attention only to suffixes attested in at least 50 verbs in the sub-database). Table 4.17 shows the share of such suffixes across prefixed (P) and unprefixed (-P) verbs in BCMS (note that here unprefixed verbs are those that contain neither a prefix nor a potential prefix; VS stands for a verbal suffix).

The suffixes in Table 4.17 fall into roughly three categories regarding their combinatorics with prefixes. The first group consists of two suffixes – *av+a, iv+a* – which almost exclusively appear in prefixed verbs (94.57% and 98.47%, respectively). This is because these suffixes serve to imperfectivize perfective verbs, which are in turn typically made perfective by prefixation (except for a small number of simple perfective verbs). On the opposite pole, suffixes traditionally regarded as purely verbalizing in BCMS – *ir+a*, *is+a* and *ov-a* – are significantly less often found in prefixed verbs (8.29%, 12% and 22.67%, respectively). Similarly in Slovenian (Table 4.18): there is a suffix that almost exclusively appears in prefixed verbs, namely, *av+a* (96.15%). Two further prefixes very frequently appear in prefixed verbs, *ov/ev+a* (76.32%) and *ev+a* (76.47%); the fact that these two have

Table 4.15: Prefixes across most frequent prefixed TV classses (only unsuffixed verbs) in BCMS

Prefix	i/i		a/a		∅/e		All_V-izers	
	Freq	%	Freq	%	Freq	%	Freq	%
u	158	13.81	42	11.86	29	11.74	277	11.43
po	145	12.67	41	11.58	25	10.12	296	12.21
o	118	10.31	25	7.06	4	1.62	195	8.04
za	112	9.79	36	10.17	23	9.31	253	10.44
iz	85	7.43	27	7.63	20	8.10	185	7.63
na	80	6.99	25	7.06	19	7.69	166	6.85
s	74	6.47	28	7.91	16	6.48	165	6.81
pro	57	4.98	33	9.32	18	7.29	149	6.15
raz	56	4.90	13	3.67	12	4.86	110	4.54
pri	55	4.81	4	1.13	9	3.64	101	4.17
pre	51	4.46	22	6.21	15	6.07	133	5.49
od	50	4.37	19	5.37	16	6.48	132	5.45
do	27	2.36	13	3.67	14	5.67	74	3.05
ob	26	2.27	5	1.41	6	2.43	44	1.82
pod	14	1.22	2	0.56	6	2.43	31	1.28
sa	12	1.05	8	2.26	11	4.45	46	1.90
uz	9	0.79	4	1.13	1	0.40	27	1.11
pred	6	0.52	1	0.28	0	0	10	0.41
nad	4	0.35	5	1.41	1	0.40	12	0.50
su	4	0.35	0	0	0	0	13	0.54
suprot	1	0.09	0	0	0	0	1	0.04
mimo	0	0.00	0	0	1	0.40	1	0.04
vaz	0	0.00	1	0.28	1	0.40	3	0.12
Total	1144	100.00	354	100.00	247	100.00	2424	100.00

lower numbers than *av+a*, and than the BCMS *iv+a*, presumably has to do with the fact that these suffixes serve not just as secondary imperfectivizers but also as verbalizers for denominal verbs (similarly to BCMS *ov+a*). The Slovenian suffix *ir+a* appears in prefixed verbs infrequently, similarly to its BCMS cognate. Another class of suffixes that combines fairly frequently with prefixes in both BCMS and Slovenian is the perfectivizing suffix *nu* in BCMS and its counterpart *ni* in Slovenian, with 71.98% and 80.99%, respectively.

Finally, the suffix *-ka*, which can be plausibly analyzed as a diminutive suffix, cf. Štarkl et al. 2024), sits somewhere in-between these two poles in both BCMS

Table 4.16: Prefixes across most frequent prefixed TV classes (only unsuffixed verbs) in Slovenian

Prefix	i/i		a/a		∅/e		All_V-izers	
	Freq	%	Freq	%	Freq	%	Freq	%
po	109	17.81	47	16.55	32	13.11	224	15.65
za	60	9.80	40	14.08	32	13.11	181	12.65
s/z/se	54	8.82	21	7.39	28	11.48	133	9.29
pre	49	8.01	27	9.51	20	8.20	123	8.60
o	55	8.99	22	7.75	11	4.51	107	7.48
u	39	6.37	10	3.52	17	6.97	77	5.38
iz	45	7.35	22	7.75	11	4.51	92	6.43
pri	35	5.72	17	5.99	18	7.38	92	6.43
na	37	6.05	12	4.23	18	7.38	84	5.87
od	29	4.74	18	6.34	19	7.79	78	5.45
raz	29	4.74	8	2.82	11	4.51	61	4.26
ob	18	2.94	10	3.52	8	3.28	45	3.14
v	18	2.94	11	3.87	8	3.28	43	3.00
do	18	2.94	9	3.17	4	1.64	41	2.87
vz	4	0.65	1	0.35	1	0.41	16	1.12
pod	6	0.98	2	0.70	3	1.23	15	1.05
pro	3	0.49	3	1.06	3	1.23	9	0.63
pred	2	0.33	2	0.70	0	0	6	0.42
so	1	0.16	1	0.35	0	0	2	0.14
nad	1	0.16	1	0.35	0	0	2	0.14
Total	612	100	284	100	244	100	1431	100

and Slovenian, with 50.91% and 60%, respectively.

While the quantitative data for secondary imperfectivizing suffixes is expected, as these typically attach to verbs previously made perfective by prefixation, the prefixation patterns with the remaining two groups are potentially more theoretically interesting. If suffixes like BCMS *-ira, -isa, -ova* and Slovenian *ira* are purely verbalizing ones, then it is interesting to consider why they are rarely prefixed, at least among the most frequent 5,300 (for BCMS) and 3,000 (for Slovenian) verbs in *WeSoSlaV*. While this can only be accomplished by including qualitative observations that are missing in *WeSoSlaV* (which makes it a task for future work), we can still quantitatively compare different types of traditional verbalizers: the suffixes like *-ira*, theme vowels like i/i, and the perfectivizing suffix *nu/ni*.

Table 4.17: Prefixation and suffixes - BCMS

Suffix	Freq	%VS	%All(5300)	P+VS	%P+VS	-P+VS	%-P+VS
av+a	258	14.72	4.87	244	94.57	12	4.65
ir+a	398	22.70	7.51	33	8.29	365	91.71
is+a	75	4.28	1.42	9	12.00	66	88.00
iv+a	326	18.60	6.15	321	98.47	3	0.92
k+a	55	3.14	1.04	28	50.91	25	45.45
n+0	389	22.19	7.34	280	71.98	107	27.51
ov+a	247	14.09	4.66	56	22.67	188	76.11

Table 4.18: Prefixation and suffixes - Slovenian

Suffix	Freq	%VS	%All(3000)	P+VS	%P+VS	-P+VS	%-P+VS
ov/ev+a	266	41.63	8.87	203	76.32	63	23.68
ev+a	34	5.32	1.13	26	76.47	8	23.53
av+a	26	4.07	0.87	25	96.15	1	3.85
ir+a	125	19.56	4.17	7	5.60	118	94.40
lj+a	2	0.31	0.07	2	100	0	0
k+a	10	1.56	0.33	6	60	4	40
et+a	5	0.78	0.17	2	40	3	60
t+e	3	0.47	0.10	2	66.67	1	33.33
ni/ne	142	22.22	4.73	115	80.99	27	19.01
v+a	26	4.07	0.87	22	84.62	4	15.38

The motivation for comparing the first two groups comes from the initial observations reported above, that prefixes are more frequently found with TVS than with verbalizing suffixes. The combinatorial potential of the BCMS suffix *nu* and Slovenian *ni/ne* with prefixes, on the other hand, is interesting because in contemporary BCMS and Slovenian these derive perfective verbs, just like prefixes, and both *nu*/*ni/ne* and prefixes are traditionally considered as markers of perfectivity (see Milosavljević 2023b for a recent overview). Moreover, *n+∅* and *ni/ne* are often analyzed as verbalizers (cf. e.g. Svenonius 2004a), so they can plausibly be compared to other traditional verbalizers. These comparisons are exactly what we turn to in the following section.

4.3.3.3 At the intersection of TVs and suffixes: Prefixes and verbalizers

In this section, we compare prefixation patterns with different verbalizers – TVs and suffixes *-ira*, *-isa*, *-ova* and *-nu* (*-nu* is spread across two TV classes in *WeSoSlaV* (∅/ne and u/e), i.e. excluding the verbs with verbal suffixes and suffix-like elements other than the verbalizing ones. This means that for the purposes of this section, we treat items like *ira* as verbalizers and not verbal suffixes.[7]

Table 4.19 summarizes the shares of different verbalizers across all unsuffixed verbs and prefixed unsuffixed verbs in BCMS, and Table 4.20 does the same for Slovenian (recall that 'unsuffixed' here refers to the exclusion of non-verbalizing suffixes such as secondary imperfectivizing ones).

Table 4.19: Prefixation across verbalizing affixes – BCMS

V-izer	Unsuffixed Vs (-VS)		Unprefixed Vs (-P-VS)			Prefixed Vs (P-VS)		
	Freq	%-VS	Freq	%-P-VS	%V-izer	Freq	%P-VS	%V-izer
∅/e	298	7.60	36	2.46	12.08	256	10.49	85.91
∅/ne	124	3.16	15	1.02	12.10	109	4.47	87.90
a/a	558	14.23	199	13.58	35.66	355	14.55	63.62
a/e	19	0.48	5	0.34	26.32	14	0.57	73.68
a/i	53	1.35	19	1.30	35.85	34	1.39	64.15
a/je	176	4.49	55	3.75	31.25	121	4.96	68.75
e/e	17	0.43	3	0.20	17.65	14	0.57	82.35
e/i	179	4.56	53	3.62	29.61	126	5.16	70.39
i/i	1520	38.76	363	24.78	23.88	1149	47.09	75.59
ira	398	10.15	365	24.91	91.71	33	1.35	8.29
isa	75	1.91	66	4.51	88.00	9	0.37	12.00
ova	247	6.30	188	12.83	76.11	56	2.30	22.67
u/e	258	6.58	92	6.28	35.66	164	6.72	63.57
Total	3922		1465			2440		

The data in Table 4.19 show that, overall, i/i is by far the most frequent verbalizer, accounting for 38.76% of all unsuffixed verbs. The pattern, however, significantly changes when prefixation is considered. Specifically, among the unprefixed verbs, the most frequent verbalizer is *ira* (24.91%), closely followed by i/i

[7] We keep here verbs with multifunctional suffixes, though, as some verbalizers often come together with certain multifunctional elements, e.g. *-ova* often comes with *-iz*, as in *verbal-iz-ova-ti* 'to verbalize'.

Table 4.20: Prefixation across verbalizing affixes – Slovenian

V-izer	Unsuffixed Vs (-VS)		Unprefixed Vs (-P-VS)			Prefixed Vs (P-VS)		
	Freq	%-VS	Freq	%-P-VS	%V-izer	Freq	%P-VS	%V-izer
a/a	626	22.81	254	31.24	40.58	372	19.25	59.42
i/i	802	29.22	189	23.25	23.57	613	31.73	76.43
a/je	348	12.68	82	10.09	23.56	266	13.77	76.44
∅/e	285	10.38	41	5.04	14.39	244	12.63	85.61
e/i	124	4.52	42	5.17	33.87	82	4.24	66.13
e/e	47	1.71	11	1.35	23.40	36	1.86	76.60
a/e	46	1.68	6	0.74	13.04	40	2.07	86.96
a/i	36	1.31	15	1.85	41.67	21	1.09	58.33
∅/ne	27	0.98	1	0.12	3.70	26	1.35	96.30
ir+a	107	3.90	101	12.42	94.39	6	0.31	5.61
ov/ev+a	237	8.63	59	7.26	24.89	178	9.21	75.11
v+a	26	0.95	4	0.49	15.38	22	1.14	84.62
ev+a	34	1.24	8	0.98	23.53	26	1.35	76.47
Total	2745		813			1932		

(24.78%) (with a difference of only two verbs). Among prefixed verbs, by contrast, i/i accounts for almost half of the verbs with 47.09%, whereas *ira* participates with only 1.35%. Looking from the perspective of the shares among unprefixed and prefixed verbs for each individual verbalizer, a clear division can be observed between traditional theme vowels (together with those that include the suffix *-n*, i.e. ∅/ne and u/e), on the one hand, and morphologically heavier verbalizers *ira, isa, ova*, on the other hand. In the former group, there are more prefixed verbs than unprefixed ones (ranging from the highest share of 87.90% of prefixed verbs in the case of ∅/ne to the lowest one of 63.57% in the case of u/e). This is in sharp contrast to *ira*, *isa* and *ova*, whose shares among prefixed verbs are 8.29%, 12% and 22.67%, respectively.

In Slovenian (Table 4.20), the difference between the most frequent verbalizers overall and the most frequent verbalizers among unprefixed verbs is less noticeable than in BCMS, though *ir+a* does also climb in frequency within unprefixed verbs compared to its frequency overall (i/i 29.22%, a/a 22.81%, a/je 12.68%, *ir+a* 3.90% vs. a/a 31.24%, i/i 23.25%, *ir+a* 12.42%, a/je 10.09%); but a similar division as in BCMS *can* also be observed in Slovenian with respect to traditional theme vowels on the one hand and the morphologically heavier verbalizer *ir+a*. Just like

in BCMS, the former group has more prefixed verbs than unprefixed ones (with the highest share of 96.30% of prefixed verbs for *∅/ne* and the lowest share of 58.33% in the case of a/i), which is in sharp contrast to *ir+a*, whose share among prefixed verbs is 8.29%.

These BCMS and Slovenian data raise the question of why prefixation patterns differ among the types of verbalizers. Putting aside the BCMS suffix *nu* and Slovenian suffix *ni* for now, one possibility may be that complex verbalizers adopt loan verbs significantly more often, and loan verbs in turn resist prefixation to a higher degree than native verbs. This possiblity remains a topic for further research, as *WeSoSlaV* currently does not annotate verbs for loan vs. native. However, evidence that the loan vs. native status of verbs is not the only factor comes from the study by Milosavljević et al. (2025), who considered minimal pairs of *i-* and *ova-* verbs in BCMS that share the base and differ only with respect to the verbalizer. These minimal pairs all include native bases, yet *i*-verbs are still far more frequent than their *ova*-counterparts. This then calls for other potential sources for the prefixation patterns, such as the type of the base (root vs. noun), aspectual status (pure imperfective vs. biaspectual), etc.

The distribution of different prefixes offers further insight into the interaction of prefixation and complex verbalizers. As can be observed from Table 4.21, for instance, the prefix *po-*, which is the most frequent prefix in both BCMS and Slovenian when all verbalizers are taken into account, is not attested in BCMS with verbs verbalized by the suffixes *ira* and *isa*, and is attested just once in Slovenian with verbs verbalized by *ira*.

Let us now move to quantitative data about the prefixation properties of verbs formed with *-nV/-ne* in both BCMS and Slovenian (which are encoded as theme classes u/e and i/e in BCMS and Slovenian sub-databases). The correlation between (im)perfectivity and the presence of a prefix is summarized in Table 4.23.

As is clear from the table, all prefixed verbs are perfective, which is expected due to the contribution of prefixes and the absence of suffix(-like) morphemes. Unfortunately, *WeSoSlaV* does not code whether the *nu-*/*ni*-bases that prefixes combine with are perfective or imperfective prior to prefixation. However, we report here the results of a follow-up quantitative study performed by Štarkl et al. (2024), which builds on *WeSoSlaV*, but additionally marks whether prefixed verbs would be perfective or imperfective if unprefixed. Specifically, out of 167 prefixed *-nV* verbs in BCMS, 95 (56.89%) combine with a perfective base, 23 (13.77%) combine with an imperfective base, while in 49 (29.34%) cases there is a bound base (i.e. a base that is not attested without a prefix). Out of the 119 prefixed *-nV* verbs in Slovenian, 43 (36.13%) combine with a perfective base, 13 (10.92%) combine with an imperfective base, while in 62 (52.1%) cases the base is bound.

Table 4.21: Distribution of prefixes across verbalizing suffixes in BCMS

Prefix	All	%All	ov+a	%ov+a	ir+a	%ir+a	is+a	%is+a
po	296	12.21	8	14.55	0	0	0	0
u	277	11.43	4	7.27	3	9.09	0	0
za	253	10.44	6	10.91	7	21.21	0	0
o	195	8.04	4	7.27	2	6.06	1	11.11
iz	185	7.63	3	5.45	7	21.21	2	22.22
na	166	6.85	3	5.45	0	0	0	0
s	165	6.81	2	3.64	2	6.06	1	11.11
pro	149	6.15	3	5.45	7	21.21	2	22.22
pre	133	5.49	3	5.45	1	3.03	3	33.33
od	132	5.45	4	7.27	0	0	0	0
raz	110	4.54	1	1.82	2	6.06	0	0
pri	101	4.17	2	3.64	0	0	0	0
do	74	3.05	3	5.45	0	0	0	0
sa	46	1.90	4	7.27	0	0	0	0
ob	44	1.82	2	3.64	0	0	0	0
pod	31	1.28	0	0	0	0	0	0
uz	27	1.11	1	1.82	0	0	0	0
su	13	0.54	2	3.64	2	6.06	0	0
nad	12	0.50	0	0	0	0	0	0
pred	10	0.41	0	0	0	0	0	0
vaz	3	0.12	0	0	0	0	0	0
mimo	1	0.04	0	0	0	0	0	0
suprot	1	0.04	0	0	0	0	0	0
Total	2424	100	55	100	33	100	9	100

The very existence of prefixed *-nV/-ne* verbs is theoretically significant since it shows that *-nV/-ne* and prefixes can be combined, contrary to some approaches reviewed in Štarkl et al. (2024). As Štarkl et al. (2024) report, the majority of such prefixes are lexical/internal prefixes, e.g. in BCMS: *pod-met-nu-ti* [UNDER-put-nV-INF] 'set up, put under', *od-gur-nu-ti* [FROM-push-nV-INF] 'push away', *s-kliz-nu-ti* [OFF-glide-nV-INF] 'slip', *u-tis-nu-ti* [IN-press-nV-INF] 'press in', *iz-tis-nu-ti* [OUT-press-nV-INF] 'press out'. Although in our main database (*WeSoSlaV*) there are no typical examples with superlexical prefixes, such verbs are possible, especially in the presence of another prefix, which is expected given that the most typical superlexical prefixes stack on top of other prefixes. Some such BCMS

Table 4.22: Distibution of prefixes across verbalizing suffixes in Slovenian

Prefix	All	%All	av+a	%av+a	ir+a	%ir+a	ov/ev+a	%ov/ev+a
po	350	15.48	6	0.24	1	14.29	33	16.26
za	244	10.79	3	0.12	0	0	19	9.36
s/z/se	207	9.16	0	0	2	28.57	12	5.91
pre	195	8.62	3	0.12	1	14.29	16	7.88
o	171	7.56	1	0.04	0	0	18	8.87
u	149	6.59	2	0.08	0	0	10	4.93
iz	144	6.37	3	0.12	0	0	16	7.88
pri	139	6.15	1	0.04	0	0	12	5.91
na	132	5.84	1	0.04	0	0	11	5.42
od	118	5.22	0	0	1	14.29	7	3.45
raz	103	4.56	0	0	0	0	9	4.43
ob	75	3.32	3	0.12	0	0	12	5.91
v	71	3.14	1	0.04	0	0	9	4.43
do	66	2.92	0	0	1	14.29	6	2.96
vz	33	1.46	0	0	0	0	2	0.99
pod	25	1.11	1	0.04	0	0	2	0.99
pro	16	0.71	0	0	0	0	1	0.49
pred	10	0.44	0	0	0	0	2	0.99
so	7	0.31	0	0	1	14.29	3	1.48
nad	6	0.27	0	0	0	0	3	1.48
Total	2261	100	25	1	7	100	203	100

examples, taken from Stojanović (2016), include: *iz-o-kre-nu-ti* [OUT-ABOUT-start-nV-INF] 'turn over all', *po-o-smeh-nu-ti* [OVER-ABOUT-laugh-nV-INF-REFL] 'laugh a little bit'. However, there are also superlexical-like prefixes, such as the attenuative *pri-*, which combine directly with *-nV* verbs, e.g. *pri-drem-nu-ti* [AT-doze-nV-INF] 'doze a little bit'. A similar picture is observed in Slovenian. An example of a LP-prefixed nV-verb is *iz-tis-ni-ti* [OUT-press-nV-INF] 'press out', whereas *po-na-tis-ni-ti* [OVER-ON-press-nV-INF] 'reprint' illustrates SLPs.

Finally, for completeness, in Table 4.24, we provide the distribution of different prefixes with two classes of verbs that contain the suffix *n*. The suffix *s-* is the most frequent one in verbs with *nu/ne*, i.e. *u/e* in BCMS, closely followed by the prefix *po-*. The prefix *po-* is by far the most frequent prefix in verbs with ∅/ne, alignining with the general pattern that displays *po-* as the most frequent prefix

Table 4.23: *-nV* verbs in *WeSoSlaV*: prefixation and (im)perfectivity

-nV verbs in *WeSoSlaV*	BCMS (258 in total, 4.87% of all the verbs in *WeSoSlaV*)		Slovenian (143 in total, 4.77% of all the verbs in *WeSoSlaV*)	
	Unprefixed	Prefixed	Unprefixed	Prefixed
All	91/258 (35.27%)	167/258 (64.73%)	24/143 (16.78%)	119/143 (83.22%)
Imperfective	9/258 (3.49%)	0 (0%)	3/143 (2.10%)	0 (0%)
Perfective	82/258 (31.78%)	167/258 (64.73%)	21/143 (14.69%)	119/143 (83.22%)

in verbs with all verbalizers. In Slovenian, the prefix *po-* is, similarly, the most frequent prefix in verbs with *ni/ne*, just as it is most frequent prefix in verbs with all verbalizers. Differently from BCMS, Slovenian verbs with ∅/ne reveal *o-* as their most frequent prefix (25%), which is only the third most frequent prefix in verbs with all verbalizers, but it should be noted that the absolute numbers are rather low here, just 6 in the case of ∅/ne verbs with *o-* (compare with 20 for BCMS ∅/ne verbs with their most frequent *po-*), so it is not clear that much can be drawn from this divergence of Slovenian from BCMS.

4.4 Summary

In this chapter, we presented some possibilities for using the database *WeSoSlaV* to explore properties of verb derivation in two Western South Slavic languages, BCMS and Slovenian. We first presented some basic descriptive statistics on different derivational categories (e.g. simplex verbs, prefixed verbs, etc.), and their interaction with aspect and TVs. Then we focused on prefixes – their frequency, as well as their interaction with other morphemes – suffixes and theme vowels. Here we summarize the most important findings that carry theoretical significance for formal investigations of the (Slavic) verb.

In both BCMS and Slovenian, simplex perfective verbs constitute a relatively small group of simplex verbs (under 5% in BCMS and under 8% in Slovenian). This supports the usual claim that such verbs are probably a closed, non-productive class.

Table 4.24: Distribution of prefixes with nV in BCMS

Prefix	nu/ne	%nu/ne	∅/ne	%∅/ne	All_V-izers	%All_V-izers
s	22	13.41	4	3.67	165	6.81
po	20	12.20	20	18.35	296	12.21
o	18	10.98	5	4.59	195	8.04
u	15	9.15	6	5.50	277	11.43
od	13	7.93	9	8.26	132	5.45
za	13	7.93	6	5.50	253	10.44
na	12	7.32	6	5.50	166	6.85
iz	10	6.10	10	9.17	185	7.63
pre	9	5.49	7	6.42	133	5.49
pri	6	3.66	12	11.01	101	4.17
uz	6	3.66	1	0.92	27	1.11
raz	5	3.05	4	3.67	110	4.54
do	4	2.44	5	4.59	74	3.05
pod	4	2.44	2	1.83	31	1.28
pro	4	2.44	5	4.59	149	6.15
ob	1	0.61	1	0.92	44	1.82
su	1	0.61	4	3.67	13	0.54
vaz	1	0.61	0	0	3	0.12
mimo	0	0	0	0	1	0.04
nad	0	0	1	0.92	12	0.50
pred	0	0	0	0	10	0.41
sa	0	0	1	0.92	46	1.90
suprot	0	0	0	0	1	0.04
Total	164	100	109	100	2424	100

Table 4.25: Distribution of prefixes with nV in Slovenian

Prefix	ni/ne	%ni/ne	∅/ne	%∅/ne	All_Vs	%All_Vs
po	18	17.65	2	8.33	350	17.04
za	7	6.86	4	16.67	244	11.88
s/z/se	0	0	0	0	0	0
pre	7	6.86	1	4.17	195	9.49
o	8	7.84	6	25.00	171	8.33
u	11	10.78	0	0	149	7.25
iz	8	7.84	0	0	144	7.01
pri	6	5.88	3	12.50	139	6.77
na	7	6.86	4	16.67	132	6.43
od	8	7.84	0	0	118	5.74
raz	3	2.94	0	0	103	5.01
ob	4	3.92	1	4.17	75	3.65
v	6	5.88	1	4.17	71	3.46
do	1	0.98	0	0	66	3.21
vz	6	5.88	2	8.33	33	1.61
pod	2	1.96	0	0	25	1.22
pro	0	0	0	0	16	0.78
pred	0	0	0	0	10	0.49
so	0	0	0	0	7	0.34
nad	0	0	0	0	6	0.29
Total	102	100	24	100	2054	100

In BCMS, prefixed verbs without additional suffixes strongly tend towards perfectivity, with only 1.81% of exceptions. In Slovenian, this tendency is much weaker, with 7.07% of imperfective verbs among prefixed unsuffixed ones. This is a quantitative confirmation of the observation made in Dickey (2003) (see also Marušič & Žaucer 2006) that this tendency is weaker in Slovenian.

In both languages, the TV a/a is the most frequent one overall and in the presence of verbal suffixes and suffix-like elements, whereas the TV i/i is the dominant one in unsuffixed verbs (i.e., simplex verbs and prefixed verbs without suffixes). These findings provide strong support for the approaches according to which the TV class assignment is not purely ornamental, but depends on several factors, including the derivational status of the base (cf. Milosavljević & Arsenijević 2022, Simonović et al. 2023, Kovačević et al. 2024).

The TV classes u/e in BCMS and its counterpart i/e in Slovenian only appear with a derivational suffix (*n*), which supports the view in Štarkl et al. (2024) that traditional suffixes *nu/ne* and *ni/ne* in BCMS and Slovenian, respectively, should be reanalyzed as bi-morphemic, consisting of the independently TV ∅/e and the suffix *-n*, with the vowels *u* and *i* being phonologically conditioned.

The fact that the frequencies of different prefixes compose a gradual cline can be taken as quantitative evidence against singling out any of the prefixes in BCMS or Slovenian as a purely grammatical perfectivizing prefix. This goes against the view in Dickey (2012), who argues that Slovenian, unlike BCMS, employs a grammaticalized purely perfectivizing prefix. In Slovenian, *po-* sticks out slightly, but clearly not with a percentage that would suggest the status of a grammaticalized perfective marker; moreover, Dickey actually identifies *s-/z-* as the partially grammaticalized empty perfectivizer in Slovenian, not *po-*.

The quantitative patterns observed with prefix stacking in both languages provide support for the approaches arguing that prefix stacking cannot be taken as any sort of clear-cut diagnostic for distinguishing between lexical and superlexical prefixes. That is, without distinguishing between specific uses of individual prefixes, almost all prefixes that function as the first prefix can also function as the second one, i.e., all except the archaic *vaz-* in BCMS and *pod-* in Slovenian. This includes prefixes that are typically considered lexical or that are never found on lists of superlexical prefixes but apparently also occur as the second, i.e., stacked prefix: *u-* 'in', *s-* 'from', *o-* 'about', *ob-* 'around', *raz-* 'apart', *uz-/vz-* 'up', and in BCMS also *pod-* 'under'.

There are significant correlations between prefixes and suffixes. As expected, traditional secondary imperfectivizing suffixes predominantly co-occur with prefixes, as secondary imperfectivization most often derives imperfective verbs from the perfective ones derived by prefixation. Somewhat surprisingly, in both BCMS and Slovenian, verbs derived by the perfective suffix *n* (traditional *nu/ne* and *ni/ne*-verbs, respectively), are predominantly prefixed in both languages, despite the fact that in contemporary BCMS and Slovenian both the suffix *n-* and prefixes tend to derive perfective verbs.

Some suffixes usually analyzed as purely verbalizing ones, e.g. *ira* in both BCMS and Slovenian, strongly resist prefixation: in both languages, more than 90% of the verbs with this suffix are unprefixed (similar tendencies are observed for *isa* in BCMS and to a lesser extent to *ova* as well). This sharply contrasts with verbs whose verbalizers are exponed by TVs, where prefixed verbs significantly prevail over unprefixed (simplex) ones. This in turn calls for exploring the effect of additional factors on prefixation, e.g. morphological complexity of the verbal-

izer, the type of the base (root vs. noun), the aspectual status of the base verb (purely imperfective vs. biaspectual), and its status as a loan or native stem.

5 Deverbal derivation

5.1 Introduction

The goal of this chapter is to overview the information in *WeSoSlaV* which is relevant for the investigation of deverbal derivations, to illustrate one possible way of using the database for this purpose, and to report and discuss the results of an explorative investigation in this direction, as well as to draw conclusions informing future research.

With the exception of two nominalization patterns to be discussed in Section 5.10, *WeSoSlaV* itself is not intended to include information about words other than verbs. Therefore, the only (deverbal) derivation that the database richly informs us about are derivations of verbs, i.e. verbs derived by the suffixes such as *-nu*, *-ta*, *-iva*, *-ova*.

Theoretical relevance of this chapter most directly concerns the notions of domains of productivity of morphological operations. A good illustration is the derivation of deverbal nominalizations by the suffix *-nje/će*. This suffix has been described as fully productive with imperfective verbs, yet idiosyncratically used with perfectives (see Arsenijević 2020, Simonović & Arsenijević 2014 for a discussion of both this generalization and some of its further theoretical consequences). The very possibility for a morphological operation to be limited to a domain of bases within a lexical class opens the question of how these domains can be determined. In particular: are they always flat, i.e., defined by one property, such as aspectual class, or can they be an intersection between multiple properties (for instance: perfective prefixed reflexives). If it may only be one property, why is it so, what does it tell us about grammar? If, rather, multiple properties may be involved, are there any restrictions what properties combine with each other? How small can domains of productivity be? Answers to these questions are crucial for the understanding of the notion of (morphological) productivity, both in the purely structural and in the cognitive perspective. As these questions are complex, and require a broad research, the current chapter does not have the ambition of offering definitive answers, but rather aims to showcase some possible empirical methods to tackle them, based on the *WeSoSlaV* database.

Another theoretical contribution of the chapter lies in its explorative character. Consider that its result indicate that a morphological operation M is fully productive in the intersection of the value V1 of property P1, value V3 of property P2 and value V3 of property P3, but otherwise not or only idiosyncratically. This prompts the question why exactly this intersection interacts with the morphological operation. What is it about the phonological, syntactic or semantic properties of the operation that is sensitive to the combination of the exact property values? Answering questions like this has a great potential of advancing not only our understanding of the particular properties and operations, but also of the most general aspects of language. Importantly, regularities of the described kind would hardly be observed without a quantitative explorative research such as that reported in this chapter.

Besides its research contribution, this chapter also has a specific role in the context of the volume as a whole, namely to illustrate how *WeSoSlaV* can fruitfully be expanded, i.e. combined with additional quantitative investigations, when the theoretical goal requires it. For instance, one can excerpt from the corpus all the deverbal derivations by a particular suffix, or set of suffixes, and then combine the result with *WeSoSlaV* to investigate the effect of the properties annotated in the latter for the application of a particular morphological operation. This of course can be done both for deverbal verbs, and for other categories derived from verbal bases. The latter is what we did for a selection of suffixes, and what we report in this chapter. In the same vein, we explore and discuss the similarities and differences between judgment-based annotation as implemented to develop *WeSoSlaV*, and corpus-based annotation which we implement as the main tool to explore deverbal derivations in this chapter.

The chapter is organized as follows. Section §5.2 presents the method. §5.3 reports the results for BCMS and §5.4 discusses them. §5.5 reports the results for Slovenian, which are discussed in §5.6. In §5.7, the comparative results for the two languages are provided. §5.9 focuses on the suffix *-aj* in BCMS, showing how the quantitative data can fruitfully be used to resolve several open questions about the derivation with this suffix. In Section §5.10, we compare two annotation methods in the domain of deverbal derivations, one based on the corpus search and the other on annotator judgments. §5.11 brings the conclusion.

5.2 Method

This section describes how the database of deverbal derivations that was used in the investigation was built to serve the research goals from §5.1. We searched

the same corpora on which *WeSoSlaV* was based (in particular, we used hrWaC and srWaC for BCMS and Gigafida 2.0 for Slovenian) for lemmas occurring more than 5 times and ending in each of the target suffixes. The sub-database for each of the suffixes was then manually cleaned from the lemmas that are not derived by adding the intended suffix to a verb. In particular, we excluded those lemmas that do not involve the suffix in the first place, as well as from those that do, but the suffix is added to another lexical category or to a root were excluded. We additionally excluded the deverbal derivations whose base verbs are not included in *WeSoSlaV*, because they could not be subject to the intended analysis due to the lack of annotation for their base verb.

Next, the remaining deverbal derivations and their attested frequencies in the corpus were entered as annotated properties for the respective base verbs in a copy of the aggregate *WeSoSlaV* in two newly introduced columns. As a result, a version of *WeSoSlaV* expanded by the annotation of corpus attestations of derivations by each of nine selected suffixes for each verb was obtained. This allowed us to investigate the potential effects of all the properties already annotated in *WeSoSlaV* on the possibility of combining the verb and the respective suffix. The expanded databases used for deverbal derivation are available as Arsenijević, Marušič, et al. (2024).

We decided to include suffixes that derive significant quantities of deverbal lexemes (at least 50 attestations with over 5 occurences in the corpus), in both BCMS and Slovenian, while also trying to have both of the other lexical categories that derive from verbs represented, namely, nouns and adjectives. A verb can also be the base for the derivation of another verb, but this direction of derivation could be investigated on the core *WeSoSlaV* alone, as it contains both simple and derived verbs. More nominal than adjectival suffixes were included in the investigation because generally there are more suffixes that derive large numbers of nouns from verbs than those that derive adjectives. The following suffixes were included for BCMS:

- *-lac*, deriving agent nouns, such as *gleda-lac* ‘watcher’
- *-telj*, deriving agent nouns, such as *gleda-telj* ‘watcher’
- *-lište*, deriving event locations, such as *gleda-lište* ‘audience’
- *-ač*, deriving agent nouns, such as *gled-ač* ‘watcher’
- *-ba*, deriving event nouns, such as *druž-ba* ‘company’
- *-((l)j)iv*, deriving possibility adjectives, such as *gled-ljiv* ‘watchable’

- *-nik*, deriving patient nouns, such as *gleda-nik* 'the one who is watched'
- *-nje/će*, deriving event nouns, such as *gleda-nje* 'watching'
- *-aj*, deriving event nouns, such as *pregled-aj* 'a round of examining'

For Slovenian, we included these suffixes:

- *-lec*, deriving agent nouns, such as *gleda-lec* 'watcher'
- *-telj*, deriving agent nouns, such as *gradi-telj* 'builder'
- *-lišče*, deriving event locations, such as *gleda-lišče* 'theater'
- *-nja*, deriving event nouns, such as *grad-nja* 'building'
- *-ava*, deriving event nouns, such as *napelj-ava* 'wiring'
- *-nik*, deriving participant nouns, such as *sod-nik* 'judge'
- *-nec*, deriving patient nouns, such as *tože-nec* 'the one who is sued (defendant)'
- *-nje*, deriving event nouns, such as *gleda-nje* 'watching'
- *-ač*, deriving agent nouns, such as *kop-ač* 'digger'

After having selected the suffixes and annotating their frequencies, we have determined a subset of the properties of verbs annotated in the aggregate database which we expected to be relevant in deciding whether a verb can be the base for the selected suffixes. These included, for BCMS: the theme vowel, the frequency of the verb, whether the verb is imperfective, whether it can take an accusative argument, whether it has a reflexive variant with a changed meaning, whether it is prefixed and whether it is suffixed. For Slovenian, these included: the theme vowel, the frequency of the verb, whether the verb is imperfective, whether it may take an accusative argument, whether it displays root allomorphy in its paradigm, whether it has a reflexive variant with a changed meaning, whether it is prefixed, whether it is suffixed, and whether it involves ablaut. Important to note is that not all of these properties are part of the fully completed databases published on CLARIN's repositories. In particular, this holds for the properties of selecting an accusative direct object and changing meaning in reflexivization. Both these properties, together with all the other included in this chapter, are part of the preliminary database published as Arsenijević, Marušič, et al. (2024).

Some of the included suffixes are underlyingly complex, i.e., they consist of two or three simple suffixes. For instance, suffix *-nje/će/tje* consists of the participial suffix *-en/an* and the mass nominal suffix *-je* (Simonović et al. 2024). Since we were interested, descriptively speaking, in morphological operations, rather than simple suffixes, we glossed over this morphological complexity and considered each of these formants as single suffixes.

In order to establish the effects of the analyzed properties on the probability that a verb undergoes derivation with a particular suffix, we fitted a Mixed Effects Logistic Regression model to the data for each suffix. The dependent variable was hence whether or not a word derived by the respective suffix was attested in the corpus for the respective verb as the base, and the predictors were those listed above for both languages. Among these, the theme vowel class of the verb was coded as a categorical variable with 14 levels (see Chapter 3 for more details), the frequency of the verb as a discrete quantitative variable. All others were binary variables coded as discrete quantitative variables with levels 0 and 1 (see Chapter 2 for how each of these properties was determined). Function *glm* from the package lme4 was used, as follows (here illustrated for the suffix *-telj* in BCMS):

(1) glm(formula = telj ~ Theme + frequency + Imperfective + Acc + Reflexive + Prefixed + Suffixed, family = binomial(link = "logit"), data = bcms)

5.3 Results, BCMS

The suffixes under discussion, illustrated in (2), turned out to be heterogeneous with respect to their productivity in the corpus. By productivity in the corpus (which we simply call productivity henceforth), we mean the percentage of verbs in *WeSoSlaV* which undergo derivation with the respective suffix such that the derived word is attested in the corpus. They roughly form three clusters: low productivity suffixes, moderate productivity suffixes, and one highly productive suffix. In BCMS, for instance, these are, respectively: suffixes with fewer than 100 attested derived words (suffixes *-nik*, *-aj*, *-ba* and *-lište*), suffixes with between 100 and 1000 attested derived words (suffixes *-lac*, *-telj*, *-iv* and *-ač*), and the suffix *-nje/će*, with its 2563 attested derived items (i.e. almost one half of all the verbs in the database derive a word that is attested in the corpus with this suffix). One can say that suffix *-nje/će* is fully productive within imperfective verbs and has restricted productivity among perfectives, that suffixes *-lac*, *-telj*, *-iv* and *-ač* have moderate productivity, and that suffixes *-nik*, *-aj*, *-ba* and *-lište* are close to fully unproductive.

(2) štiće-nik pokuš-aj kosit-ba igra-lište čita-lac uči-telj
protect-*nik* try-*aj* mow-*ba* play-*lište* read-*lac* teach-*telj*
'protégé.N' 'attempt.N' 'mowing.N' 'playground.N' 'reader.N' 'teacher.N'
zamisl-iv pev-ač snima-nje
imagine-*iv* sing-*ač* record-*nje*
'imaginable.ADJ' 'singer.N' 'recording.N'

In tables 5.1 and 5.2, we report paired results. On the one hand, the cells include for each relevant variable level the number of verb members of that level which undergo derivation for the respective suffix such that the derived word is attested by our corpus search. On the other hand, paired with the raw quantities, we summarize the results of the statistical analysis, i.e. we code significant effects and their direction for the variables included in the research. The effects are specified using the standard coding: p<.05 is marked by one asterisk, p<.01 by two, and p<.001 by three asterisks. We also mark what can be indicating a tendency (i.e. where p<.1), by a point. Considering sample sizes, which in all cases other than the highly productive suffix *-nje/će* range between 61 and 303 observations, these are very likely to indicate actual effects which were not captured due to the scarcity of data. A minus sign as a prefix codes that the effect is negative: the property in question decreases the probability that the verb will combine with the given suffix.

Table 5.1 does this for theme vowels, for which the effects are reported for each level of the variable (each theme vowel), due to its categorical nature. Theme a/i is taken as the baseline. For the remaining properties, which are all binary, the results are reported in table 5.2. Here the reported quantities are those where the verbs bearing the property in question (e.g. the imperfective verbs, those compatible with an accusative etc.) are attested as bases of deverbal derivations. Totals of variable levels (row totals, column *All*) report the aggregate size of the variable level in *WeSoSlaV* (e.g. there are 1702 a/a verbs in the database, and 2730 imperfective verbs, all together). As is obvious from the tables, no variable level has 100% productivity for any suffix.

Additionally, we specify the effect size of the model in terms of R^2 (as the test is run on all the variables at the same time, the effect sizes are the same in both tables), the sample size (*Sample*) - i.e. the aggregate number of deverbal derivations by the particular suffix attested in the corpus, and the meaning of their derived word (label *Sem*, with values *event*, participant – coded *prtc*, property – coded *prop*, and *place*) for each suffix. As the *theme vowel* variable is categorical, the effect of each level is given separately, with the theme vowel a/i taken as the baseline. The reason for taking this theme as the baseline is that it was ordered

as an extreme by the heatmap analysis, and that it is not grouped with the large, productive theme vowel classes (so that we get explicit results for each of these classes).

As expected, suffixes with smaller sample sizes also have fewer significant effects. This suggests that due to the sample sizes, possibly not all the actual factors among the included variables are captured as statistically significant (but see the discussion in §5.7 showing that an even smaller database gives results almost identical to those from very large data sets). However, suffixes with a small number of derivations are typically more idiosyncratic rather than regular in their choice of bases. Their idiosyncrasy too predicts that these suffixes will show a smaller number of effects even with larger samples (to the extent that larger samples are possible, considering their productivity), i.e. that the scarcity of effects is not due to small samples, but due to the irregularity of these derivations. Further research is hence necessary to obtain a definite picture of their behavior. Effect size too, as expected, shows a correlation with sample sizes.

To get an overview of the relative productivity of thematic classes, we divided each section of the table within the theme vowel by the size of the thematic class. This yielded the quantities in Table 5.3 (the defective class was too small for an analysis and has been removed), with the portion of each thematic class that combines with the respective suffix. Then we used R to compute heatmaps including information about clustering by similarity (function *heatmap.2* from the package gplots) for the matrix we got. The result is provided in Table 5.1.

The heatmap shows that the suffix *-nje/će* is clearly the most productive suffix overall, as it is by far the most productive one for all theme classes except for class e/i. This class sticks out, as it is more productive with the suffix *-iv*.

Other than that, theme classes show a clear split to those involving *a* in the infinitive stem plus theme i/i and the rest, i.e. those involving a zero theme, *e* or *u* in the infinitive (with the exception of class e/i, they all have an *e* in the present tense stem). The former group includes the three largest theme classes, a/a, i/i and a/je, i.e., the only two classes that are still productive (a/a and i/i).[1]

The other five theme classes are characterized by a high productivity of the suffix *-iv*, and for most of them also a particularly low productivity with a few suffixes.

[1]New verbs are also derived in classes u/e and a/je via suffixation, as verbal suffixes *-ov/ev*, -iv and one version of *-av* take the theme a/je, and the semelfactive suffix *-n* takes u/e. We assume here that verbs derived by verbal suffixes do not count for theme productivity, because each suffix has a fixed theme, and should therefore count as one verb in all the new verbs it derives, i.e. those verbs are to be seen as compounds of the suffix-verb with other verbal stems.

Table 5.1: Raw quantities and effects of BCMS theme vowel classes on the productivity of individual suffixes

TV	nik	ište	aj	ba	lac	telj	iv	ač	nje/će	All
a/a	7	30	36 –**	5	25 –*	26 –***	22 –***	150	1114	1702
i/i	52	21	4 –**	58	72	111	145 –.	27 –.	576	1601
a/je	2	14	28 –**	2	11 –**	22 –**	11 –***	114	780	1029
∅/e	0	1	1	2	2 –*	1 –*	6 –***	6	19 –***	298
u/e	0	0	0	0	1	1 –*	8 –**	0	21 –**	258
e/i	0	1	0	0	1 –.	0	10 –.	0	4 –***	184
∅/ne	0	0	0	1	0	0	4 –*	1	9 –*	124
a/i	0 bs-ln	2 bs-ln	3 bs-ln	0 bs-ln	2 bs-ln	3 bs-ln	7 bs-ln	3 bs-ln	28 bs-ln	62
a/e	0	0	0	0	0	0	1	2	8	19
e/e	0	0	0	0	0	0	1	0	3 –*	17
defective	0	0	0	0	0	0	0	0	1	6
R^2	.028	.023	.040	.039	.055	.058	.042	.080	.443	
Sem	prtc	place	event	event	prtc	prtc	prop	prtc	event	
Sample	61	68	69	72	114	164	215	303	2563	

Table 5.2: Raw quantities and effects of the selected properties of BCMS verbs on the productivity of individual suffixes

	nik	ište	aj	ba	lac	telj	iv	ač	nje/će	All
Ifv	14	62	72	36	76	94	67	287	2130	2730
	_*	*		.	**	*		***	***	
Acc	54	52	54	58	106	152	188	269	2075	4027
				.	***	***	***	***	***	
Refl	25	23	23	24	29	48	69	71	624	1217
	.	**	*						**	
Pref	44	22	61	33	60	99	162	182	1442	3660
	_*	_**	***							
Suff	0	17	48	0	21	28	19	153	783	966
			*					***		
Freq		*			***	***	**	*	**	
R^2	.028	.023	.040	.039	.055	.058	.042	.080	.443	
Sem	prtc	place	event	event	prtc	prtc	prop	prtc	event	
Sample	61	68	69	72	114	164	215	303	2563	

Table 5.3: Relative quantities of verbs undergoing derivations by each suffix per theme vowel class

TV	nik	ište	aj	ba	lac	telj	iv	ač	nje/će
∅/e	0,00	0,00	0,00	0,01	0,01	0,00	0,02	0,02	0,06
e/i	0,00	0,01	0,00	0,00	0,01	0,00	0,05	0,00	0,02
i/i	0,03	0,01	0,00	0,04	0,04	0,07	0,09	0,02	0,36
a/a	0,00	0,02	0,02	0,00	0,01	0,02	0,01	0,09	0,65
a/je	0,00	0,01	0,03	0,00	0,01	0,02	0,01	0,11	0,76
∅/ne	0,00	0,00	0,00	0,01	0,00	0,00	0,03	0,01	0,07
a/e	0,00	0,00	0,00	0,00	0,00	0,00	0,05	0,11	0,42
u/e	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,00	0,08
a/i	0,00	0,03	0,05	0,00	0,03	0,05	0,11	0,05	0,45
e/e	0,00	0,00	0,00	0,00	0,00	0,00	0,06	0,00	0,18

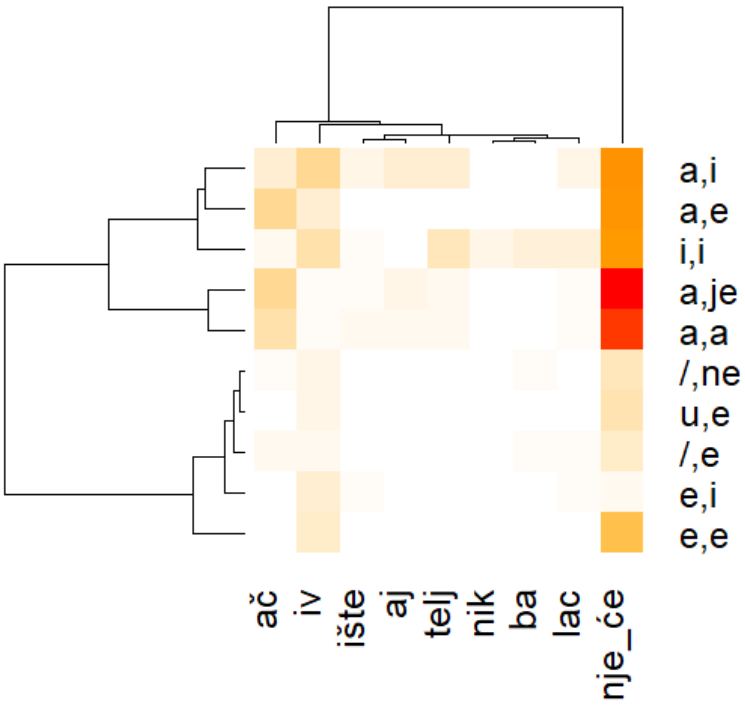

Figure 5.1: The heatmap for BCMS theme vowel classes

Extreme pairwise closeness is attested in the heatmap analysis for three pairs of theme vowels. One is a/i and a/e , two relatively small classes combining *a* in the infinitive stem with a front vowel in the present stem. Next, very similar are classes a/a and a/je, both of which independently show similar properties. In addition to sharing the vowel in the infinitive stem, they are both still productive in the derivation of new verbs, and both are characteristic for a set of verbal suffixes (verbs derived by suffixes *-av*, *-k*, *-ir* take a/a, and those derived by suffixes *-ov*, *-iv*, *-t*, *-is* the theme a/je. Finally, themes ∅/ne and u/e show extreme similarity. These themes are often seen as one class, since both have the sequence *ne* in the present tense, and a majority of the ∅/ne verbs can also be used in the u/e paradigm.

Looking at the suffixes, suffix *-nje/će* sticks out as by far the most productive one, but other suffixes show neither clear-cut divisions nor extreme similarities. If we compare the suffixes pairwise in terms of sensitivity to theme classes, more similar than others are suffix pairs *-nik* and *-ba*, and *-ište* and *-aj*. It is hard,

however, to identify any other properties which connect these pairs, in order to indicate some deeper reality of the observed similarities.

5.4 Discussion, BCMS

In this section we focus on the big picture: the comparison of the selection of suffixes regarding the effects of investigated properties of base verbs on their productivity. As an illustration of how this method can be used to explore individual suffixes, we engage in a somewhat more detailed discussion of the suffix *-aj*, which derives nominalizations denoting instantaneous episodic events and is theoretically relevant due to its varying selectional restriction between verbs of the perfective and imperfective class.

Clearly, whether a derived word will be attested in the corpus or not does not only depend on whether it is well-formed, but also on a range of mostly pragmatic factors determining its frequency in use. We work under the assumption that these factors do not interact with the properties of verbs we investigate as predictors of the derivational capacity of a verb, and hence affect all the relevant classes in approximately the same way.

However, probably largely due to this fact, that the derivation of a word needs to be pragmatically motivated, the quantitative data could not lead to the kind of insights we were looking for, namely to niches of full productivity of certain suffixes. For instance, even within the crossing of three strong factors of the productivity of the suffix *-ač*, selection of an accusative direct object, imperfectivity and suffixation, the productivity of this suffix based on attested forms in the corpus remains at only 10.80%, which is double its productivity with the rest of the verbs, but still far from full productivity. Nevertheless, the insight that these three properties have such strong effects on the prodctivity of suffixation by the suffix *-ač* calls for further theoretical engagement in explaining the interaction.

5.4.1 Theme vowels

Regarding productivity, BCMS theme vowels can be divided in three classes. One includes the directly productive themes, those that can be added in present day BCMS to a base without the mediation of a verbal suffix to derive verbs, and these include themes a/a and i/i, as in (3a). Next, there are themes which can derive new verbs, but only mediated by verbal suffixes such as *-iv/uj*, *-ov/uj*, *-n*, or *-av*, as in (3b). This mainly applies to the themes represented here as a/je and u/e, and while a/a also shows this pattern of behavior, it is, as already mentioned, also

directly productive. Finally, there are themes, such as ∅/e and ∅/ne, which are not productive any more, and only live on old verbs, on which they are inherited from older stages in the development of BCMS, as in (3c).

(3) a. u-anksioz-i-m lajk-a-m
in-anxious-TV-PRES.1.SG like-TV-PRES.1.SG
'make anxious' 'like'

b. stalk-u-je-m restart-n-e-m
stalk-V-TV-PRES.1.SG restart-V-TV-PRES.1.SG
'stalk' 'restart'

c. ber-e-m uzm-e-m
pick-TV-PRES.1.SG take-TV-PRES.1.SG
'pick' 'take'

Overall, the data in 5.3 show that the moderately productive suffixes and the suffix *-nje/će* display a rather even distribution of productivity across theme vowels. The unproductive suffixes display a higher degree of specialization. This may indicate that when these suffixes used to be productive, they were bound to certain theme classes, while other suffixes (or allomorphs) achieved the same effect in other theme classes.

It is insightful to look at how themes behave regarding the three suffixes which have similar semantics and hence compete with each other, namely the suffixes tending towards the agent interpretation: *-lac*, *-telj* and *-ač*. Looking only at the three large theme classes, a/a and a/je (both with an *a* in the infinitive) strongly favor suffix *-ač* (three times more *-ač* derivations than the sum of *-lac* and *-telj*) with an initial *a*, unlike theme i/i, which shows the exact inverse: each of *-lac* and *-telj* are 3-4 times larger than *-ač*. This fits in the view expressed about the unproductive suffixes, namely that they are remnants from a state where they were in competition with other suffixes with similar semantics, and were favored by some themes.

5.4.2 Frequency

We expected that frequency will play a role in the sense that derivations from more frequent verbs will also be more likely to be frequent in the same corpus. This was indeed confirmed for a majority of the BCMS suffixes. For suffixes *-lište*, *-lac*, *-telj*, *-iv*, *-ač*, *-nje/će*, frequency of the verb showed a positive effect on the number of attested derivations. The list includes the five most productive suffixes, i.e. the five suffixes with the largest sample, which suggests that the

effect might emerge with all suffixes, provided a larger sample. However, see §5.6.2, where Slovenian data suggest that no across the board effect of frequency can be generalized.

5.4.3 Compatibility with an accusative argument

Verbs that can take an accusative argument are generally more productive in derivation with the suffixes included in the investigation. Again, positive effects are confirmed for the five most productive suffixes, with a tendential effect for the sixth (suffix *-ba*). Interestingly, the effect was not significant for the suffix *-nik*, which typically denotes the theme or patient of the eventuality denoted by the verb, and hence is expected to correlate with the compatibility with a direct object.[2] The reason for this is likely in the small number of observations for this suffix (only 61 derived noun, the least of all the suffixes).

5.4.4 Meaning shift with the reflexive marking

Verbs were annotated to have this property if they could be used both with and without the reflexive clitic, yet not as a minimal pair of a reflexive and non-reflexive use, but with clearly different meanings. The presence of the meaning shift was confirmed by the fact that the same verb that is used with the reflexive clitic, which is generally ambiguous between marking reflexivity and argument structure effects (unaccusatives, middles, passives, impersonals etc.), could not be used with the strong reflexive pronoun *sebe*, which is restricted to the reflexive meaning. This is illustrated for BCMS in (4).

(4) a. nasmejati, izleteti
smile.INF fly_out.INF
'make smile' 'fly out'

b. nasmejati se, izleteti se
smile.INF REFL fly_out.INF REFL
'smile' 'let the cat out of the bag'

Within BCMS data, this factor had a significant positive effect on deverbal derivations for three suffixes, and a marginal tendency for one. Curiously, these

[2] Suffix *-nik* is underlyingly complex, and consists of the passive participle ending *-en*/*-an* (depending on the theme class), and the more general nominalizer *-ik*. While the latter suffix derives many different kinds of nouns, in combination with the passive participle, it compositionally derives patient nouns. For a discussion of both its morphological complexity and semantics, see Klajn (2002), Babić (2002).

include the three least productive suffixes, *-nik*, *-lište* and *-aj*, and the most productive one: *-nje/će*. We are not in a position to provide an explanation for these results at this point, but we consider it a result that calls for further research into the interaction of verbal properties and derivational operations taking them as bases. Different results are obtained on Slovenian data, as discussed below in §5.6.4.

5.4.5 Imperfectivity

Imperfectivity seems to generally facilitate deverbal derivations. Positive effects are attested with five out of the nine suffixes (*-lište*, *-lac*, *-telj*, *-ač* and *-nje/će*), and a positive tendency with one more (*-ba*). These include five out of the six most productive suffixes (*-iv* is the only one among them that did not reach at least $p<0.1$).

However, while in all other cases the effect was positive, with the suffix *-nik*, it was negative. Considering that, as discussed below, this suffix is also inhibited by the verb having a prefix, and that prefixation typically perfectivizes, the explanation might be that this suffix is bad with secondary imperfectives. This hypothesis is undermined by the absence of effect of suffixation of the base verb on this suffix, as secondary imperfectives are all annotated in our database as suffixed, even when the relevant morphological contrast is only in the theme vowel or only its prosody. It is therefore another question for future research why derivation by the suffix *-nik* is negatively sensitive to imperfectivity and(/or) prefixation.

5.4.6 Prefixed verbs

Like reflexivity, prefixation only showed effects on suffixes with small samples. It had a negative effect on the derivation with the suffix *-nik*, already discussed, as well as with the suffix *-lište*, and a positive effect with the suffix *-aj*, the three least productive suffixes. No other significant effects are attested.

Assuming with Arsenijević (2023) that prefixation correlates with telicity, or with Milosavljević (2023b) that it presupposes telicity and reduces it to singularity, all prefixed verbs, perfective or imperfective, include in their semantic composition a telic predicate. The positive effect in the case of the suffix *-aj*, which derives strictly telic event nominalizations, is thus likely because not all verbs easily license a telic interpretation, but those including a telic predicate certainly do.

The negative effect with the suffix *-lište* is more difficult to explain. One possible source is the tendency of the derived nouns to have the phonological template involving four syllables and the default prosody. Prefixed verbs usually result in nominalizations longer than the template, and often also with a lexically specified prosodic pattern.

5.4.7 The base verb includes a verbal suffix

Availability of a verbal suffix on the base verb had a significant effect for only two deverbal derivational suffixes: *-aj* and *-ač*. The case of *-aj* is discussed from the perspective of its competition with verbal suffixes in the base in section 5.9. For the suffix *-ač*, the effect probably has to do with its competition with suffixes *-lac* and *-telj*. All three suffixes derive *agent nouns*, and they compete for verbal stems. The other two suffixes, however, seem to avoid suffixed verbs. The absence of negative effects for the two verbs is probably due to their smaller samples.[3]

5.5 Results, Slovenian

Similarly to what was the case for BCMS, Slovenian suffixes included in the investigation, illustrated in (5), ranged from suffixes with a very low productivity to suffixes of extreme productivity, that is, from suffixes represented by only 26 attested derived words to suffixes represented by almost 2500 derived words from the set of 3000 high frequency verbs. In Table 5.4, we report the raw numbers of members of a particular theme vowel class which undergo derivation with the respective suffix, and in Table 5.5 for the remaining properties of verbs considered, in both cases together with the statistically significant effects (as in Section 5.3, significance of effects is marked with dots and asterisks: . for < .1, * for < .05, ** for < .01, *** for < .001; a minus in front marks that the effect is negative, i.e. decreases the probability that the verb will undergo derivation with the given suffix).

(5)

grad-nja	napelj-ava	gleda-lišče	kop-ač	gradi-telj	sod-nik
build-*nja*	lead-*ava*	watch-*lišče*	dig-AČ	build-*telj*	judge-*nik*
‘construction’	‘wiring’	‘theater’	‘digger’	‘builder’	‘judge’

gleda-lec	gleda-nec	gleda-nje
watch-*lec*	watch-*nec*	watch-*nje*
‘watcher’	‘one who is watched’	‘watching’

Table 5.4: Raw quantities and effects of Slovenian theme vowel classes on the productivity of individual suffixes

TV	nja	ava	lišče	ač	telj	nik	lec	nec	nje_tje	All
a/a	19	34 ***	51 .	95	13 _***	83 _***	191 _**	151 _.	891 ***	1044
i/i	20 bs-ln	12 bs-ln	11 bs-ln	0 bs-ln	102 bs-ln	185 bs-ln	90 bs-ln	228 bs-ln	5 bs-ln	863
a/je	1 _.	11 **	15 .	12	10 _**	35	199 ***	95 *	299 ***	378
∅/e	15 **	0	2	0	6 _***	41	15	26 _***	114 ***	285
i/e	1	0	1	0	1 .	0	0	33	0	143
e/i	3	0	1	0	0	20	4 _**	9 _*	53 ***	128
e/e	2	0	1	0	1 _*	6 _*	1 .	2 _*	29 ***	47
a/e	1	5 **	0	0	0	4 _**	6	13	8 ***	46
a/i	0	4 ***	1	3	1	6 _.	2	6	22 ***	36
∅/ne	0	5 ***	0	0	1	3 .	1	0	12 ***	27
Defective	0	0	0	0	0	0	0	0	3	3
R^2	.001	.003	.003	.133	.006	.010	.001	.076	.213	
Sample	62	71	83	110	135	383	509	563	2469	
Sem	event	event	place	prtc	prtc	prtc	prtc	prtc	event	

Just as we did for BCMS, we also report the data relative to theme class sizes in Table 5.6 (leaving the defective class out, as it barely derives any new words) by dividing each combination of a theme and a suffix with the aggregate size of the theme class (the latter is given in column *All* in 5.4). As expected, not all suffixes are equally productive and there are also some differences in their productivity

[3]We had originally also included root allomorphy among the investigated properties, but it showed very few effects in both languages, so we decided to leave it out.

Table 5.5: Raw quantities and effects of the selected properties of Slovenian verbs on the productivity of individual suffixes

	nja	ava	lišče	ač	telj	nik	lec	nec	nje/tje	All
Ipfv	44 **	27 _**	77 ***	107 ***	57	222 ***	470 ***	211 _***	1144 ***	1382
Acc	45	60 *	49 _.	74	127 ***	276	419 ***	515 ***	1041 ***	2242
Refl	2	1	2	1	2	12	10	5	30	68
Pref	24 _.	52 _.	43	55 _**	78 _**	182 _***	300	415 _***	971	2259
Suffix	6 _*	17	30 _**	33 _***	10 _.	27 _***	19 _*	160	722 ***	1039
Ablaut	0	0	1	6	2 *	3	19 ***	3	21	28
√allom	0	5	1	0	4	19 **	7	11	41 _*	104
Freq		*			***	***	_**			
R^2	.001	.003	.003	.133	.006	.010	.001	.076	.213	
Sample	62	71	83	110	135	383	509	563	1436	
Sem	event	event	place	prtc	prtc	prtc	prtc	prtc	event	

depending on the theme vowel class. Only one suffix, *-nje/tje*, is fully productive, and even this one does not appear to be equally productive with all theme-vowel classes. It is the most productive suffix with all theme vowels except for the i/i and i/e verbs. On the basis of this measure, the suffixes *-nja*, *-ava*, *-lišče*, *-ač*, and *-telj* all seem unproductive, whereas the suffixes *-nik*, *-lec*, and *-nec* are somewhere in-between, at least for certain theme classes.

However, these numbers probably do not tell the whole story. The suffix *-lišče* can be used to derive novel words from novel verbs or from loanwords quite easily, (6). It may be that even though this suffix is productive, we do not find many occurrences in the corpus as the words derived with it are simply not used very frequently (as low frequency is characteristic of productively derived words). Also, in addition to *-lišče* there also exists *-išče* (a suffix taking the verbal root rather than the active participle), as shown in (7), which likely blocks *-lišče*-derivatives where an *-išče*-term already exists.

Table 5.6: Relative quantities of Slovenian verbs undergoing derivations by each suffix per theme vowel class

	nja	ava	lišče	ač	telj	nik	lec	nec	nje/tje
a/a	0,02	0,03	0,05	0,09	0,01	0,08	0,18	0,14	0,85
i/i	0,02	0,01	0,01	0,00	0,12	0,21	0,10	0,26	0,01
a/je	0,00	0,03	0,04	0,03	0,03	0,09	0,53	0,25	0,79
∅/e	0,05	0,00	0,01	0,00	0,02	0,14	0,05	0,09	0,40
i/e	0,01	0,00	0,01	0,00	0,01	0,00	0,00	0,23	0,00
e/i	0,02	0,00	0,01	0,00	0,00	0,16	0,03	0,07	0,41
e/e	0,04	0,00	0,02	0,00	0,02	0,13	0,02	0,04	0,62
a/e	0,02	0,11	0,00	0,00	0,00	0,09	0,13	0,28	0,17
a/i	0,00	0,11	0,03	0,08	0,03	0,17	0,06	0,17	0,61
∅, ne	0,00	0,19	0,00	0,00	0,04	0,11	0,04	0,00	0,44

(6) a. tvitati – tvitališče
'to tweet (on Twitter)' 'location (from) where people tweet'

b. kajtati – kajtališče
'to kitesurf' 'location where people kitesurf'

c. skejtati – skejtališče
'to skateboard' 'skatepark'

(7) a. smučati – smučišče
'to ski' 'a ski resort'

b. igrati – igrišče
'to play' 'playground'

c. križati – križišče
'to cross' 'a crossing'

Again fully parallel to what we did for BCMS, we generated a heat map using R's function *heatmap.2*. The heatmap is given in Figure 5.2.

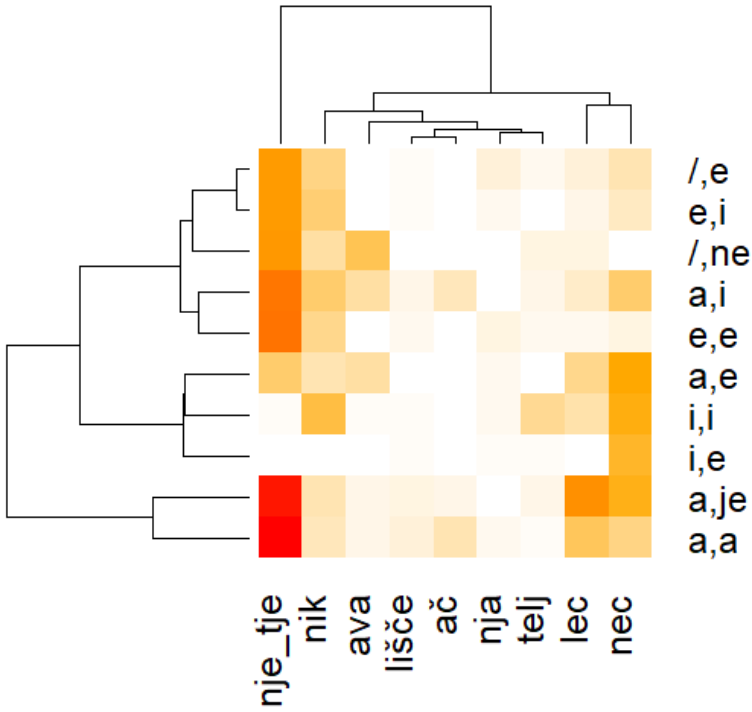

Figure 5.2: The heatmap for Slovenian theme vowel classes

5.6 Discussion, Slovenian

In this section we look at the Slovenian data, repeating the methodology discussed in the previous section. Mimicking the BCMS-focused section, this section is split into comparable sections discussing more or less the same issues in Slovenian.

5.6.1 Theme vowels

Regarding productivity, Slovenian theme vowels too can be divided in three classes, as was the case with BCMS. The most productive class includes themes, a/a and i/i (also highly productive in BCMS), which can be added to a base without the mediation of an additional verbal suffix, as in (8a). Next, there is the theme i/e, which albeit with the help of the suffix *n* (the composite that is elsewhere often treated as a semelfactive theme ni/ne), can also derive new verbs, as in (8b)

(cf. Marušič & Žaucer 2006: 131). All the other theme vowels, like a/je, e/i, ∅/e, e/e, a/i, and ∅/ne, are not productive. They are used on closed classes of verbs, two of which are given in (8c).

(8) a. žamar-i-m tvit-a-m
jumar-TH-PRES.1.SG tweet-TH-PRES.1.SG
'ascend using jumar' 'tweet'

b. blek-n-e-m tvit-n-e-m
blurt-V-TH-PRES.1.SG tweet-V-TH-PRES.1.SG
'blurt out' 'tweet'

c. ber-e-m vzam-e-m
pick-TH-PRES.1.SG take-TH-PRES.1.SG
'pick' 'take'

Looking from the perspective of suffixes, suffix *-nje/tje*, just like in BCMS, sticks out as by far the most productive one, deriving nominalizations from verbs with all theme vowels; see Table 5.4. The three most productive theme vowels overall, i.e., the three theme vowels that derive the most verbs, are also the three most productive theme vowels in the derivation of nominalizations with each of the suffixes. One thing that sticks out from the perspective of suffixes is that the i/i class is the only one that shows no significant effects with any of the suffixes. Also, the suffix *-nja* stands out in being the only suffix not deriving nouns from the verbs of the a/i class, though this class is not a productive class and is actually very rare, so we are dealing with a very small number of verbs to begin with.

Comparing the suffixes pairwise, in terms of sensitivity to theme classes, it is not obvious that we can identify suffixes that form a pair (or group) showing a higher degree of similarity to each other than to other suffixes. The suffixes *-lec* and *-nec* are rather close in their overall frequency, and are rather close in terms of sensitivity to some theme vowels, though with a reversal with class i/i and a/je, *-lec* being considerably more frequent than *-nec* with a/je and *-nec* being considerably more frequent than *-lec* with i/i. Also, unlike *-nec*, the suffix *-lec* is completely absent within the i/e class, but this is not surprising given that the i/e class contains semelfactives (traditionally ni/ne) and *-lec* being known to shun perfective bases (Marvin 2002). Overall, it does not seem to be possible to identify clear pairs (or groups) of suffixes in terms of sensitivity to theme vowels.

5.6.2 Frequency

Just like in the case of BCMS above, we expected that frequency will play a role to the effect that derivations from more frequent verbs will also be more likely to

be frequent in the same corpus. For a subset of the suffixes, this was confirmed, i.e., for the relatively highly frequent suffixes *-nik*, and *-telj*, frequency of the verb showed a positive effect on the number of attested derivations. Among the somewhat less frequent suffixes, the same effect is observed for the suffix *-ava*, though no effect is observed for *-nja*, *-lišče*, and *-ač*. But as a bigger deviation from what we found for BCMS, there is no such effect for the most productive suffix, i.e., *-nje/tje*, and there is a negative effect (higher verb frequency correlating with lower frequency of derivations) for the third most productive suffix, *-lec*. As the sample of the Slovenian *-lec* is bigger than the sample of the BCMS *-lac*, unless an explanation is provided for the behavior of this suffix, this falsifies the above suggestion, in the context of BCMS, that a positive correlation effect might emerge with all suffixes if the sample was larger.

5.6.3 Compatibility with an accusative argument

Verbs that can take an accusative argument are generally more productive in derivation with the suffixes included in the investigation. Positive effects are confirmed for four out of the five most productive suffixes. The fourth most productive suffix, *-nik*, sticks out here in the sense that it does not show such an effect. Interestingly, this lack of effect mirrors the situation in BCMS, where *-nik* was likewise found to not show such an effect. However, the BCMS *-nik* ranks lower in frequency than the Slovenian *-nik*, relatively speaking, so they may not be directly comparable. Moreover, whereas the BCMS *-nik* when deriving from the passive participle most typically denotes the theme or patient of the eventuality denoted by the verb and would as such be expected to correlate with the compatibility with a direct object, the Slovenian *-nik*, when part of a deverbal derivative, most frequently denotes the agent of the eventuality denoted by the verb (Toporišič 2000), and would as such not be expected to correlate with the compatibility with a direct object. In this sense, BCMS *-nik* on passive participle bases is rather equivalent to the Slovenian suffix *-nec*. The possibility for investigating correlations such as this one–between the frequency of a verb's requiring an accusative argument and its serving as a base for derivation with a specific suffix–is clearly a productive way in which *WeSoSlaV* can be used for verifying theoretical claims made about the nature of specific derivational processes.

5.6.4 Meaning shift with the reflexive marking

As noted in the parallel section for BCMS, verbs were annotated as having this property if they can be used both with and without the reflexive clitic and if

this correlates with clearly different meanings (rather than simply with a reflexive versus non-reflexive use. The meaning shift was confirmed by the fact that the same verb that is used with the reflexive clitic, which is ambiguous between marking reflexivity and argument structure effects (unaccusatives, middles, passives, impersonals etc.), could not be used with the strong reflexive pronoun *sebe*, which is restricted to the reflexive meaning. This is illustrated in (4).

(9) a. nasmejati / obleteti
laugh.INF fly_around.textscinf
'make laugh' 'fly around'

b. nasmejati se / obleteti se
laugh.INF REFL fly_around.textscinf REFL
'laugh' 'shed leaves'

Unlike in BCMS, where this factor had a significant positive effect on deverbal derivations for the three least productive suffixes, *-nik*, *-lište* and *-aj*, and for the most productive one, *-nje/će*, we find no such effect in Slovenian. A significant effect is not found for any of the nine suffixes. The discrepancy between BCMS and Slovenian is something for which we currently have no explanation, and which clearly calls for further research on possible interactions between verbal properties and specific derivational operations.

5.6.5 Imperfectivity

Imperfectivity shows a strong effect on deverbal derivations with almost all suffixes. Out of the nine Slovenian suffixes surveyed, only *-telj* shows no effect on this condition. Positive, facilitating effects of a verb's imperfectivity are attested with six out of the nine suffixes: *-nja*, *-lišče*, *-ač*, *-nik*, *-lec* and *-nje/tje*. Unlike in BCMS, these suffixes include *-nik*, which thus comes out as a suffix that shows divergent results between BCMS and Slovenian on several conditions, and may be a good candidate for a comparative study dissecting the functions of this affixal "false friend" in BCMS on the one hand and in Slovenian on the other. Next to the six out of nine suffixes that show strong positive effects of imperfectivity, two suffixes, *-ava* and *-nec*, show a strong negative effect of imperfectivity, i.e., they clearly prefer perfective bases.

Considering the option of further decomposition of our suffixes, it is interesting that *-nec* and *-nje/tje*, which (in deverbal derivations) are both often assumed to contain the same past-participle *-n*, show strong opposite effects on the factor of imperfectivity, and also, as explained below, on the related factor

of prefixation, which is a strongly significantly hindering factor for *-nec* but a non-factor for *-nje/tje*. The obvious difference between *-nec* and *-nje/tje* is that *-nec* denotes participants and *-nje/tje* denotes events, so an explanation for the strong difference on the factor of imperfectivity should likely be searched for in the domain of participant-noun (or specifically, theme/patient-noun) derivation versus event-noun derivation.

5.6.6 Prefixed verbs

Among the nine suffixes in our sample, prefixation did not have a positive effect on any of them, and had a negative effect on six, four of which with a strong negative effect. All four of the latter derive participant nouns, and they include both suffixes that typically derive agent nouns, *-ač*, and suffixes that typically derive theme/patient nouns, *-nec*. At least for some of these, such as *-nik*, the negative effect may stem from the fact that a number of such derivations, even though relatable to a verb from the same root and as such counted here as deverbal, may not actually be deverbal but rather derived from an adjective or sometimes perhaps a root, such as *športnik* 'sportsman', *krožnik* 'plate', etc. More generally, since the prefixed verb sample includes a large number of perfectives, it does not seem surprising that participant-deriving suffixes will tend to be negatively correlated with prefixation given that participant-denoting nouns generally rely on genericity, which, at the level of derivation, is closely related to imperfectivity.

A *prima facie* unusual situation presents itself in the context of *-nec*, however. A verb is considerably less likely to derive a *-nec* noun if it is imperfective, but at the same time, it is also considerably less likely to derive a *-nec* noun if it is prefixed. Given that prefixation is the most common way of deriving perfectivity, this may seem surprising. What is presumably at stake here is that whereas *-nec* prefers a perfective base, and even if prefixation is the main way of deriving perfectives, the prefixed verb sample includes nevertheless about 45% imperfectives (secondary imperfectives), and this then enables the seemingly contradictory negative effect of prefixation on the likelihood of deriving a *-nec* noun.

5.6.7 The base verb is derived by suffixation

The presence of a verbal suffix on the base verb had a significant negative effect for five of the nine deverbal derivational suffixes and a negative tendency for another one, and it also had a significant positive effect for one–the most productive–suffix, *-nje/tje*. Four out of the six suffixes whose affixation is less likely in the presence of another suffix include agent-denoting uses, with which

a negative effect could already be expected just on the basis of competition with the rest of the surveyed suffixes. Quite generally, however, the results of this condition should be interpreted rather carefully, since most of what we coded here as suffixes, such as *-telj*, *-nik*, *-lec*, *-nec*, would typically be decomposed into further subunits, and so while most of our suffixes apparently prefer not to stack over other suffixes, it is not necessarily obvious whether these could be effects on those specific suffix complexes or just on the innermost subunit of those suffixes. To make inferences of this type, comparisons of *-nec* and *-nje/tje*, which presumably both contain the same passive participle *-n*, could be done, or comparisons of *-lišče* and *-lec*, which presumably both contain the same *-l*. For any firm conclusions along these lines, or about suffix combinability, further study is necessary.

5.7 Discussion, comparison between the two languages

After looking at a number of properties of deverbal derivations that can be probed with *WeSoSlaV* in BCMS and Slovenian separately, we now look at the results of the two languages together, contrastively. Some of the contrasts have already been pointed out in the discussion of the Slovenian results, and they will not be repeated in this section. Rather, here we focus on the big picture similarities and differences.

BCMS and Slovenian are close relatives, i.e., they are the two languages constituting the Western South Slavic subgroup. They display a high degree of similarity, but there are also huge differences. These general evaluations extend into the domain of deverbal derivations. An overlap of six suffixes obtains between the sets of suffixes investigated for the two languages: *-ač*, *-lišče/lišće*, *-telj*, *-nik*, *-lec*, and *-nje/će/tje*. These six suffixes roughly derive the same type of nouns in both languages: event nouns for *-nje/će/tje)*, event location for *-lišče/lišće* and agent nouns for *-telj*, *-lec* and *-ač*. The only difference, as already commented is that the suffix *-nik* derives mainly patient deverbal nominalizations in BCMS, and predominantly agent nominalizations in Slovenian. Regarding the variables included, those looked at for BCMS are a proper subset of those considered for Slovenian. Common components thus make the core of the comparison.

Similar groupings obtain between theme vowels in the two languages. The larger, productive themes show one pattern of behavior, the smaller, unproductive ones another. Suffix *-nje/će/tje* showed extremely high productivity in both languages, and in both languages only theme classes deriving predominantly strictly perfective verbs departed from this generalization. In both languages,

theme classes show a stronger grouping by similarity than suffixes – which is expected, as theme classes are a more uniform set. In both languages, theme classes formed two groups with five classes in each. However, BCMS showed a stronger grouping of theme classes, especially pairwise, than Slovenian. Moreover, the groupings in BCMS have stronger correlations with other independent properties (i.e., groupings based on other properties result in similar groups) than in Slovenian.

Looking at other variables, similarities can be observed regarding the argument-structure related properties: compatibility with an accusative argument facilitates derivations with all the suffixes, and the semantic shift with the reflexive marker has none or few significant effects. In both languages, frequency has an overall positive effect, but not in all the cases, and availability of a verbal suffix tends to have a negative effect on the derivational productivity of the verb (each with one exception in Slovenian: *-nec* for frequency and *-nje/tje* for the availability of a verbal suffix).

A striking difference is obtained in the effect of prefixation. While in BCMS, prefixed verbs only show effects of higher productivity in derivation than prefixless verbs, the situation is exactly the inverse in Slovenian. This observation seems to be revealing a difference between the two languages when it comes to prefixation, which may be related to the difference observed regarding (im)perfectivity. Namely, Dickey (2000) observes that perfective verbs are much more freely used in Slovenian than in BCMS (for instance, in Slovenian, they can make present tense performatives, and in BCMS this is impossible).

A small difference can also be observed in the effect of imperfectivity. In BCMS, it facilitates derivations with all but one suffix (*-nik*), while in Slovenian, there are two other suffixes on which it has a negative effect (*-ava* and *-nec*), while the effect on the productivity with the suffix *-nik* is positive. This difference is somewhat relativized by the fact that the semantic equivalent of the BCMS *-nik* in Slovenian is *-nec* (deriving patient nouns) rather than the cognate *-nik* (deriving agent nouns).

Looking at individual suffixes, some further similarities can be observed. Suffix *-lište/lišče* is sensitive to only two of the analyzed factors in both languages, and these are the same: it is facilitated by imperfectivity and hinderd by the availability of a verbal suffix on the base. Suffix *-lac/lec* combines more readily with verbs which can take an accusative argument, are imperfective, and do not have a verbal suffix in both BCMS and Slovenian. Suffix *-nje/će/tje* is facilitated by imperfectivity and the compatibility with an accusative argument in both languages. Finally, in both languages, theme class i/i turns out to be the most productive one in derivations by suffixes *-telj* and *-nik*.

The three suffixes that compete for the derivation of agent nouns, *-ač*, *-telj* and *-lac/lec*, show inverse productivity in the two languages, if productivity is taken as the size of the base, i.e. the number of bases that the suffix combines with. In BCMS, *-ač* is most productive, *-telj* is second, and *-lac* is the least productive, such that *-ač* combines with twice as many bases as *-telj*, and three times as many as *-lac*. In Slovenian, *-ač* is the least productive suffix, *-telj* is second, and *-lec* is the most productive one, such that *-lec* combines with almost four times as many bases as *-telj*, and almost five as many as *-ač*.

Overall, effect sizes measured in terms of R^2 are significantly smaller in Slovenian. This is probably partly due to the smaller overall size of the Slovenian database in *WeSoSlaV* (5300 to 3000 verbs), but may also have to do with the more varying behavior of the included variables regarding their effects on derivation in Slovenian, compared to BCMS.

5.8 Overall discussion

The nature of the reported research is exploratory. The goal is a general exploration of quantitative interactions between the properties of verbs annotated in *WeSoSlaV* and their productivity in a range of morphological operations, as well as more concretely answering the question how small, i.e., how complex niches of productivity can be.

Regarding the latter goal, quantitative analyses indicate that while certain operations are strongly sensitive to a range of investigated properties, domains of (nearly) full productivity are not likely to be found in their intersections. In other words, there are not likely to be morphological regularities hidden from observation because of their high level of complexity.

In terms of general exploration, we have identified a range of effects and interactions which warrant further exploration. At this level, it may have the appearance of a phonebook, as noted by a reviewer of the volume, to whom we are grateful for valuable comments also in this very regard, but we do believe to have opened some new questions which may lead to theoretically relevant insights.

5.9 BCMS suffix *-aj*

Following up on the sections that provided illustrations for ways in which the method presented in the chapter can be used for quick initial, exploratory probing into individual suffixes, or relations between individual suffixes, this section now presents a more in-depth case study investigating a single suffix with the

related general methodological goal of illustrating how the method presented in the chapter can be used to investigate an individual suffix in detail. We chose the BCMS suffix *-aj* primarily because of a problem that it posited for annotation regarding the aspectual status of its base. At first, we assigned all these nominalizations to perfective bases, in agreement with the traditional view of the suffix as deriving perfective nouns (see Klajn 2002, Babić 2002), but we ended up changing this systematically into their imperfective, i.e. atelic variants. Before discussing the examples that led us to this decision, let us first provide some more general quantitative data.

Suffix *-aj* is, to our knowledge, the only one of all the deverbal nominalizing suffixes in BCMS that derives telic event nouns (see the comparison in (10)).

(10)

plivati	tvoriti	žudeti
swim.IPFV	make.IPFV	long.IPFV
pliva-nje	tvor-ba	žud-nja
‘swimming (atelic event)’	‘making (atelic event)’	‘longing (atelic event)’

postaviti	doživeti
install.PFV	experience.PFV
postavlje-nje	doživlj-aj
‘appointment (result)’	‘experience (telic event)’

For this reason, it has been generally assumed that this suffix derives nouns from perfective, i.e. telic verbs, and that their telicity survives derivation and surfaces on the derived nouns. Indeed, a large number of examples morphologically support this view. Consider the examples in (11), in which the derived noun lacks the imperfective suffix of the imperfective candidate for the base, and thus likely derives from the perfective member of the pair. Indeed, the *lj* in *doživljaj* is shared with the imperfective verb, but this may be a more general consequence of the relatively unpredictable capacity of the suffix to trigger iotation of the final consonant of the base – a property observed in a range of suffixes in BCMS. In support of this view, observe that in *izbačaj* the final consonant of the base (*č*) has undergone iotation, without occurring palatalized either in the perfective or in the imperfective variant of the base verb.

(11)

zadrž-a-ti : zadrž-av-a-ti	doživ-e-ti : doživ-lja-vati	izbac-i-ti : izbac-iv-a-ti
keep.PFV:IPFV	experience.PFV:IPFV	throw_out.PFV:IPFV
zadrž-aj	doživlj-aj	izbač-aj
‘retention (telic event)’	‘experience (telic event)’	‘release (telic event)’

The aspectual status of these nouns can be tested by combining them with the preposition *tokom* 'during', which combines well with atelic event nouns, and is ungrammatical with the telic ones.

(12)

tokom plivanja	tokom tvorbe	??tokom treptaja
during swimming	during making	during blink
'during the swimming'	'during the making'	

However, there are certain *-aj* nouns which do not fit this picture as they derive from verbs which do not even have a perfective variant, at least not without prefixes being added (and in many cases no prefix can be added without a meaning change).

(13)

vap-i-ti	jec-a-ti	titr-a-ti
becry.IPFV	weep.IPFV	vibrate.IPFV
vap-aj	jec-aj	titr-aj
'outcry (telic event)'	'weep (telic event)'	'vibration (telic event)'

Moreover, there are cases where the derived noun lacks the suffix which appears in the perfective variant of the base. In such cases too, the *-aj* noun is often additionally phonologically more similar to the imperfective variant of the verb, as with *poticaj* and *uzdisaj* below.

(14)

trg-nu-ti : trz-a-ti	potak-nu-ti : potic-a-ti	uzdah-nu-ti : uzdis-a-ti
tug.PFV:IPFV	stimulate.PFV:IPFV	sigh.PFV:IPFV
trz-aj	potic-aj	uzdis-aj
'tug (telic event)'	'stimulus (telic event)'	'sigh (telic event)'

If the suffix *-aj* universally deletes the verbal suffix of the base (and the preceeding section has showed that it in fact prefers bases involving a suffix), the possibility emerges that deverbal *-aj* nouns derive from imperfective verbs, but in some cases differ from them and look more similar to their perfective counterparts, because the imperfectivizing suffix has been erased. This effect may for instance be the result of a competition between the two suffixes for the same morphological slot.

To establish the correct analysis regarding both the aspectual variants selected by the suffix *-aj* and the issue of deletion of the verbal suffix, we conducted a more detailed quantitative analysis of the deverbal *-aj* BCMS nouns derived from the verbs available in *WeSoSlaV*. We established the quantities of the following classes:

- cases where once the suffix *-aj*, as well as a potential verbal suffix, are deleted, the stem matches the perfective variant of the base verb, and cases where it matches the imperfective one;
- the same two classes, except that the stem arrived at differs from the stem of the verb in being iotated;
- cases where the final consonant of the stem is iotated, but the base verb has a theme vowel which does not iotate;
- cases where a verbal suffix is deleted, and those where a verbal suffix survives, including the particular suffixes involved;
- cases where the base verb has no perfective variant and those where no imperfective variant is in use (imperfective variants can always be productively derived), and finally
- cases where the perfective and the imperfective variant of the base verb differ in terms of the Ablaut of the root (i.e. the root shows a different vowel in the two verbs).

The results are summarized in Table 5.7.

Table 5.7: Quantitative analysis of *-aj* nominalizations

Class	Percentage of cases
Stem possibly imperfective	60.00%
Stem possibly imperfective + iotated	15.45%
Stem possibly perfective	27.27%
Stem possibly perfective + iotated	49.09%
Stem iotated + non-iotating theme	1.82%
No perfective variant	10.91%
No existing imperfective variant	4.55%
Likely deletes a verbal suffix	35.45%
Preserves a verbal suffix	5.45%
Has a prefix	90.00%
Ablaut	19.09%

Twice as many *-aj* nominalizations match the imperfective (60%) as the number of those matching the perfective variant (27.27%) of the base verb. Once iotation of the final consonant of the stem, which is encountered in a subset of *-aj*

nominalizations, is ignored, they reach almost the same quantities, with a slight advantage for the perfective variants (78.18% matching the perfective and 75.45% matching the imperfective). This raises the question of the source of iotation: is it part of the stem, hence iotated stems differ from the respective verbal stems, or is it induced by the derivational process, and should thus indeed be ignored? This possibility appears relevant especially since the iotation often also occurs in the operation of secondary imperfectivization and might therefore be exactly an indication that the imperfective base is selected by the suffix.

Effecting iotation of the final consonant of some stems, and not of others, is a relatively widespread property among BCMS derivational suffixes (see e.g. Janić 2018 for the suffix *-iv*). One possible source of the iotation is the theme vowel of the base verb. It has been shown that themes i/i, ∅/e and e/i, i.e. those involving front vowels, do trigger this phonological operation in derivational processes (Simonović 2022b, see also chapter 3 and the references therein). Indeed, almost all the nominalizations showing iotation of the final consonant of the stem derive from verbs with one of these themes, with only two exceptions, *gutljaj* from *gutati* 'swallow' and *zavežljaj* from *zavezati* / *zavezivati* 'tie.PERF/IMPERF'.

The comparison by the overall match therefore supports the view where *-aj* selects for imperfective verbs. There are, however, other properties of these derivations that may be informative regarding this question.

Almost 11% of the verbs deriving *-aj* nominalizations have no perfective counterpart whatsoever, which supports the view that they are derived from imperfective variants. However, close to 5% of the nominalizations derive from verbs without an imperfective variant, indicating that either *-aj* may select either variant (in a predictable or unpredictable way), or it selects neither and the surface form is regulated by an independent mechanism operating at the same level as those deriving the realization of the perfective and imperfective variant of the verb.

In a number of verbs, the imperfective and the perfective variant of the verb differ in the realization of the vowel in the final syllable of the stem in the way of phonological alternation known as Ablaut, or apophony. Among the verbs deriving *-aj* nouns, 19.09% are of this type. In all cases but one (21 verbs, 10 different roots), the realization in the *-aj* nominalization fits the imperfective, as in (15a). The one case in the opposite direction actually involves one of the ten roots above, hence this root manifests both directions, as in (15b). This comparison again favors an analysis on which *-aj* selects imperfective bases, with the one exceptional case showing the opposite pattern for an idiosyncratic reason (possibly influenced by the adjective *ugodan* 'pleasant' from the same verb).

(15) a. u-tek-ti : u-tik-a-ti do-god-i-ti : do-gad-i-a-ti
flow_in.PFV:IPFV happen.PFV:IPFV
u-tik-aj /uticaj/ do-gad-i-aj /događaj/
'influence' 'event'

b. u-god-i-ti : u-gad-i-a-ti
suit.PFV:IPFV
u-god-i-aj /ugođaj/
'ambience'

On the assumption that *-aj* derivations target imperfective verbs, 35.45% of the derivations involve deletion of an imperfectivizing verbal suffix. However, 5.45% of them surface with an imperfectivizing suffix. Of all the verbal suffixes, only one has been attested surviving in *-je* nominalizations: the frequentative, i.e., de-semelfactive imperfectivizing suffix *-ta* (cf. trep-n-u-ti 'to blink.PFV' vs. trep-t-a-ti 'to blink.IPFV'). It is one of the less frequent suffixes, monosyllabic (together with its competitor *-ka* with the equivalent semantics and the very infrequent *-sa*, as well as the much more frequent perfective semelfactive *-nu*), compatible with two themes (a/a and a/je, where all the attested *-aj* nominalizations come from verbs taking the latter; no other verbal suffix in BCMS is compatible with two themes).

Further investigation is needed to establish whether any of these properties are relevant for the (avoidance of) suffix deletion. What can certainly be concluded is that those verbs whose bases surface with both a verbal suffix and the suffix *-aj* cannot be analyzed as deriving from perfectives. If the suffix *-aj* selects one aspectual class, then this speaks in favor of an analysis where it selects imperfectives. One possible generalization is that if the verb suffix *-aj* combines with has a verbal suffix, the suffix can only survive if it realizes semelfactivity without imposing perfectivity. Plausibly, suffix *-aj* by default realizes both the semantics of semelfactivity and the nominal category, yet it can also nominalize if semelfactivity is independently realized (by virtue of a subset or superset principle, depending on the approach, Caha 2018). As suffix *-nu* expresses semelfactivity in perfective verbs, a view that does not assume that the suffix selects imperfective verbs would have a problem explaining why *-nu* never surfaces in deverbal *-aj* nominalizations.

While generally in *WeSoSlaV*, 69.85% of verbs are prefixed, among those deriving *-aj* nominalizations, this percentage goes to 90%, an affinity whose significance was confirmed by the statistical test (see table 5.2). As indicated there, suffix *-aj*, which derives telic event nouns, plausibly needs or at least prefers that

the base includes a telic event predicate (which may be embedded in an atelic one).

In conclusion, it follows from the discussion that the analysis where suffix *-aj* selects imperfective verbs is somewhat superior to the one assuming perfective bases, even though it is also not without issues. In particular, a more sophisticated analysis is required to capture the deletion of certain suffixes, and the preservation of others, the semelfactive (hence telic) semantics of the nominalizations (unexpected with atelic bases, Arsenijević 2023), and the nominalizations derived from verbs without an imperfective variant. To reconcile all these issues, an analysis in terms of realizational morphology is advantageous, as it can capture the particular realizations in terms of structure-induced allomorphy, without seeing them as necessarily originating either from a perfective or from an imperfective verb.

5.10 Comparing results obtained by two different annotation methods

The method used in the investigation presented in this chapter (in section 5.2) differs from the method used to build the rest of *WeSoSlaV* (presented in Chapter 2). In *WeSoSlaV* generally, we relied on annotator judgments, while in the research reported in this chapter and the additional annotation it involves, we annotated verbs for combining with a suffix based on corpus evidence (i.e., if an attestation was found in the corpus with a frequency higher than a certain threshold). Considering that judgment- and corpus-based data have been observed to not always fully align (e.g. Schütze 1996), a question that this difference raises is how similar the results of the two methods are, i.e., what differences they display. The answer to this question should be informative regarding the overall properties of *WeSoSlaV* as well as regarding its possible combination with other methods in further investigations. A difference that one would certainly expect is that judgment-based annotation will result in greater subsets of the verbs from *WeSoSlaV* being annotated as undergoing the relevant derivations, since not all derivations are sufficiently frequent in use to necessarily enter the corpus. The highest relevance lies in the sub-question of whether both methods show the same effects when the investigated factors are concerned, at least as long as the factors concern grammatical and semantic phenomena. Pragmatic, world-knowledge-related factors are more likely to display different patterns, as they play a bigger role in the corpus than in annotators' decisions, and there may even be a qualitative difference in these roles. The comparison may thus also be

informative with respect to the goal of disentangling formal from usage-related properties of verbs as bases for derivation.

For this purpose, we did the following. First, we took two suffixes, the BCMS *-nje/će* and its Slovenian counterpart *-nje/tje*, and the BCMS *-lac* and its Slovenian counterpart *-lec*, and annotated *WeSoSlaV* for speakers' judgments regarding the existence of nouns derived with these suffixes from the verbs in the database. These suffixes were selected because they stand in an opposition that is interesting for their morphological analyses, namely *nje/će/tje* derive from the passive and *lac/lec* from the active participle of the verb; at the same time these suffixes display the highest overall productivity in the corpus based investigation. The annotation procedure was the same as for the rest of *WeSoSlaV*, except for one additional step. It turned out to be quite difficult to think about various narrowed, metaphorical, metonymic, or other non-prototypical contexts and meanings that may license the derivation of a noun. To avoid false negatives, when the annotator found that the derived noun sounds possible, but not very salient – the annotators checked the corpus for examples. They then identified those examples that do plausibly represent the intended derived noun, and judged it as acceptable or not. Hence, corpus data were not used as a criterion, but as an enrichment of the data that were judged.

To test the comparative properties and mutual relations of the two annotation methods, we have therefore compared the statistical analyses of effects of the same properties considered in the rest of the chapter. To simplify the picture, we included only those factors that showed a considerable number of effects in the analysis of corpus-annotated data: the theme vowel, imperfectivity, prefixation and suffixation. Three possible outcomes of the comparison are foreseen: attesting identical effects, attesting in the analysis of judgment-base annotation a superset of the effects attested on the corpus based data, and attesting a set of effects such that both differences are non-empty: there are effects attested in corpus-annotation data and not on judgment-annotation data, as well as vice versa.

As we aimed to maximize the differences, we first fitted models with different levels of the variable theme vowel as the base line, and then chose the one that maximized the difference, which turned out to be ∅/e in both languages. Considering this, and that the models were fit for a different set of factors, the effects reported for the corpus-annotated data are different from those reported in sections 5.3 for BCMS, i.e., 5.5 for Slovenian.

As we can see from table 5.8, the effects captured on the corpus annotation are a proper subset of those captured on judgment annotation, except for the effect of suffixation on *-lac* suffixation. While the presence of a suffix shows a strong

negative effect in the corpus-annotated data, it has an even stronger positive effect in the annotation based on judgments. It is important to note here that we established a higher than usual disagreement between annotators with respect to suffixed verbs, where one tended to rejected them for a majority of suffixes, and the other accepted a nouns derived by the suffix *-lac* for all suffixes, pointing out that suffixed verbs as bases for derivation by *-lac* require unlikely, but possible contexts. After the regular procedure of negotiating disagreeing annotations and consulting a third annotator, a majority of such derivations were accepted. The discrepancy between the two annotations thus probably reflects the fact that the corpus contains data mainly from frequent, typical contexts, while the annotators could construct also the unlikely, hypothetical ones. The insight to be further investigated is that suffixation of the base verb is a factor of deverbal derivation which interacts with the likeliness of the context.

In the Slovenian dataset in 5.9, no subset relation can be established between the two methods of annotation. For both suffixes, there are effects captured under one, but not under the other, as well as vice versa. Still, for Slovenian too, there is only one effect which goes in opposite directions for the two methods of annotation. In the Slovenian dataset, this is the effect of the theme a/a on the derivation by the suffix *-nje/tje*. In the corpus-annotated dataset, this theme has a negative, and in the judgment-based one positive effect on this derivation. We were unable to explain this difference by any specific property of this theme or the suffix. Both highlighted differences, as well as that from BCMS, indicate that there are stronger underlying differences between the two methods of annotation.

This raised the question of whether the differences in the two methods of annotation are indeed sensitive to the factors included in the analysis, i.e. whether e.g. when verbs are imperfective, the tendency to accept more derivations based on the judgments becomes even stronger, or perhaps weakens. To provide just the most superficial *yes-no* answer to this question, we did the following. We added an additional variable for each suffix (*je_dif, lc_dif*) in both languages, in which we coded the relation between the two methods for each observation. When they were the same, we coded *0*, when the judgment-based annotation accepted the derived noun, but the corpus did not confirm it, we annotated *1*, and when the corpus had it, but the judgment was negative *–1*. We then fitted a generalized mixed model for each of these variables as the dependent and the same set of factors above. The effects are summarized in tables 5.10 and 5.11 for BCMS and 5.12 and 5.13 for Slovenian.

Obviously, in both languages, effects are found – in fact, the only case where a variable shows no effect is suffixation in *-lec* derivations in Slovenian. Partic-

Table 5.8: Comparison of raw quantities and effects for the two methods of annotation for BCMS

TV	lac_judg	lac_corp	nje_će_judg	nje_će_corp
a/a	970 ***	25	1423 ***	1114 ***
i/i	863 ***	72 *	1184 ***	576 ***
a/je	610 ***	11	909 ***	780 ***
∅/e	12 bs-ln	2 bs-ln	71 bs-ln	19 bs-ln
u/e	0	1	152 ***	21
e/i	15	1 bs-ln	98 ***	4
∅/ne	0	0	60 ***	9 *
a/i	17 **	2	44 ***	28 ***
a/e	6 ***	0	11 **	8 ***
e/e	1	0	8	3
Defective	0	0	3 _***	1
Ipfv	1884 ***	76 ***	2724 ***	2130 ***
Pref	1541 ***	60	2419 ***	1442
Suff	956 ***	12 _**	1514	783
R^2	−.28	.76	.225	.443
Sem	prtc	prtc	event	event

Table 5.9: Comparison of raw quantities and effects for the two methods of annotation for Slovenian

TV	lec_judg	lec_corp	nje_tje_judg	nje_tje_corp
a/a	309 **	191	1003 ***	891 _*
i/i	140 ***	90	619	5 _***
a/je	130 ***	199 ***	352 *	299
∅/e	7 bs-ln	15 bs-ln	200 bs-ln	114 bs-ln
i//e	0	0	86 _*	0
e/i	5	4 _*	89	53 _**
e/e	1	1	38	29 .
a/e	8 ***	6 *	33	8 _**
a/i	6 .	2	29	22
∅/ne	1	1	20	12 –.
Defective	0	0	1 _**	3
Ipfv	650 ***	211 ***	1348 ***	1144 ***
Pref	457 ***	415 .	1784	971
Suffix	367 _**	160 _**	832 .	722 ***
R^2	.478	.548	.47	.658
Sem	prtc	prtc	event	event

ularly interesting are the variables that show a negative difference, i.e. which include more cases where annotator judgments rejected derivations found in the corpus, such as with imperfectivity, prefixation and suffixation in Slovenian *-je* derivations.

While we can speculate that the observed effects show either pragmatic effects

of certain variable levels requiring less natural contexts, or possibly also strategies unconsciously adopted by the annotators, these issues require thorough further analysis in order to provide more accurate answers. The point of our discussion was mainly to point them out, and thus indicate a possible use of *WeSoSlaV* in providing fine empirically based answers to methodological questions, such as that of the differences between judgment and corpus-based annotation.

Table 5.10: Effects of the analyzed factors on differences in annotation of *-lc* derivations in BCMS

BCMS-lc	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-0.15275	0.02834	-5.390	7.35e-08	***
∅/ne	-0.13767	0.04821	-2.855	0.00431	**
a/a	0.19308	0.02839	6.801	1.16e-11	***
a/e	0.23280	0.09948	2.340	0.01931	*
a/i	0.06451	0.05893	1.095	0.27372	
a/je	0.13273	0.03097	4.286	1.85e-05	***
Defective	-0.32326	0.17387	-1.859	0.06305	.
e/e	-0.05190	0.10485	-0.495	0.62058	
e/i	-0.03418	0.03952	-0.865	0.38716	
i/i	0.40375	0.02661	15.174	<2e-16	***
u/e	-0.12166	0.03983	-3.055	0.00226	**
Ipfv	0.47601	0.01746	27.261	<2e-16	***
Pref	0.14994	0.01543	9.716	<2e-16	***
Suffix	0.15863	0.01703	9.312	<2e-16	***

Wrapping up the section, our assessment of the two different methods of annotation, one based on judgments and the other on corpus search, showed that they tend to give highly congruent results, but in rare cases also contradictory ones. Their differences are effected by all four variables included in the analysis in the section: theme vowels, imperfectivity, prefixation and suffixation. This opens possibilities for the exploration of methodological questions central for the research of language, while at the same time warning that generalizations we make about language, including those based on *WeSoSlaV*, should be factored by the empirical method of acquiring the empirical insights.

Table 5.11: Effects of the analyzed factors on differences in annotation of *-je* derivations in BCMS

BCMS-je	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.02683	0.02766	0.970	0.332182	
∅/ne	0.12828	0.04706	2.726	0.006440	**
a/a	-0.52292	0.02771	-18.868	<2e-16	***
a/e	-0.36200	0.09710	-3.728	0.000195	***
a/i	-0.52469	0.05752	-9.121	<2e-16	***
a/je	-0.55452	0.03023	-18.343	<2e-16	***
Defective	-0.41627	0.16972	-2.453	0.014209	*
e/e	0.06915	0.10234	0.676	0.499256	
e/i	0.21443	0.03858	5.558	2.86e-08	***
i/i	0.17087	0.02597	6.579	5.20e-11	***
u/e	0.30550	0.03888	7.858	4.69e-15	***
Ipfv	0.38945	0.01704	22.849	<2e-16	***
Pref	0.10150	0.01506	6.738	1.77e-11	***
Suffix	0.07019	0.01663	4.221	2.47e-05	***
	-0.14144	0.02452	-5.768	8.84e-09	***

5.11 Conclusion

The main goal of the chapter was to show how *WeSoSlaV* can be used to investigate various deverbal derivational operations and to compare the two Slavic languages along these dimensions. In the relatively superficial, purely exploratory overview, we have established that most of the properties selected from *WeSoSlaV* for this purpose do show an effect in some of the suffixes, and that some of the properties tend to have effects in the same direction for all suffixes (e.g. frequency and the ability to take accusative arguments), while others show different effects across different suffixes (e.g. imperfectivity and prefixation). It is to be expected that properties not included in this overview, such as the prosodic patterns, or the availability of a second, or third prefix, would also be found to behave in one of these two ways. Moreover, it was shown that theme vowel classes behave differently, and show some systematicity, which indicates that they do not take arbitrary assemblies of verbs, but classes with common underlying properties. Finally, we did not calculate interactions of the included properties, which is another possible source of insights into the nature of verbal vocabulary.

Another possible avenue for further research would be to modify our method

Table 5.12: Effects of the analyzed factors on differences in annotation of *-lc* derivations in Slovenian

Slo-lc	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-0.094530	0.022381	-4.224	2.48e-05	***
∅,ne	0.037062	0.060168	0.616	0.537960	
a/a	0.068270	0.021655	3.153	0.001634	**
a/e	0.065532	0.045626	1.436	0.151031	
a/i	0.115624	0.050957	2.269	0.023335	*
a/je	0.091736	0.025561	3.589	0.000337	***
Defective	-0.364568	0.167056	-2.182	0.029164	*
e/e	0.018057	0.045636	0.396	0.692375	
e/i	0.018314	0.030791	0.595	0.552027	
e/i	-0.031234	0.287779	-0.109	0.913578	
i/i	0.075905	0.019714	3.850	0.000120	***
ni/ne	0.048339	0.034498	1.401	0.161251	
Imperf	0.125764	0.017353	7.247	5.39e-13	***
Prefix	0.064483	0.015478	4.166	3.19e-05	***
Suffix	-0.009642	0.017088	-0.564	0.572612	

in the following way. In the process of preparation, as already specified, we removed from the database all the legitimate deverbal derivations whose bases were verbs not included in *WeSoSlaV*. This amounted to 5-10% of the final sets of deverbal derivations. If instead, the respective verbs had been added to the database, and annotated like the verbs already in it, this would have somewhat expanded the sample sizes, and potentially led the observed tendencies to become significant effects.

We hope that this chapter provided a clear and insightful illustration, with just a selection from the available deverbal derivations, and with a relatively superficial overview without diving into individual suffixes, of the great opportunities that *WeSoSlaV* offers for a quantitative exploration of the properties of verbs in the perspective of deverbal derivations, as well as of the actual morphological structures and operations underlying these derivations. The few proposed answers, and especially numerous new questions that have emerged from this investigation, additionally support this conclusion.

The chapter has also provided some theoretically relevant empirical insights regarding the effects of various verbal properties on their capacity to serve as bases for further morphological derivations. Similarities and differences between

Table 5.13: Effects of the analyzed factors on differences in annotation of *-je* derivations in Slovenian

Slo-je	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.37493	0.03212	11.674	<2e-16	***
∅,ne	0.12502	0.08634	1.448	0.147722	
a/a	-0.02481	0.03107	-0.798	0.424793	
a/e	0.24627	0.06547	3.761	0.000172	***
a/i	-0.07129	0.07312	-0.975	0.329656	
a/je	0.03862	0.03668	1.053	0.292457	
Defective	-0.86827	0.23972	-3.622	0.000297	***
e/e	-0.05091	0.06549	-0.777	0.437027	
e/i	0.01788	0.04418	0.405	0.685747	
e/i	-1.20160	0.41296	-2.910	0.003644	**
i/i	0.42388	0.02829	14.984	<2e-16	***
ni/ne	0.42736	0.04950	8.633	<2e-16	***
Imperf	-0.17333	0.02490	-6.961	4.15e-12	***
Prefix	-0.06339	0.02221	-2.854	0.004349	**
Suffix	-0.14144	0.02452	-5.768	8.84e-09	***

theme vowel classes, but also individual derivational suffixes have been highlighted, indications provided about the productivity of the latter, and comparisons given between the two Western South Slavic languages.

6 Multifunctional suffixes

6.1 Introduction

This chapter focuses on multifunctional suffixes, i.e., suffixes that appear in different, (prima facie) unrelated contexts, often in different categories, and have little meaning common to all these contexts. We base the discussion on the columns in *WeSoSlaV* that mark verbal and multifunctional suffixes: the Verbal Suffix group of columns, the column Multifunctional Suffixes and the column Suffix Like. The discussion will occasionally extend to items annotated in other columns which meet some of the criteria of multifunctionality, but also have a clearly identifiable role in the verbal domain.

This chapter takes a somewhat different approach from the preceding ones. Rather than offering extensive quantitative analysis, it provides a detailed overview of specific cases of multifunctional elements. By doing so, we demonstrate that multifunctional suffixes are not rare or exceptional in language but instead form a crucial part of linguistic systems. Similarly to the other chapters, this chapter addresses the challenges involved in describing and analyzing multifunctional affixes, both in general and in relation to specific elements of Western South Slavic. Additionally, it discusses the rationale behind the chosen annotation solutions and details their implementation, so that this annotation can serve as a blueprint for annotating other multifunctional elements in the targeted systems and in related languages.

The remainder of the chapter is organized as follows. In 6.2 we discuss the definition of multifunctionality and the main theoretical questions that multifunctionality raises, whereas in 6.3 we discuss the annotation of multifunctional elements in *WeSoSlaV*. Finally, 6.4 brings three case studies which focus on specific multifunctional elements.

6.2 Definition

Multifunctional suffixes, sometimes also called multipurpose, multicategorial or polyfunctional suffixes, are suffixes that can appear in different, seemingly un-

related contexts and have little meaning of their own. Such affixes are traditionally often treated as separate, unrelated, accidentally homophonous items, see for example the many instances of *-ov* in Toporišič (2004). While multifunctionality is typically considered in the context of derivational affixes, e.g., De Belder (2011), Lowenstamm (2014), Creemers et al. (2018), some seemingly multifunctional derivational suffixes can also appear in inflection. Simonović & Arsenijević (2020) argue that the latter instantiate *maximal* multifunctionality. Furthermore, while some multifunctional affixes can be categorially flexible, others serve different functions only within one category, see, e.g., Salvadori & Huyghe (2023).

The full scope of multifunctionality can be observed if we consider the Slovenian suffix *-ov*, first approached from the perspective of multifunctionality in Simonović & Mišmaš (2020). This item is multifunctional as it appears in different categories – it can be found in verbs, adjectives and nouns, but it is also multifunctional within specific categories. In adjectives it can derive kind, (1), or possessive adjectives, (2a), though the possessive function is unavailable for feminine-gender nouns, (2b). In verbs, *-ov* is, among other uses, employed as an imperfectivizing suffix, (3), and in deriving denominal verbs, (4). In nouns it can act as an obligatory root extension in the sense of Acquaviva (2008), as in (5), and even serve as an inflectional ending, (6).

(1) a. bezeg-∅, bezg-ov (sok) (Slo)
elder-flower-M.SG.NOM elder-flower-*ov*.ADJ (juice)
'elder-flower' 'elder-flower (juice)'

b. jagod-a, jagod-ov (sok)
strawberry-F.SG.NOM strawberry-*ov*.ADJ (juice)
'strawberry' 'strawberry (juice)'

(2) a. brat-∅, brat-ov (sok) (Slo)
brother-M.SG.NOM brother-*ov*.ADJ (juice)
'brother' 'brother's (juice)'

b. sestr-a, *sestr-ov, sestr-in (sok)
sister-F.SG.NOM sister-*ov*.ADJ sister-*in*.ADJ (juice)
'sister' 'sister's (juice)'

(3) del-a-ti^{IPFV}, pre-del-a-ti^{PFV}, pre-del-ov-a-ti^{IPFV} (Slo)
work-TV-INF re-work-TV-INF re-work-*ov*-TV-INF
'to work' 'to re-work' 'to re-work'

(4) kmet-∅, kmet-ov-a-ti$^{\text{IPFV}}$ (Slo)
farmer-M.SG.NOM farmer-*ov*-TV-INF
'farmer' 'to farm'

(5) bank-a, *bank-ec, bank-ov-ec (Slo)
bank-F.SG.NOM, bank-*ec*, bank-*ov*-*ec*
'bank' 'banknote'

(6) hrib-∅, hrib-i, hrib-ov (Slo)
hill-M.SG.NOM hill-M.PL.NOM hill-*ov*.M.PL.GEN
'hill' 'hills' 'hills'

As mentioned above, traditional grammars (e.g., Toporišič 2004) treat cases like this as different, accidentally homophonous suffixes or parts of suffixes. This approach assumes the existence of a verbal *-ov*, an adjectival *-ov*, and a nominal *-ov* (potentially with multiple instances per category). A challenge for this approach lies in the fact that even though some of the functions of *-ov* are clearly definable, e.g., the possessive in (2a), many instances of *-ov* cannot be assigned a clear meaning, which means that we would need to assume several separate semantically empty/light suffixes with the same phonological form. It therefore seems more sensible to adopt the view that we are dealing with a single semantically empty or light item.[1]

From a formal theoretical perspective, the existence of multifunctional items across categories has significant implications for the treatment of derivational suffixes, particularly in syntax-centric approaches to morphology, such as Distributed Morphology (Halle & Marantz 1993). In "classic" Distributed Morphology, derivational affixes are viewed as exponents of categorial heads, see, e.g., Lowenstamm (2014) and Simonović (2022a) for discussion. However, this view has been challenged by various empirical observations. First, derivational affixes

[1] Simonović & Mišmaš (2020) argue that *-ov* is semantically empty on the basis of the use of *-ov* with other derivational suffixes. Consider the minimal pairs of *je*-nominalizations and *en*-adjectives, taken from Simonović & Mišmaš (2020: 88), in (i), in which the two items have slightly different meanings but there is no clear (or consistent) contribution of *-ov*.

(i) a. grm, grm-ov-je; list, list-je (Slo)
bush bush-*ov*-*je* leaf, leaf-*je*
'bush' 'shrubbery' 'a leaf' 'leaves/foliage'

b. jezik, jezik-ov-en, jezič-en
tongue/language tongue/language-*ov*-*en*.ADJ tongue/language-*en*.ADJ
'tongue/language' 'relatedto language' 'related to tongue'

do not just determine the category of the root without contributing anything to the meaning (i.e., there are no real pure categorizers, see Acedo-Matellán & Real-Puigdollers 2019, De Belder 2011, Lowenstamm 2014). Second, as illustrated above for Slovenian *-ov*, some affixes appear in different categories. Detailed descriptions of multifunctional suffixes in different categories can be found in De Belder (2011) for Dutch and Lowenstamm 2014: 233 for English. An example from English would be *-ory*, that can appear in nouns, such as *laboratory*, or adjectives, such as *accusatory*, or in items that can be either nouns or adjectives, such as *advisory*. In the latter case the affix also has different additional meanings, so fully matching the definition of multifunctional affixes. As for an analysis of multifunctional affixes in Western South Slavic, Simonović & Arsenijević (2020) discuss *-je* in BCMS, whereas Mišmaš et al. (2020) and Arsenijević (2020) bring a more general discussion of *-je* in Slavic. Simonović & Arsenijević (2020) discuss *-at*, *-en*, *-an* in BCMS, whereas Simonović & Kovačević (2022) present an analysis of the different adjectival instances of *-ov* in Serbo-Croatian. For Slovenian, Simonović (2022a) analyzes Slovenian *-av* and *-ov* across categories, Simonović et al. (2024) discuss the multifunctional suffix *-l* and Simonović & Mišmaš (2022) discuss the multifunctional suffix *-e*.

Identifying, describing and analyzing multifunctional elements in all contexts of their use is a pivotal step toward an adequate theory of derivational affixes. Here the accounts in the literature vary from determining their status in the syntactic structure (head vs. non-head in Gouskova & Bobaljik 2022), to positioning them in the functional hierarchy (Cinque 2015), to arguing that all derivational affixes are roots (Lowenstamm 2014) or that some derivational affixes are roots (Creemers et al. 2018). Much of the research based on the *WeSoSlaV* database so far has taken as it starting point the assumption that multifunctional affixes are roots, see, for example, Simonović (2020), Simonović & Mišmaš (2020). Independently of the selected approach, the *WeSoSlaV* database provides an initial empirical overview of potentially multifunctional items in the verbal domain in WSS. This overview can be further enriched with additional annotations.

6.3 Treatment of multifunctional affixes in *WeSoSlaV*

Verbs that feature multifunctional items are marked with "1" in the columns Multifunctional Suffixes and Verbal Suffix. The column Multifunctional Suffixes is reserved for suffixes that do not not appear in any verbal function, potentially originate from a different category and/or have a vague meaning. The suffix itself is then annotated in the Root column, separated from the root

with a plus sign, see Table 6.1 (all examples in the table are from Slovenian). The column Verbal Suffix is used for marking those suffixes that have a verbalizing, aspectual or verb-modifying function and is followed with several columns that list individual verbal suffixes and the typical theme vowel they appear with (e.g, *-av+a*, *-ir+a* etc.). If a verb has a specific verbal suffix, this was marked with a 1 in the column with the relevant suffix. Since some of these verbal suffixes can also be used in other functions, they should also be included when identifying multifunctional suffixes, as shown with the suffix *-av* in Table 6.1 and discussed below in reference to examples (7), (8) and (9).

Table 6.1: Multifunctionality in *WeSoSlaW*

Verb	Root	Multifunctional suffixes	Verbal suffix	av-a	Translation
sanjariti^{IPFV}	sanj+ar	1	0	0	'to dream'
priznati^{PFV}	zn	0	0	0	'to admit'
priznavati^{IPFV}	zn	0	1	1	'to admit'
krvaveti^{IPFV}	krv+av	1	0	0	'to bleed'

As evident from Table 6.1, some affixes are annotated as verbal for some verbs and as multifunctional (but never both for the same verb), see also *-ov* in Section 6.4.1. Note also, that in the table (as well as *WeSoSlaV*) *-av* is listed as a verbal suffix together with the theme vowel it typically appears with. The decision which function a suffix has, and hence whether it will be annotated as a multifunctional or a verbal suffix, was made on a case by case basis. In annotation we considered the function of the affix in the verb (is it, e.g., aspectual), its presence in the base of the verb (if the verb is, for example, denominal) and theme vowel selection.

Consider the suffix *-av* in Slovenian. In some verbs, the suffix participates in secondary imperfectivization – it combines with a perfective prefixed base and renders it imperfective, (7), see Simonović (2022a), Simonović & Mišmaš (2023a) for an analysis of this particular suffix.[2] In other verbs, *-av* seems to be part of the nominal, (8), or the adjectival base (9). Consequently, due to its aspectual function, *-av* is marked as a verbal suffix in the case of secondary imperfectives, e.g., *priznavati*^{IPFV} 'to admit' in Table 6.1 and example (7), and as a multifunctional suffix in the other two examples (as also shown with *krvaveti* 'to bleed'

[2]The literature on the topic of secondary imperfectives in Slavic is extensive, see, for example, Biskup (2024) and Biskup (2023) for a recent work on Russian and Czech, respectively, or Klimek-Jankowska & Błaszczak (2022) for Polish.

in Table 6.1), since it here originates from a different category. The difference is also reflected in the glosses in examples (7) vs. (8) and (9). The two functions of *-av* are also distinguished by the theme vowel selection: in secondary imperfectives *-av* consistently appears with the theme vowel a/a, (7), while *-av* from nominal bases shows up with the theme vowel i/i, (8), and *-av* from adjectival bases appears with the theme vowel e/i, (9).

(7) zn-a-ti$^{\text{IPFV}}$, pri-zn-a-ti$^{\text{PFV}}$, pri-zn-av-a-ti$^{\text{IPFV}}$ (Slo)
know-TV-INF, at-know-TV-INF, at-know-SI-TV-INF
'to know' 'to admit' 'to admit'

(8) diš-av-a, diš-av-i-ti$^{\text{IPFV}}$ (Slo)
breath-*av*.F.SG.NOM breath-*av*-TV-INF
'perfume' 'to scent'

(9) krv-av, krv-av-e-ti$^{\text{IPFV}}$ (Slo)
blood-*av*.ADJ blood-AV-TV-INF
'bloody' 'to bleed'

Regarding suffixes with no purely verbal function, consider the Slovenian suffix *-ač*. This suffix can appear in (agentive) nouns, (10), and (less frequently) in adjectives, (11). Both can serve as the base for verbalization. Since the suffix *-ač* is itself never used as a verbalizing, aspectual or verb-modifying suffix, and since (10) and (11) show the only two contexts in which *-ač* appears in *WeSoSlaV*, the suffix is only annotated as a multifunctional suffix in the database.

(10) ber-e-mo, ber-ač, ber-ač-i-ti$^{\text{IPFV}}$ (Slo)
collect-TV-1PL, collect-*ač*, collect-*ač*-TV-INF
'we collect' 'beggar' 'to beg'

(11) dom, dom-ač, u-dom-ač-i-ti$^{\text{IPFV}}$ (Slo)
home, home-*ač*.ADJ, in-home-*ač*-TV-INF
'home' 'domestic' 'to domesticate'

Finally, it should be noted that for the purposes of the database, the deciding factor in annotating an item as multifunctional was its origin in a non-verbal category. Consider BCMS suffix *-išt*, (12), or its Slovenian equivalent *-išč*, (13). These suffixes are never annotated as verbal, as they never derive verbs. They rather derive nouns, and their meaning contribution is typically one of location. Some of these nouns can then get verbalized (the verbal morphology is applied

only after the nominal *-išč/-išt* had already been introduced into the structure). Despite this single meaning/function, the suffix *-išč/-išt* was annotated as multifunctional. It is worth noting that *-išč/-išt*-nominalizations can themselves be deverbal, as *-išč/-išt* can select a verbal base (L/past participle, see Simonović et al. 2024 for an alternative analysis). To the best of our knowledge, (15)) is the only relatively well-established example where a deverbal *-išč*-nominalization gets verbalized.

(12) sklad-išt-e, sklad-išt-i-ti (BCMS)
stack-*išt*-N.SG.NOM, stor-*išt*-TV-INF
'storage' 'to store'

(13) sklad-išč-e, sklad-išč-i-ti (Slo)
stack-*išč*-N.SG.NOM, store-*išč*-TV-INF
'storage' to'store'

(14) gled-a-ti, gled-a-l-išč-e (Slo)
look-TV-INF, look-TV-*l*-*išč*-N.SG.NOM
'to look' 'theater'

(15) gled-a-l-išč-e, u-gled-a-l-išč-i-ti (Slo)
look-TV-*l*-*išč*-N.SG.NOM, in-look-TV-*l*-*išč*-TV-INF
'theater' 'to adapt for theater'

This means that suffixes present in verbalized bases were annotated as multifunctional in the *WeSoSlaV* database even if they had a stable meaning and were only associated with a single category.[3]

We end this section with lists of multifunctional suffixes in BCMS and Slovenian. While these two lists are most probably not exhaustive, they cover the multifunctional items found in the verbal domain based on the verbs included in the *WeSoSlaV* database. Suffixes are listed for each sub-base in Table 6.2 and Table 6.4. As a comparison of the two tables will make apparent, we identified different multifunctional suffixes in the two sub-bases and found more in the BCMS sub-base. There are several potential reasons for this. One is simply that

[3] As a reviewer appropriately notes, this leads to a paradoxical situation in which some items annotated as multifunctional, in fact, exhibit only one function. We discuss one such example in subsection 6.4.3. Nonetheless, we maintain this terminology, following the initial decision that the determining factor for annotating an item as multifunctional was its origin in a non-verbal category. We do not, however, wish to suggest that items with a single function should generally be considered multifunctional.

languages differ in the inventory of derivational affixes, e.g. the suffix *-olj* can not be found in Slovenian. The other reason is the size of the Slovenian sub-base, which includes fewer verbs than the BCMS sub-base. Consequently, some affixes do not appear in the list in Table 6.4 as they appear in less common verbs. One such example is *-os*, which was identified for BCMS but not Slovenian, despite being found in Slovenian verb *žigosati* 'to stamp' from *žig* 'a stamp' which did not appear in the Slovenian sub-base because it is not frequent enough.

6.4 Three case studies

Since one of the purposes of this chapter is to illustrate the challenges we encounter in the description and analysis of multifunctional items, we will present three case studies in this section, where we will explore three multifunctional items that appear in the verbal domain. In doing so, we will also address the methodological issues concerning the annotation of multifunctional elements, which can be instrumental when using these and other annotations in *WeSoSlaV*, as well as in performing supplementary annotations.

6.4.1 Example 1: The suffix *-ov*

We start with the suffix *-ov*, which we have already mentioned as an example of a multifunctional suffix in Section 6.2. We chose this suffix precisely because of its many functions which allow us to show some of the challenges that arise when exploring multifunctionality of suffixes, but also show how identifying these functions can contribute to ongoing debates in morphology. In what follows, we show the variety of *-ov*'s functions we can identify using *WeSoSlaV*, Section 6.4.1.1, and then turn our attention to prosody of verbs with *-ov* in Section 6.4.1.2 to explore how the annotation in the database can help us test predictions of accounts that center on the interaction of stress and structure of words.

Before turning to the intricacies of its behavior, a short note is in order regarding its surface allomorphs. The suffix *-ov* has a phonologically predictable allomorph *-ev*, which is realized if the suffix follows a set of consonants traditionally termed soft (i.e., *j*, *c* [ʦ], *č* [ʧ], *ž* [ʒ] and *š* [ʃ] in Slovenian).[4] One such example is the verb *dež-ev-a-ti* 'to rain' (from *dež* 'rain'). This allomorphy occurs regardless of the categorial context in which the affix appears (e.g., in the adjective *češnj-ev* 'made of cherries', or in the nominal paradigm *boj-i boj-ev* 'fight-NOM.PL, fight.GEN.PL'). This note becomes especially relevant in the context of

[4] In BCMS also ć [ʦ̑], lj [ʎ], nj [ɲ], dž [ʤ] and đ [ʤ̑].

Table 6.2: Multifunctional suffixes annotated in the BCMS sub-base

Suffix	Example	Translation
(a)c	u-nov-č-i-ti [unoʊtʃiti]^{PFV}	‘to cash out’
ać	o-dom-ać-i-ti^{PFV}	‘to domesticate’
ač	ras-tum-ač-i-ti^{PFV}	‘to clarify’
aj	pre-za-log-aj-i-ti^{PFV}	‘to have a snack’
(a)k	peš-ak-i-ti [peʃatʃiti]^{IPFV}	‘to hike’
(a)lj	češ-lj-a-ti^{IPFV}	‘to comb’
(a)n	ob-zn-an-i-ti^{PFV}	‘to announce’
ar	krv-ar-i-ti^{IPFV}	‘to bleed’
at	svoj-at-a-ti^{IPFV}	‘to appropriate’
av	s-mrš-av-i-ti^{PFV}	‘to lose weight’
c	jur-c-a-ti^{IPFV}	‘to run around’
en	o-kam-en-i-ti^{PFV}	‘to petrify’
ev	o-duš-ev-i-ti^{PFV}	‘to enthuse’
er	trep-er-i-ti^{IPFV}	‘to flicker’
ic	gran-ic-i-ti [granitʃiti]^{IPFV}	‘to border’
ič	sumnj-ič-i-ti^{IPFV}	‘to suspect’
ij	rob-ij-a-ti^{IPFV}	‘serve a sentence’
ik	jad-ik-ov-a-ti^{IPFV}	‘to moan’
in	bašt-in-i-ti^{IPFV}	‘to inherit’
iš	jur-iš-a-ti^{IPFV}	‘to rush’
išt	s-klad-išt-i-ti^{IPFV}	‘to store’
iv	u-milost-iv-i-ti^{PFV}	‘to propitiate’
iz	simbol-iz-ov-a-ti^{IPFV}	‘to symbolize’
j	jak-j-a-ti [jatʃati]^{IPFV}	‘to strenghten’
k	piš-k-i-ti^{IPFV}	‘to pee’
lj	kap-lj-a-ti^{iPFV}	‘to drop’
n	bes-n-i-ti^{IPFV}	‘to rage’
ok	s-ved-ok-i-ti [svedotʃiti]^{IPFV}	‘to testify’
olj	mig-olj-i-ti^{IPFV}	‘to squirm’
os	žig-os-a-ti^{iPFV}	‘to stigmatize’
oš	o-pust-oš-i-ti^{PFV}	‘to devastate’
ot	sram-ot-i-ti^{IPFV}	‘to disgrace’
ov	u-tjel-ov-i-ti^{PFV}	‘to embody’
š	mek-š-a-ti^{IPFV}	‘to soften’

Table 6.3: Multifunctional suffixes annotated in the BCMS sub-base cont.

Suffix	Example	Translation
telj	s-prija-telj-i-ti$^{\text{PFV}}$	'make a friend'
uć	o-mog-uć-i-ti$^{\text{PFV}}$	'to enable'
ulj	smej-ulj-i-ti$^{\text{IPFV}}$	'to smile'
uš	pev-uš-i-ti$^{\text{IPFV}}$	'to croon'
ut	skak-ut-a-ti$^{\text{IPFV}}$	'to jump around'

imperfectivizing suffixes, since there also exists a secondary imperfectivizer *-ev*. To distinguish *-ov*'s allomorph *-ev* from the suffix *-ev*, we need to consider stress in the non-finite forms of the verb. While *-ev* verbs have the stress on the first vowel of the suffix, (18), *-ov* verbs never have this stress pattern. Moreover, *-ev* verbs have the same exponent in finite and non-finite forms, (18), while *-ov* verbs have the *-u* allomorph in finite forms, (16) and (17).

(16) dež-ev-á-ti, dež-ú-je-∅ (Slo)
rain-*ov*-TV-INF, rain-*ov*-TV-3.SG
'to rain' 'it rains'

(17) ob-jok-ov-á-ti, ob-jok-ú-je-∅ (Slo)
at-cry-*ov*-TV-INF, at-cry-*ov*-TV-3.SG
'to mourn' '(s)he mourns'

(18) za-rd-év-a-ti, za-rd-év-a-∅ (Slo)
behind-red-*ev*-TV-INF, behind-red-*ev*-TV-3.SG
'to blush' '(s)he blushes'

Having addressed the allomorphy pattern, in 6.4.1.1 we present an overview of the uses of *-ov* in the verbal domain, whereas in 6.4.1.2 we discuss the prosodic behavior of *-ov* as an indicator of its structural position in the different contexts.

6.4.1.1 The suffix *-ov* in the verbal domain

Turning to *-ov* in the verbal domain and focusing on Slovenian for a start, *-ov* in verbs typically appears as a part of a string *-ova*, which traditional literature considers a verbal or infinitival suffix, see, e.g., Toporišič (2004: 346) and Šekli (2010: 157), respectively. Traditional analyses then essentially see *-ova* as a conjugation

Table 6.4: Multifunctional suffixes annotated in the Slovenian sub-base

Suffix	Example	Translation
ač	u-dom-ač-i-ti^{PFV}	'to domesticate'
ak	en+ak-i-ti [enačiti]^{IPFV}	'to equate'
ar	koles-ar-i-ti^{IPFV}	'to bike'
av	krv-av-e-ti^{IPFV}	'to bleed'
b	u-glas-b-i-ti^{PFV}	'to set music to'
ec	v-nov-ec-i-ti [vnovčiti]^{IPFV}	'to cash in'
eh	bol-eh-a-ti^{IPFV}	'to suffer'
elj	s-po-prijat-elj-i-ti^{PFV}	'to befriend.'
ic	kop-ic-i-ti [kopičiti]^{IPFV}	'to accumulate'
ik	svet-l-ik-a-ti^{IPFV}	'to glisteb'
il	krm-il-i-ti^{IPFV}	'to steer'
in	koren-in-i-ti^{IPFV}	'to root'
išč	sklad-išč-i-ti^{PFV}	'to store'
it	o-plemen-it-i-ti^{PFV}	'to refine'
iz	stabil-iz-ir-a-ti^{PFV}	'to stabilize'
j	z-oz-j-a-ti [zožati]^{IPFV}	'to narrow'
k	slad-k-a-ti^{IPFV}	'to sweeten'
l	raz-svet-l-i-ti^{PFV}	'to illuminate'
lj	kap-lj-a-ti^{IPFV}	'to dribble'
n	bes-n-e-ti^{IPFV}	'to rage
oč	o-mog-oč-i-ti^{PFV}	'to enable'
oš	pust-oš-i-ti^{IPFV}	'to ravage'
ot	o-sram-ot-i-ti^{PFV}	'to disgrace'
ov	raz-pol-ov-i-ti^{PFV}	'to halve'
š	z-manj-š-a-ti^{PFV}	'to reduce'
telj	s-po-prija-telj-i-ti^{PFV}	'make a friend'

class marker, so that the entire sequence can be viewed as a theme vowel. In *WeSoSlaV*, however, we follow Quaglia et al. (2022: 5) in treating items like *-ova* as containing a derivational affix followed by a theme vowel, as illustrated by (16) and (17). The sequence then consists of the suffix *-ov*, which has the allomorph *u* in the finite forms of the verb, and the independently attested theme vowel *a/je* (cf. *or-a-ti, or-je* 'to plough, (s)he ploughs').

Whereas the form of *-ov* in the verbal domain (the allomorphs and the preferred theme vowel) are relatively straightforward,[5] the functions or uses that *-ov* can have in the verbal domain are somewhat harder to define. Table 6.5 sums up the uses of *-ov* recorded in the Slovenian sub-database.

Table 6.5: Multifunctional *-ov* as annotated in the Slovenian sub-base

Context/Function	Number	Example
Imperfectivizing *ov*	174	
ov in SI	166	*izboljševati* 'to improve'
ov w. no PFV counterpart	2	*občudovati* 'to admire'
ov w. prefixless PFV counterpart	6	*kupovati* 'to buy'
Verbalizing *ov*	89	
nP-base	70	*pojmovati* 'to regard'
• nP-base no prefix	49	*deževati* 'to rain'
• nP-base further prefixed	21	*podedovati* 'to inherit'
AdvP-base	6	*nasprotovati* 'to object'
AdjP-base	2	*hudovati* 'to be mad'
Roots	10	*varovati* 'to guard'
Phrasal base	1	*vsebovati* 'to contain'
Total (imperf. & verbal.)	263	

As is clear from table 6.5, *-ov* can function as an imperfectivizing suffix. In this function, it can imperfectivize a simplex perfective verb, (19), or a prefixed perfective verb, (3). In fact, as shown in Simonović & Mišmaš (2023a), *-ov* is the second most common secondary imperfective suffix in Slovenian. This is in stark contrast to BCMS where *-ov* only imperfectivizes one verb.[6]

[5] As with *-av*, discussed in Section 6.3, the theme vowel selection is different in deadjectival items: *sur-ov-e-ti* 'roughen' (from *sur-ov* 'rough'), *za-got-ov-i-ti* 'ensure' (from *got-ov* 'certain'), *iz-jal-ov-i-ti* 'fall through' (from *jal-ov* 'barren').

[6] In addition to the fact that BCMS *-ov* can be used as an imperfectivizer on a single verb, *kupiti*^{PFV}, *kupovati*^{IPFV} 'to buy', the BCMS *-ov* is also different in that it serves to integrate

(19) kup-i-ti$^{\text{PFV}}$, kup-ov-a-ti$^{\text{IPFV}}$ (Slo/BCMS)
buy-TV-INF, buy-*ov*-TV-INF
'to buy' 'to buy'

As reported in Table 6.5, our sub-database also contains two instances of so-called orphan secondary imperfectives (see Polančec 2018 for a discussion). In these verbs *-ov* seems to act as a secondary imperfectivizing suffix, however, the prefixed perfective base is not independently attested, as illustrated in (20). We separate these *-ov* verbs verbs both from 'proper' secondary imperfectives and from imperfective pairs of simplex perfectives.

(20) žal-i-ti$^{\text{IPFV}}$ *ob-žal-i-ti$^{\text{PFV}}$ ob-žal-ov-a-ti$^{\text{IPFV}}$ (Slo)
sorry-TV-INF at-sorry-TV-INF at-sorry-*ov*-TV-INF
'to offend' 'to regret'

It should also be noted that *WeSoSlaV* contains instances of verbs with *-ov* in which *-ov* has the function of a secondary imperfective, but in which the form with *-ov* is characterized by an additional meaning. Consider example (21). While both the perfective prefixed and the imperfective verb have the meaning 'to pull', only the latter verb also has the second meaning, i.e., 'to compete'.

(21) po-teg-ni-ti$^{\text{PFV}}$, po-teg-ov-a-ti$^{\text{IPFV}}$ (Slo)
on-pull-TV-INF on-pull-*ov*-TV-INF
'to pull' 'to pull repeatedly/to compete'

In the second large group of examples, *-ov* acts as a verbalizer, which, as shown in Simonović & Mišmaš (2020), can take as its base nouns, (22), adjectives, (23), phrases, (24), and roots which do not surface as independent words, (25). These verbs are consistently imperfective (or biaspectual). Examples below are adapted from Simonović & Mišmaš (2020: (14)).

(22) pot, pot-ov-a-ti$^{\text{IPFV}}$ (Slo)
path path-*ov*-TV-INF
'path' 'to travel'

(23) moder, modr-ov-a-ti$^{\text{IPFV}}$ (Slo)
wise.ADJ wise-*ov*-TV-INF
'wise' 'to theorize'

new loanwords, e.g., in *googlovati/guglovati* 'to google', in some varieties (which Kovačević et al. 2024 term Eastern SC) – a function that *-ov* does not have in Slovenian.

(24) v sebi, vseb-ov-a-ti$^{\text{IPFV}}$ (Slo)
in self in-self-*ov*-TV-INF
'in self' 'to contain'

(25) *var, var-ov-a-ti$^{\text{IPFV}}$ (Slo)
guard-*ov*-TV-INF
'to guard'

Some instances of *-ov* in table 6.5 are marked as *further prefixed*. These are instances of *-ov* in prefixed counterparts of verbs with the verbalizing *-ov*. We separate *further prefixed* verbs from imperfectivizing *-ov* because of aspect. While the addition of an imperfectivizing *-ov*, as the name implies, results in an imperfective verb, unprefixed verbs with the verbalizing *-ov* are imperfective/biaspectual (as are most simplex verbs in Slavic), but their prefixed pairs are perfective (as are prefixed simplex verbs). One such example is the perfective *pre-potovati*^{PFV} 'to cross by traveling', cf. (22). Taking perfectivity as a cue, this indicates that the prefixes in these examples are merged above *-ov*, similarly to superlexical prefixes that attach above the secondary imperfective suffix, see for example Svenonius (2004b).

A subset of *-ov* verbs are one part of a minimal pair of imperfective forms in which *-ov-a* alternates with a single theme vowel. These are included in the *nP-base*-line of table 6.5. One such example is the pair *gost-i-ti - gost-ov-a-ti* 'to host - to be hosted'.[7] In such items, *-ov* does not seem to have one single function. While one could argue that *-ov* is an intransitive suffix in these verbs (as in *gostiti - gostovati* 'to host - to be hosted', *delati - delovati* 'to work/do - to act', *miriti - mirovati* 'to calm somebody down - to rest' where the second of the pair is always intransitive), this is not always the case. Consider the pair *dvigati - dvigovati* in (26) – both are imperfective, both are transitive, both can be used in the same context (in fact, they also have comparable frequencies: the former appears 22.440 times in the *Gigafida 2.0* (2019) corpus, the latter 21.114 times):

[7] These *ov*-verbs can sometimes show surface overlaps with secondary imperfectives. As example (i) shows for *pre-del-ov-a-ti*$^{\text{IPFV}}$ 'to re-work', this item cannot be taken as a further prefixed version of *del-ov-a-ti*$^{\text{IPFV}}$ 'to act' because *pre-del-ov-a-ti* is imperfective. Rather *pre-del-ov-a-ti* is a secondary imperfective pair of *pre-del-a-ti*$^{\text{PFV}}$ 'to re-work':

(i) del-o, del-a-ti$^{\text{IPFV}}$, del-ov-a-ti$^{\text{IPFV}}$, pre-del-ov-a-ti$^{\text{IPFV}}$
work-NEUT.NOM.SG work-TV-INF work-*ov*-TV-INF, re-work-*ov*-TV-INF
'work, to work, to act, to re-work'

(26) Miha je dvigal$^{\text{IPFV}}$ / dvigoval$^{\text{IPFV}}$ denar skoraj vsak dan.
Miha AUX raise / raise money almost every day
'Miha withdrew money almost every day.'
(Slo)

We leave it to future research to provide a more accurate analysis of *-ov* in these examples.

Turning to *-ov* annotated as multifunctional in *WeSoSlaV*, as already mentioned, *-ov* tends to appear in the string *-ova*, where *-a* is the theme vowel *a/je*.[8] But as mentioned in footnote 5, there are a few examples in the database in which *-ov* does not appear with the theme vowel *a/je* but instead with other theme vowels, which also speaks in favor of the segmentation of *-ova* into *-ov* and *-a* (in both Slovenian and BCMS). For Slovenian, one such example is the verb *ustanoviti*$^{\text{PFV}}$ 'to establish', which does not have an unprefixed imperfective counterpart (**stanoviti*) and in which the suffix *-ov* does not have the allomorph *-u* (*ustanovi* '(s)he establishes'). Put differently, *-ov* in this example does not have the function we associate with verbal suffixes, nor does it have the form of the "verbal" *-ov*. In BCMS one such example is the verb *u-mir-ov-i-ti* 'to retire somebody'.

The root *mir* 'peace' can appear in three different verbs as shown in (27). Note also that the unprefixed *mir-ov-i-ti* is not attested.

(27) mir-i-ti – mir-ov-a-ti – u-mir-ov-i-ti (BCMS)
peace-TV-INF peace-*ov*-TV-INF peace-*ov*-TV-INF
'to soothe' 'to rest' 'to retire'

It is not the case that *-ov* in *umiroviti* is needed to verbalize the root/noun, as *miriti* without *-ov* exists. It is also not the case *-ov* is needed to imperfectivize the verb in *mirovati*, as *miriti* is imperfective – as most simplex verbs in Slavic. It rather seems that *-ov* serves to modify the base and this results in adding a hard-to-pinpoint meaning. Similar occurrence of *-ov* also happens in adjectives. Consider the Slovenian pair *mir-en* 'peaceful' and *mir-ov-en* 'peace.ADJ' as in

[8] Note, however, that Kovačević et al. (2024), for example, start from the assumption that *-ova* is a theme vowel in BCMS, even though the BCMS *-ova* has the same properties as the Slovenian *-ova* when it comes to allomorphy – it too can be decomposed into *ov* (that gets an allomorph *u* in finite forms) and the independently attested theme vowel a/je. However, the view that *-ova* is a single theme vowel in BCMS is substantiated by the fact that it is in competition with another element that much more clearly qualifies as a theme vowel – *a* – in the context of loanword integration and denominal verbalisation. That is, *a* and *ova* are possible in new verbs, such as *guglati* and *guglovati* 'to google' in Eastern BCMS.

mirovna konferenca ‘peace conference’. The function of *-ov* here is only to modify the base. Such uses of *-ov* were annotated as multifunctional in the database.

6.4.1.2 The prosodic behavior of *-ov*

In this section we will consider the prosodic behavior of *-ov* across the domains, discuss a theoretical account based on this behavior (Simonović 2022a) and consider some ways in which data from *WeSoSlaV* can be used to test the prediction of such accounts.

The discussion on multifunctional suffixes is closely related to the discussion about the interaction of stress and structure, first illustrated in Marvin (2002). The following minimal pairs from Marvin (2002: 130) show the relevant contrast:

(28) a. stár, stár-ost, star-ást (Slo)
old.ADJ old-ness.N old-ness.N
‘old’ ‘oldness’ ‘old age’

b. visɔ́k, visɔ́k-ost, visok-ást
tall.ADJ tall-ness.N tall-ness.N
‘tall’ ‘tallness’ ‘royal highness’

In Slovenian, the suffix *-ost* ‘-ness’ is a nominalizer which can, together with the base it chooses, result in a predictable, compositional meaning (in these examples ‘property of adjective’) or an unpredictable/idiosyncratic meaning. The former is correlated with preserved stress of the adjective, the latter with a stress shift to the suffix. Both the stress and the meaning can be accounted for if we assume, following Marvin (2002), that the suffix *-ost* can be merged either with a phrase (in this case an aP) or with a root. As the meaning and the prosodic pattern of roots is stable once they are categorized (Marantz 1997, Arad 2003), assuming phasal spell-out that occurs at categorial heads (so already at the aP in deadjectival nominalisations) yields the correct result. In the rightmost examples of (28a) and (28b), which are analyzed as deradical (i.e., derived from a root), *-ost* is merged with an acategorial root, the meaning of which is not stable. The same goes for prosody, and so the prosodic pattern of these nouns can be influenced by the suffix.

Building on Marvin’s model, Simonović (2022a) argues that the stress pattern in deradical derivations can be analyzed as the default in Slovenian and the lexical specification of stress on affixes can be dispensed with. Simonović (2022a) also argues that similar systematic contrasts also emerge with multifunctional suffixes appearing in different categories. The multifunctional suffix *-ov* is a relevant example, as Simonović (2022a)’s comparison of *-ov* in verbs and *-ov* in

adjectives illustrates. That is, *-ov* tends to appear with the theme vowel *a/ja* and the stress is virtually always on the theme vowel in the infinitive form, regardless of *-ov*'s function. In adjectives, on the other hand, stress is virtually always on the material preceding it.

(29) a. dolg, dolg-ov-á-ti$^{\text{IPFV}}$ (Slo)
debt.N long-*ov*-TV-INF
'debt' 'to owe'

b. pís-a-ti$^{\text{IPFV}}$, pre-pís-a-ti$^{\text{PFV}}$, pre-pis-ov-á-ti$^{\text{IPFV}}$
write-TV-INF re-write-TV-INF re-write-*ov*-TV-INF
'to write' 're-write' 're-write'

(30) a. bámbus, bámbus-ov (Slo)
bamboo bamboo-*ov*.ADJ
'bamboo' 'made of bamboo'

b. limón-a, limón-ov
lemon-F.SG.NOM lemon-*ov*.ADJ
'lemon' 'made of lemon'

c. Jákob, Jákob-ov
Jakob Jakob-*ov*.ADJ
'Jakob' (name) 'Jakob's'

We see that the suffix *-ov* is stress-preserving in adjectives, which indicates that in these examples, *-ov* selects for a category, specifically, nP. As expected, *ov*-adjectives thus also have a compositional meaning: if the noun is a masculine or neuter animate/human noun, the adjective will be interpreted as possessive, but with other nouns (regardless of the gender), the derived adjective will be interpreted as kind.

The situation is somewhat more complex with verbs. That is, at least when *-ov* is in the role of a secondary imperfectivizer, we should expect the prosody of the verbal base to be preserved. That is, there are no reasons to assume that it is anything but categorized, and as such a spelled-out base that is being imperfectivized. And yet, as shown by data from *WeSoSlaV*, which are reported in Table 6.6, the stress is shifted to the theme vowel of the imperfectivized verb. This means that an analysis extended from Marvin (2002) would not be able to account for the stress in these examples.

However, Simonović (2022a), assuming the proposal in Quaglia et al. (2022), which was already mentioned above and which treats *-ov-a* in secondary imperfectives as miniverbs, proposes a different solution. Specifically, he observes

that Slovenian verbs generally have stress falling on the theme vowel or on the syllable preceding it. Simonović (2022a) therefore argues that in verbs stress is controlled by theme vowels, which, unlike derivational affixes, have lexical stress and are able to enforce stress shifts. In imperfectivized verbs with *-ov*, stress always falls on the theme vowel in the infinitive and on the syllable preceding the theme vowel in finite forms (as we saw in (16) and (17) above). This enables Simonović (2022a) to group this class together with simplex *a/je* verbs (such as *or-á-ti, ór-je* 'to plough, (s)he ploughs') and to conclude that stress is controlled by the theme vowel in both. This proposal is couched in a specific view of derivational affixes that treats affixes as acategorial transitive roots (transitive in that they can take complements), first proposed in Lowenstamm (2014). In this approach the suffix *ov* has an Encyclopedia entry for its verbalizing function and a different one for the adjectivizing function (and also others, which we leave aside here).

The Slovenian sub-base of *WeSoSLaV* seems to yield some support to this proposal, given that - looking beyond deverbal *-ov* verbs - we can observe that there are two possible stress patterns. First and foremost, stress can always be as described above: on *-ov*'s theme vowel *-a* in the infinitive and on *-ov*'s allomorph *-u* in the finite form. However, in a limited set of verbs, some speakers also allow a second pattern, one that consistently stresses the root (the syllable before *-ov*). Table 6.6 sums up the data, and example (31a) shows the relevant contrast.

Table 6.6: Stress in *ov* verbs as annotated in the Slovenian sub-base

Context/Function	Total N	TV	Root
Imperfectivizing *ov*	174		
ov in SI	166	166	0
ov w. no PFV counterpart	2	2	0
ov w. prefixless PFV counterpart	6	6	0
Percentage per stress pattern		100%	0%
Verbalizing *-ov*	89		
nP-base	70	70	14
Adv-base	6	6	1
Adj-base	2	2	2
Roots	10	10	1
Phrasal	1	1	0
Percentage per stress pattern	89	100%	20,2%

(31) a. ded-ov-á-ti, ded-ú-je-mo
b. déd-ov-a-ti, déd-u-je-mo (Slo)
grandfather-*ov*-TV-INF, grandfather-*ov*-TV-1.PL
'to inherit' 'we inherit'

Simonović (2022a) models the variation between the two prosodic patterns in (31a) and (31b) in Optimality Theory in terms of a competition between two forces which regulate stress in Slovenian: faithfulness to noun stress and faithfulness to theme vowel stress. In speakers who only allow the former pattern, theme vowel stress always wins, whereas for speakers who allow both patterns noun stress is equally strong. The data presented in Table 6.6 show that the reference to faithfulness to nouns stress may be too restrictive, as root stress is attested in other domains as well, which then begs the question whether there are independent arguments that all the other categories get nominalized before *-ov* is attached.

6.4.2 Example 2: The suffix *-k*

While *-ov* displays a variety of functions across categories, the suffix *-k* is, at first glance, most prominently associated with a diminutive interpretation in the verbal domain, (32). Given that in this function *-k* modifies the verb, in examples parallel to (32) *-k* was annotated as a verbal suffix.

(32) a. tap-k-a-ti$^{\text{IPFV}}$ (BCMS)
tap-*k*-TV-INF
'to tap a little/gently'
b. lež-k-a-ti$^{\text{IPFV}}$ (Slo)
lie-*k*-TV-INF
'to lie a little'

In fact, *-k* associated with a diminutive interpretation is found in other categories as well. This is demonstrated by (33a) for nouns and by (34) for adjectives (note that not all speakers accept the noun *črtka* 'a small/short line', hence the question mark in the example).

(33) a. cv(ij)et-a, cv(ij)et-k-a (BCMS)
flower-M.SG.GEN flower-*k*-M.SG.GEN
'a flower' 'a small flower'
b. fant-a, fant-k-a (Slo)
boy-M.SG.GEN boy-*k*-M.SG.GEN
'a boy' 'a little boy'

(34) črt-a, ?črt-k-a, črti-ic-a, črt-ast,
line-F.SG.NOM line-*k*-F.SG.NOM line-*ic*-F.SG.NOM line-*ast*.ADJ
'a line' 'a small/short line' 'a small/short line' 'covered in lines'
črt-k-ast (Slo)
line-*k*-*ast*.ADJ
'covered in short lines'

This is not to say that *-k* has no other function. In BCMS and Slovenian nouns, for example, it can be used to derive feminine counterparts to masculine nouns denoting humans:

(35) a. Talijan-∅, Talijan-k-a (BCMS)
Italian-M.SG.NOM Italian-*k*-F.SG.NOM

b. Italijan-∅, Italijan-k-a (Slo/BCMS)
Italian-M.SG.NOM Italian-*k*-F.SG.NOM
'Italian' 'Italian'

Furthermore, *-k* is also a nominalizer. As observed in Toporišič (2004) for Slovenian, combing *-k* with a verbal base results in a noun with an event, (36a), or a result, (36b), interpretation and *-k* can also nominalize adjectives, (37). Note that the Slovenian nominalizing suffix *-ek* (as in in (36)) is really just *-k*, with the vowel in *-ek* being just an epenthetic vowel inserted to avoid a complex coda. As such, the vowel is absent in many forms of each paradigm, such as the genitive singular form of the noun in (36b), i.e., *ostanka* 'leftover'.

(36) a. raz-misl-i-ti, raz-misl-ek (Slo)
over-thought-TV-INF over-thought-*k*
'to think over something' 'deliberation'

b. o-st-a-ti, o-st-a-n-ek
about-stand-TV-INF about-stand-TV-PASS.PTCP-*k*
'to stay' 'leftover'

(37) megl-a, megl-en, megl-en-k-a (Slo)
fog-F.SG.NOM fog-*en*.ADJ fog-*en*-*k*-F.SG.NOM
'fog' 'foggy' 'fog-light'

These uses of the suffix *-k* show that it qualifies as multifunctional, as it has several functions across different categories. But the multifunctionality of *-k* is also evident if we zoom into verbs. That is, considering the verbal domain more

closely shows that *-k* is not just a diminutive suffix in verbs, but rather also has other functions.

If we consider the data from *WeSoSlaV*, summarized in Table 6.7, we can observe that in BCMS the diminutive *-k* is dominant among the uses of *-k*. In such examples, *-k* is annotated as a verbal suffix (in accordance with our definition of verbal suffixes having a verbalizing, aspectual or verbal-modifying function). Note that numbers in Table 6.7 for BCMS do not add up to the total of 55 because some diminutive verbs are also counted as verbs with an imperfectivizing *-k*, as for instance the example (32a). The latter group of verbs is characterized by having a perfective counterpart with a *nV/ne*-suffix.[9] We return to this group below.

The Slovenian sub-database, on the other hand, contains no verb that can be analyzed as diminutive. Of course, this is not to say that such verbs are not otherwise attested in Slovenian. Traditional grammars, such as Toporišič (2004), do take *-k* to be a typical diminutive suffix, as in (38)

(38) bož-a-ti$^{\text{IPFV}}$, bož-k-a-ti$^{\text{IPFV}}$ (Slo)
stroke-TV-INF stroke-*k*-TV-INF
'to caress' 'to caress.DIM'

Such verbs are not an artefact of the traditional grammars, the verb *božkati* in (38) has three hits in *Gigafida 2.0* (2019)). This shows that the number of verbs included in the Slovenian sub-database is not large enough to detect all functions of different suffixes (and, in fact, to identify all suffixes in the verbal domain in the language).

While there is no diminutive *-k* in the Slovenian sub-database, other uses are attested. Firstly, in some instances *-k* seems to be a part of the nominal or adjectival base that is then verbalized, (39) and (40). Given that we have taken multifunctional suffixes to be those that do not not appear in any verbal function, potentially originate from a different category and/or have a vague meaning, instances of *-k* as a part of the nominal or adjectival base of the verb were annotated as multifunctional in *WeSoSlaV*.

(39) tip-∅, tip-k-a, tip-k-a-ti$^{\text{IPFV}}$, tip-k-ni-ti$^{\text{PFV}}$ (Slo)
touch-M.SG.NOM touch-*k*-F.SG.NOM touch-*k*-TV-INF touch-*k*-*ni*-INF
'touch' 'key' 'type' 'to type once'

[9] We follow Štarkl et al. (2024) and use the annotation *nV/ne* as a shorthand for the Slovenian *ni/ne* and BCMS *nu/ne* sequence.

Table 6.7: The suffix *-k* in the sub-databases

Context/Function	Number of Vs with *-k*	Roots
-k in BCMS		
Diminutive	36	24
In the verbal base	4	3
Imperfectivizing *-k*	9	9
Other *-k*	15	9
Sub-total	55	
-k in Slovenian		
Diminutive	0	0
In the verbal base	2	1
Imperfectivizing *-k*	6	1
Other *-k*	2	2
Sub-total	10	

(40) slad-∅, slad-ek, slad-k-a-ti$^{\text{IPFV}}$
malt-M.SG.NOM sweet-*k*.ADJ sweet-*k*-TV-INF
'malt/sugar' 'sweet' 'to sweeten'

While both types of bases result in imperfective verbs, only verbs with nominal bases can be perfectivized with the semelfactive *nV/ne*, (39). In addition, *-k* in the nominal base of the verb and the diminutive *-k* allows suffix stacking on top of it – not just with *nV/ne*, but also with imperfectivizing suffixes, as (41) shows for BCMS, see Štarkl et al. (2024) for futher data and an analysis.

(41) za-pit-k-iv-a-ti$^{\text{IPFV}}$ (BCMS)
for-ask-*k*-SI-TV-INF
'to ask repeatedly (and in an irritating way)'

Importantly for us, suffix stacking in (39) would then imply that we can treat the seemingly nominal *-k* on a par with a diminutive *-k* (or, in other words, to treat them as a single suffix). To confirm that this is indeed the case and to show if this can be extended to *-k* in adjectival bases as well, more examples with *-k* in their bases are needed. Finally, before we proceed to other functions, it is worth noting that *-k* can also stack on top of other (diminutive) suffixes:

(42) rad-i-ti$^{\text{IPFV}}$, rad-uc-k-a-ti$^{\text{IPFV}}$ (BCMS)
work-TV-INF work-*uc*-*k*-TV-INF
'to work' 'work a bit'

(43) drem-a-ti$^{\text{IPFV}}$, drem-uc-k-a-ti$^{\text{IPFV}}$ (Slo)
nap-TV-INF nap-*uc*-*k*-TV-INF
'to nap' 'nap a bit'

In addition to verbs like (39), which can be perfectivized with the addition of *nV/ne*, there are also verbs with the *k*-suffix which behaves as the imperfective counterpart to a semelfactive *nV/ne* verb. According to Štarkl et al. (2024), this happens in almost 30% of all unprefixed BCMS *nV/ne* verbs in *WeSoSlaV* (and no such verbs in Slovenian), and in 7% of prefixed *nV/ne* verbs in Slovenian (only 1.2% in BCMS).

(44) gric-nu-ti, gric-k-a-ti (BCMS)
nibble-*nu*-INF, nibble-*k*-TV-INF
'to nibble once' 'to nibble'

(45) s-tis-ni-ti, s-tis-k-a-ti (Slo)
with-press-*ni*-INF, with-press-*k*-INF
'to press once' 'to press'

In Table 6.7 these are counted in the IMPERFECTIVIZING *-K* row and in *WeSoSlaV* these occurrences of *-k* are annotated as having a verbal *-k*. But this is just an initial approximation. That is, while the resulting verb with *-k* is indeed imperfective, the suffix does not behave as suffixes we can find in secondary imperfectives. For example, *-ov*, which is often found in secondary imperfectives, does not not appear in a zero nominalizations (e.g., *iz-pis-ov-a-ti*$^{\text{IPFV}}$ 'to sign out', *iz-pis* 'a print out', **iz-pis-ov*), *-k*, on the other hand, does (*s-tisk* 'a squeeze', cf. (45)). We leave it to future research to propose a more detailed account of such examples, whereby a database such as *WeSoSlaV* can help in identifying relevant examples.

Finally, there are verbs which we currently counted in the OTHER *-K* line in Table 6.7. The two such Slovenian examples are *prečkati*$^{\textit{PFV/IPFV}}$ 'to cross' and *razburkati*$^{\textit{PFV}}$ 'to agitate (usually a body of water)'. Both verbs have a counterpart without *-k*, *prečiti*$^{\textit{PFV/IPFV}}$ 'to cross' and *razburiti*$^{\textit{PFV}}$ 'to agitate', respectively. As a comparison of these pairs show, the suffix *-k* in these cases dos not result in an imperfective verb, nor do the verbs without *-k* in these examples include the semelfactive *nV/ne*. This is why we separated them from verbs in the IMPERFECTIVIZING *-K* row. Furthermore, unlike in verbs in the IMPERFECTIVIZING *-K* row, the two verbs in the OTHER *-K* row in Slovenian do not form zero nominalizations. It is also not the case that *-k* acts as a diminutive suffix in these examples.

Rather, *-k* simply slightly modifies the meaning of the verb, but it is unclear at this point how this is achieved.

Although we started this section by noting that the suffix *-k* is generally associated with a diminutive function, data from *WeSoSlaV* reveals additional uses of *-k* in the verbal domain that still require analysis. We leave this to future work.

6.4.3 Example 3: the suffix *-š*

We offer one final example of a suffix that was annotated only as multifunctional in *WeSoSlaV* – the suffix *-š*. However, since the inquiry into this suffix in the Western South Slavic verbal domain is very much in its beginning stages (Butschety & Mišmaš 2024), we will only touch briefly on it in order to highlight some of the issues that emerge when considering it.

Unlike *-ov* and *-k*, the suffix *-š* has a single function (meaning) outside of the verbal domain. It also never changes the category of the base (again, unlike *-ov* and *-k*) and it has a predictable meaning regardless of its context. Specifically, in Slovenian *-š* is used in the comparative (and then superlative, which we here leave aside) form of some monosyllabic adjectives, as in (46), see, for example, Toporišič (2004) for a more detailed description. It is similarly used to derive comparatives in BCMS, however, this is the case for only three adjectives (*lak, lakši* 'easy, easier'; *mek, mekši* 'soft, softer', *l(ij)ep, l(j)epši* 'pretty, prettier', see Klajn (2005) for Serbian). In both Slovenian and BCMS *-š* is also used in adverbs derived from comparative forms of adjectives that use *-š*, e.g, *lepše* 'more beautiful.ADV' in both Slovenian in BCMS. Similarly, *-š* also appears in verbs derived from comparative forms of adjectives that include *-š*. Examples (46) and (47) show two deadjectival verbs of the relevant kind in Slovenian that can be found in *WeSoSlaV*. Note that in addition to the comparative *-š*, example (47) also exhibits root suppletion.[10]

[10] We do not offer an analysis of comparative forms here, but see Caha et al. (2019), Vanden Wyngaerd et al. (2020) for similar forms in Czech and Slovak, respectively, but it should be mentioned that *-š* is not the only comparative suffix in Slovenian and BCMS. In Slovenian, the other comparative suffixes are *-j* and *-ejš* and in BCMS *-j* and *-ij*. As for deadjectival verbs, comparative forms with *-ejš* and *-ij* never serve a verbal base, *-j*, on the other hand, does. In fact, the sub-bases of *WeSoSlaV* also include verbs derived from comparatives with *-j*, as (i) shows for BCMS and (ii) for Slovenian. These verbs tend to behave the same as verbs with *-š* – they fall into the a/a TV-class and some require a prefix in BCMS.

(i) jak, jači(=jak-ji), o-jač-a-ti (BCMS)
strong, stronger, around-stronger-TV-INF
'strong' 'stronger' 'to strengthen'

(46) lep, lep-š-i, naj-lep-š-i,
beautiful beautiful-*š*-M.SG.NOM most-beautiful-*š*-M.SG.NOM
'beautiful' 'more beautiful' 'the most beautiful'
lep-š-a-ti^{IPFV} (Slo)
beautiful-*š*-TV-INF
'to embellish'

(47) dober, bolj-š-i, najbolj-š-i; bolj-š-a-ti^{IPFV} (Slo)
good, good-*š*-M.SG.NOM most-good-*š*-M.SG.NOM good-*š*-TV-INF
'good' 'better' 'the best' 'to better'

The situation is similar in BCMS, in which, as already mentioned, *-š* only appears in three comparatives. All three also derive a verb, but some such deadjectival verbs require a prefix, (49).

(48) mek, mek-š-i, naj-mek-š-i, mek-š-a-ti^{IPFV} (BCMS)
soft, soft-*š*-M.SG.NOM, most-soft-*š*-M.SG.NOM, soft-*š*-TV-INF
'soft' 'softer' 'the softest' 'to soften'

(49) lep, lep-š-i, naj-lep-š-i,
beautiful, beautiful-*š*-M.SG.NOM, most-beautiful-*š*-M.SG.NOM,
'more beautiful' 'the most beautiful'
*lep-š-a-ti, u-lep-š-a-ti^{PFV} (BCMS)
beautiful-*š*-TV-INF in-beautiful-*š*-TV-INF
'to embellish'

As verbs in examples (46) to (49) already show, *-š* in verbs is arguably the comparative suffix and verbs in which it appears have a comparative form of the adjective as their base.[11] This in general holds for both the Slovenian and the BCMS sub-database of *WeSoSlaV* which include several deadjectival verbs with the suffix *-š*. However, only the Slovenian verbs always have a corresponding

(ii) nizek, niž-j-i, naj-niž-j-i; niž-a-ti (Slo)
low, low-*j* most-low-*j*-M.SG.NOM good-er-TV-INF
'good' 'better' 'the best' 'to better'

[11] As a side note, verbs from comparative bases do not block verbs from positive forms. For example, in addition to *slabšati* 'to worsen' from *slabši* 'worse', the verb *slabiti* 'to weaken' from *slab* 'bad' also exists in Slovenian. As indicated by the translations of the two verbs, verb from the positive and the comparative form of the adjective also tend to have different meanings assigned to the root.

attested comparative *-š* form of the adjective (for all 9 roots with which *-š* appears in the Slovenian sub-base). But the situation is somewhat more complex for BCMS.

In the BCMS sub-database we find the three deadjectival verbs corresponding to the three *š*-comparatives in BCMS, but in addition to these we also find two verbs from "comparative" forms that are not attested in the language and one *š* verb that is not de-adjectival, see (52) below. The two relevant verbs have as their base adjectives with root suppletion in the comparative form and unexpectedly get the suffix *-š* only in the verb.

(50) dobar, bolji, najbolji, po-bolj-š-a-ti (BCMS)
good better best at-better-*š*-TV-INF
'good' 'better' 'the best' 'to better'

(51) zao, gori, najgori, po-gor-š-a-ti (BCMS)
bad worse the-worse at-worse-*š*-TV-INF
'bad' 'worse' 'the worst' 'to worsen'

Given that the suppletive form of the root in the verb is the same to the one in the comparative form of the adjective and that the meaning of the verbs is parallel to those of the comparatives, we can at this point still assume that *-š* has a single function in Slovenian and BCMS. Despite this stability, *-š* was classified as multifunctional in *WeSoSlaV* because it does not have a verbalizing, aspectual or modifying function and potentially originates from a different category, cf. Section 6.3. But given its single function, this is a considerable shortcoming in the annotation process. The only exception that could justify *-š* being truly multifunctional consists of the previously mentioned BCMS verb annotated as having a multifunctional suffix with *-š* that is not de-adjectival:

(52) tap-, tap-š-a-ti^{IPFV}, tap-nu-ti^{PFV}, ?tap-a-ti^{IPFV},
tap tap-*š*-TV-INF tap-*nu*-INF tap-TV-INF
'tap' 'to gently clap/tap' 'to tap' 'to move around while tapping'
tap-k-a-ti^{IPFV} (BCMS)
tap-*k*-TV-INF
'to tap'

It is, however, unclear if the root of the verb is in fact *tap-*, as annotated in *WeSoSlaV*. On the one hand, we could claim this is the case based on similar forms, given in (52) (the question mark with *tapati* indicates that not all speakers have this form). On the other hand, according to HJP (2006), the verb *tapšati* is

borrowed from the Hungarian *taps* 'a clap' which implies that the *-š* is simply a part of the root *tapš-*.

We leave the analysis of this example for future work and note that verbs with *-š* not only show a potential issue with the annotation of the database, but also open a variety of questions related to the analysis of deadjectival verbs. For example, it remains to be determined why some BCMS verbs require additional modification with the comparative suffix, (50), and why *-š* is the suffix to do it, but see Butschety & Mišmaš (2024) for some initial suggestions.

6.5 Conclusion

Unlike other chapters of this monograph, which focus on *WeSoSlaV* to show how it can be used to explore correlations between different annotated properties, we rather focus on the material gathered in *WeSoSlaV* to discus the phenomenon of suffix multifunctionality, whereby multifunctionality is a property annotated in the database. In doing so we focus on items that were identified as multifunctional in the database using the initial definition (i.e., suffixes that do not not appear in any verbal function, potentially originate from a different category and/or have a vague meaning) but also consider these suffixes when they have a verbalizing, aspectual or verb-modifying function (annotated separately in the database). This crucially allows us to show how annotating a property allows us to build an empirical base that is then in turn used to explore a single phenomenon.

By building an empirical basis we show that multifunctionality as a phenomenon is not uncommon in the languages under investigation, but we also show that focusing on a single category (such as verbs) does not allow one to capture the full scope of the functions a suffix can have. We focus on three items that were annotated as multifunctional in the database and that help pin-point some of the complexities that emerge when discussing multifunctionality, as well as some issues in our annotation of the *WeSoSlaV* material. Perhaps most importantly for the latter point, the initial definition of multifunctionality which was used when annotating, was overly vague. We assumed multifunctional suffixes to be those suffixes that do not not appear in any verbal function, potentially originate from a different category and/or have a vague meaning and consequently annotated *-š* as a multifunctional item despite the fact that *-š* only appears in BCMS (with one potential exception) and Slovenian with a single function. Secondly, we demonstrate that the size of *WeSoSlaV* is problematic. As the discussion of *-k* shows, not all functions of the suffix in the verbal domain can be observed when only the verbs included in the sub-database are considered.

As for multifunctionality as a phenomenon, *-k* and *-ov* show the range of functions an item can have within the category and beyond it. At the same time, these suffixes help us demonstrate how challenging it is to capture all the functions and point to some issues that should be addressed in future work. Furthermore, the discussion of *-ov* also reveals how important it is to consider prosody when discussing multifunctionality and verbal/word structure in general. That means that, while our central focus was not on correlations between annotated properties in the database, such correlations still have an important role in understanding specific phenomena.

7 Conclusion

This monograph presents the Western South Slavic Verbs (*WeSoSlaV*) database, a database containing annotations of 5300 verbs from Bosnian, Croatian, Montenegrin, Serbian (BCMS) and 3000 verbs from Slovenian in a large number of prosodic, inflectional, derivational, syntactic and semantic properties, primarily aimed at exploring their morphology, and provides a shallow quantitative description, as well as some deeper quantitative research on theoretically relevant questions based on the materials contained in *WeSoSlaV*. To our knowledge, this is the only database of its size to feature manual annotations across so many grammatical properties, not only in Slavic linguistics but in linguistics more broadly, offering unprecedented opportunities for the study of the verbal system, especially at the level of morphology.

WeSoSlaV was developed as the main deliverable of the bilateral research project between researchers from the universities of Graz and Nova Gorica titled *Hyperspacing the verb: The interplay between prosody, morphology and semantics in the Western South Slavic verbal domain* (FWF I 4215, ARIS N6-0113). The central aim of the database was to bridge the gap between theoretical and quantitative methodology in morphology. Theoretical linguists are primarily interested in the attested and unattested types of expressions, and in categorical regularities that they display. Quantitative linguistics offers valuable measurable insights, yet these are often shaped by the inherent properties of the data being quantified. The results therefore either fail to directly address theoretical questions or, when tailored to do so, addresses them very narrowly. Consequently, some central theoretical questions are not adequately explored.

Such is the case, for instance, with theme vowels. This class of morphophonological items has a long tradition of puzzling linguists, and being answered in an incomplete way. Various prominent phenomena of the verbal or nominal domain in which these occur have been investigated for a possible association with the occurrence and selection of the exact theme vowels, either with a weak confirmation or with none at all (from Brugmann 1892 to Oltra-Massuet 2020, Taraldsen Medová & Wiland 2019 and Milosavljević & Arsenijević 2022). Considering that the weak association has been observed for a variety of verbal features, from argument structure to aspect, to prosody, the possibility remains that a model

including various properties of verbs and their interaction might show that the selection of theme vowels is strongly, or at least weakly grammatically determined and thus carries information about their grammatical properties. This can only be done if a sufficient number of verbal properties are available for quantitative investigation. *WeSoSlaV* was aimed at enabling this kind of research: the development and quantitative testing of complex models, including nontrivial relations between a larger number of properties.

The purpose of the present monograph is threefold: to introduce *WeSoSlaV* to researchers in linguistics, to illustrate the main ways in which it can be used and to contribute original research of the kind introduced above.

In Chapter 2, based on the documentation of the process of development of *WeSoSlaV*, we provided the core description of the database. We explained how it was developed, specified the criteria for annotation both for the purpose of its users understanding the quality of the information it contains and for the possibility of its further development by the community. The chapter described the concrete components of the database that have been published, the properties annotated in them, the principles of annotation, problems that have emerged, and ways how they have been resolved.

The following four chapters were organized by the domains of morphology: inflection (Chapter 3) and two chapters on derivation, one for the derivation of verbs (Chapter 4), and another for the derivation of nouns and adjectives from verbal bases (Chapter 5). Each of these chapters presents general quantitative insights from *WeSoSlaV* regarding the respective domain and incorporates original research addressing a prominent research question in the field, based on the database. The purpose of the latter was both to contribute to the theory of verbal morphology and to manifest ways in which the database can be used. In Chapter 5, we also include some possible complementary methods, as well as a comparative methodological discussion.

Chapter 6 was a bridge towards the follow-up bilateral research project between the same two institutions titled *Multifunctionality in morphology* (ARIS J6-4614, FWF I6258). Its purpose was, rather than discussing the database with respect to an entire discipline of morphology as in the Chapters 3-5, to contribute novel research regarding one narrower research question: the issue of multifunctionality (i.e. syncretism in the broadest sense), simultaneously illustrating the use of *WeSoSlaV* in such a perspective.

As the careful reader has observed, a variety of topics have been discussed, from the locus of prosodic prominence in the verbal paradigm to the multifunctionality of the suffix *-ov*, yet one issue pervades the entire book, prominently presenting itself in almost all chapters. This is the issue of theme vowels. The

question of their syntactic, semantic, morphological and prosodic status naturally emerged as a central aspect of the book, manifesting relevance not only as a matter of inflection class, which is its natural place, but also as a major tool in the derivation of verbs, including in secondary imperfectivization, as well as in deverbal derivation, as a strong factor regarding the capacity of a verb to combine with a certain derivational suffix. While all these dimensions cast doubts on the view that theme vowels are purely ornamental, the monograph does not tackle the issue of the prediction of theme vowel assignment based on the full range of annotated properties: this is left as a task for systematic large-scale research in the future.

In sum, while this monograph touches on many key aspects of Western South Slavic verbs, much work remains to be done. The exploration of theme vowels, which emerged as a recurring focus throughout the chapters, serves as a reminder of the complexity and depth of these linguistic phenomena. The work presented here lays a solid foundation for future investigations, offering both a resource in *WeSoSlaV* and a roadmap for further research. As new insights emerge and the field continues to evolve, the themes addressed in this monograph will hopefully inspire continued exploration and refinement.

Abbreviations

1	first person
2	second person
3	third person
ACC	accusative
ADJ	adjective
BCMS	Bosnian/Croatian/ Montenegrin/Serbian
DIM	diminutive
F	feminine
GEN	genitive
XXX^{IPFV}	imperfective
INF	infinitive
INS	instrumental
IPFV	imperfective
M	masculine
N	neuter
NEUT	neuter
NOM	nominative
XXX^{PFV}	perfective
PFV	perfective
PL	plural
PRS	present
PTCP	participle
REFL	reflexive
SG	singular
SI	secondary imperfective
S.INF	short infinitive
TV	thematic vowel
WSS	Western South Slavic

References

Acedo-Matellán, Víctor & Cristina Real-Puigdollers. 2019. Roots into functional nodes: Exploring locality and semi-lexicality. *The Linguistic Review* 36(3). 411–436. DOI: 10.1515/tlr-2019-2019.

Acquaviva, Paolo. 2008. Roots and lexicality in distributed morphology. Ms., University College Dublin/Universität Konstanz. http://ling.auf.net/lingBuzz/000654.

Alexiadou, Artemis & Terje Lohndal. 2017. The structural configurations of root categorization. In Leah Bauke & Andreas Blümel (eds.), *Labels and roots*, 203–231. Boston / Berlin: De Gruyter Mouton. DOI: 10.1515/9781501502118-009.

Anderson, Stephen. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Arad, Maya. 2003. Locality constraints on the interpretation of roots: the case of Hebrew denominal verbs. *Natural Language & Linguistic theory* 21(4). 737–778. DOI: https://doi.org/10.1023/A:1025533719905.

Aronoff, Mark. 1994. *Morphology by Itself: Stems and Iinflectional Classes*. Cambridge, Massachusetts / London, England: The MIT Press.

Arsenijević, Boban. 2007. A unified analysis of two classes of Slavic verb prefixes. In Sylvia Blaho, Luis Vicente & Erik Schoorlemmer (eds.), *Console xiv proceedings*, 21–36. Leiden: ConSOLE. http://www.hum2.leidenuniv.nl/pdf/lucl/sole/console14/console14-arsenijevic.pdf.

Arsenijević, Boban. 2018. Serbo-Croatian suffix *–va* and the illusion of the imperfective specification. presented at the conference AMBIGO, held at the University of Göttingen. https://rb.gy/iqblpq.

Arsenijević, Boban. 2020. Deverbal nouns in -ie and their variation across the South Slavic area. *Linguistica* 60(1). 7–29. DOI: 10.4312/linguistica.60.1.7-29.

Arsenijević, Boban. 2023. Specification of telicity in Serbo-Croatian, without null prefixes. In Petr Biskup, Marcel Börner (né Guhl), Olav Mueller-Reichau & Iuliia Shcherbina (eds.), *Advances in formal Slavic linguistics 2021*. Berlin: Language Science Press. DOI: https://doi.org/10.5281/zenodo.10123627.

Arsenijević, Boban, Katarina Gomboc Čeh, Franc Lanko Marušič, Stefan Milosavljević, Petra Mišmaš, Jelena Simić, Marko Simonović & Rok Žaucer. 2024.

Database of the Western South Slavic Verb HyperVerb 2.0 – WeSoSlaV. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1846.

Arsenijević, Boban, Franc Lanko Marušič, Stefan Milosavljević, Petra Mišmaš, Marko Simonovic & Rok Žaucer. 2024. *Database of the Western South Slavic Verb HyperVerb (WeSoSlaV) – Deverbal Nominalizations*. Dataset. Zenodo. DOI: 10.5281/zenodo.14230589.

Arsenijević, Boban & Stefan Milosavljević. 2021. Serbo-Croatian theme vowels carry functional features. Presented at Theme vowels in V(P) Structure and beyond, University of Graz, April 22–23, 2021.

Arsenijević, Boban, Stefan Milosavljević & Marko Simonović. 2025. Secondary imperfectivisation is reverbalisation. *Morphology* 35. 271–304. DOI: https://doi.org/10.1007/s11525-025-09440-7.

Babić, Stjepan. 2002. *Tvorba riječi u hrvatskome književnome jeziku*. Zagreb: HAZU – Nakladni zavod Globus.

Barić, Eugenija, Mirko Peti, Mijo Lončarić, Dragica Malić, Slavko Pavešić, Vesna Zečević & Marija Znika. 1997. *Hrvatska gramatika*. Zagreb: Školska knjiga.

Biskup, Petr. 2019. *Prepositions, Case and Verbal Prefixes : The Case of Slavic*. Amsterdam / Philadelphia: John Benjamins Publishing Company. DOI: 10.1075/la.255.

Biskup, Petr. 2023. Aspect separated from aspectual markers in Russian and Czech. In Petr Biskup, Marcel Börner (né Guhl), Olav Mueller-Reichau & Iuliia Shcherbina (eds.), *Advances in formal Slavic linguistics 2021* (Open Slavic Linguistics 8), 61–98. Berlin: Language Science Press. DOI: 10.5281/zenodo.10123631.

Biskup, Petr. 2024. On imperfective suffixes in Russian. *Russian Linguistics* 48(1). 14. DOI: 10.1007/s11185-024-09299-y.

Borer, Hagit. 2005. *The normal course of events / structuring sense vol. ii*. Oxford: Oxford University Press.

Borer, Hagit. 2009. Roots and categories. Presented at the 19th Colloquium on Generative Grammar, University of the Basque Country, Vitoria-Gasteiz, April 2009. http://hborer.sllf.qmul.ac.uk/Downloads/Borer%202009%20roots_and_categories.pdf.

Borer, Hagit. 2013. *Taking Form / Structuring Sense Vol. III*. Oxford: Oxford University Press.

Borer, Hagit. 2014. The category of roots. In Artemis Alexiadou, Hagit Borer & Florian Schäfer (eds.), *The syntax of roots and the roots of syntax*, 112–148. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199665266.003.0006.

Brugmann, Karl. 1892. *Grundriss der vergleichenden grammatik der indogermanischen sprachen. volume ii, part ii: numerals and pronouns, verb*. Strassburg: Trübner.

Butschety, Madeleine & Petra Mišmaš. 2024. For *better* or *worse*, we need to talk about *-š*. Presentation given at the Adjectives, categorization and argument structure workshop at the 21st International Morphology Meeting. https://adjectives-imm21.com/wp-content/uploads/2024/08/butschety-mismas_for-better.pdf.

Caha, Pavel. 2009. *The nanosyntax of case*. University of Tromsø. (Doctoral dissertation). https://hdl.handle.net/10037/2203.

Caha, Pavel. 2018. Notes on insertion in distributed morphology and nanosyntax. In *Exploring nanosyntax*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780190876746.003.0002.

Caha, Pavel, Karen De Clercq & Guido Vanden Wyngaerd. 2019. The fine structure of the comparative. *Studia Linguistica* 73(3). 470–521. DOI: 10.1111/stul.12107.

Caha, Pavel & Tobias Scheer. 2008. The syntax and phonology of czech templatic morphology. In Andrei Antonenko, John F. Bailyn & Christina Y. Bethin (eds.), *Annual workshop on formal approaches to slavic linguistics: the stony brook meeting 2007* (FASL 16), 68–83. Ann Arbor: Michigan Slavic Publications.

Cinque, Guglielmo. 2015. Augmentative, pejorative, diminutive and endearing heads in the extended nominal projection. In Elisa Di Domenico, Cornelia Hamann & Simona Matteini (eds.), *Structures, strategies and beyond: studies in honour of Adriana Belletti*, 67–81. Amsterdam: John Benjamins Amsterdam. DOI: 10.1075/la.223.04cin.

Creemers, Ava, Jan Don & Paula Fenger. 2018. Some affixes are roots, others are heads. *Natural Language & Linguistic Theory* 36(1). 45–84. DOI: 10.1007/s11049-017-9372-1.

De Belder, Marijke. 2011. *Roots and affixes: eliminating lexical categories from syntax*. Netherlands Graduate School of Linguistics. (Doctoral dissertation).

Dickey, Stephen M. 2003. Verbal aspect in Slovene. *STUF-Language Typology and Universals* 56(3). 182–207. DOI: 10.1524/stuf.2003.56.3.182.

Dickey, Stephen M. 2000. *Parameters of Slavic aspect:A cognitive approach* (Dissertations in Linguistics). CSLI Publications.

Dickey, Stephen M. 2012. Orphan prefixes and the grammaticalization of aspect in South Slavic. *Jezikoslovlje* 13(1). 71–105. https://hrcak.srce.hr/86246.

Embick, David. 2021. The motivation for roots in distributed morphology. *Annual Review of Linguistics* 7(1). 69–88. DOI: 10.1146/annurev-linguistics-040620-061341.

Endresen, Anna, Laura A Janda, Julia Kuznetsova, Olga Lyashevskaya, Anastasia Makarova, Tore Nesset & Svetlana Sokolova. 2012. Russian "purely aspectual" prefixes: not so "empty" after all? *Scando-Slavica* 58(2). 231–291. DOI: 10.1080/00806765.2012.740247.

Gehrke, Berit. 2008. Goals and sources are aspectually equal: Evidence from Czech and Russian prefixes. *Lingua* 118(11). 1664–1689. DOI: 10.1016/j.lingua.2007.05.011.

Gigafida 2.0. 2019. Corpus of Written Standard Slovenian. viri.cjvt.si/gigafida.

Gouskova, Maria & Jonathan David Bobaljik. 2022. The lexical core of a complex functional affix: Russian baby diminutive *-onok*. *Natural Language & Linguistic Theory*. 1–41. DOI: 10.1007/s11049-021-09530-1.

Gribanova, Vera. 2015. Exponence and morphosyntactically triggered phonological processes in the Russian verbal complex1. *Journal of Linguistics* 51(3). 519–561. DOI: https://doi.org/10.1017/S0022226714000553.

Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Kenneth Hale & S. Jay Keyser (eds.), *The view from building 20*, 111–176. Cambridge, MA: MIT Press.

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40(3–4). 225–276. DOI: 10.1515/tl-2014-0010.

HJP. 2006. *Hrvatski jezični portal [the Croatian language portal]*. Zagreb. https://hjp.znanje.hr.

Istratkova, Vyara. 2004. On multiple prefixation in Bulgarian. *Nordlyd* 32(2). 301–321. DOI: 10.7557/12.71.

Jabłońska, Patrycja. 2007. *Radical decomposition and argument structure*. University of Tromsø. (PhD dissertation).

Janić, Aleksandra. 2018. *Morfologija i semantika prideva na -(lj)iv u srpskom jeziku*. University of Kragujevac. (PhD Dissertation).

Jurgec, Peter. 2022. Tanja Mirtič & Marko Snoj (eds.). 1. slovenski pravorečni posvet. razprave ii. razreda 25. Ljubljana: Slovenska akademija znanosti in umetnosti, 2021. iSBN 978-961-268-077-0. 272pp. *Slovene Studies* 44(1). [Prepublication version], 99–107.

Klaić, Bratoljub. 2013. *Naglasni sustav standardnoga hrvatskoga jezika [The standard Croatian pitch-accent system]*. Božidar Smiljanić (ed.). Zagreb: Nova knjiga Rast.

Klajn, Ivan. 1998. Gramatički i leksikografski status glagolskih imenica od nesvršenih glagola. *Naučni sastanak slavista u Vukove dane* 27(2). 149–157.

Klajn, Ivan. 2002. *Tvorba reči u savremenom srpskom jeziku. prvi deo: slaganje i prefiksacija*. Belgrade / Novi Sad: The SASA Institute for Serbian Language / Matica srpska.

Klajn, Ivan. 2005. *Gramatika srpskog jezika*. Zavod za udžbenike i nastavna sredstva.

Klimek-Jankowska, Dorota & Joanna Błaszczak. 2022. The status of secondary imperfectivization in Polish: evidence from VP idioms. *Journal of Slavic Linguistics* 30(3). 1–19. https://ojs.ung.si/index.php/JSL/article/view/99.

Kovačević, Predrag, Stefan Milosavljević & Marko Simonović. 2024. Theme-vowel minimal pairs show argument structure alternations. *Journal of Linguistics*. 1–30. DOI: 10.1017/S0022226723000415.

Langston, Keith. 1997. Pitch accent in Croatian and Serbian: towards an autosegmental analysis. *Journal of Slavic linguistics*. 80–116. https://www.jstor.org/stable/24599009.

Łazorczyk, Agnieszka Agata. 2010. *Decomposing Slavic Aspect: the Role of Aspectual Morphology in Polish and Other Slavic Languages*. University of Southern California. (Doctoral dissertation). DOI: 10.9793/elsj1984.8.104.

Leskien, August. 1871. *Handbuch der altbulgarischen (altkirchenslavischen) sprache*. German. Weimar: Hermann Böhlau.

Ljubešić, Nikola & Filip Klubička. 2014. bs,hr,srWaC - web corpora of Bosnian, Croatian and Serbian. In *Proceedings of the 9th web as corpus workshop (WaC-9)*, 29–35. Gothenburg, Sweden: Association for Computational Linguistics. DOI: 10.3115/v1/W14-0405. https://aclanthology.org/W14-0405.

Ljubešić, Nikola, Peter Rupnik & Taja Kuzman. 2024a. *Croatian web corpus CLASSLA-web.hr 1.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1929.

Ljubešić, Nikola, Peter Rupnik & Taja Kuzman. 2024b. *Serbian web corpus CLASSLA-web.sr 1.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1931.

Lowenstamm, Jean. 2014. Derivational affixes as roots: phasal spell-out meets English stress shift. In Artemis Alexiadou, Hagit Borer & Florian Schäfer (eds.), *The syntax of roots and the roots of syntax*, 230–258. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199665266.003.0011.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In Alexis Dimitriadis, Laura Siegel, Clarissa Surek-Clark & Alexander Williams (eds.), *University of Pennsylvania working papers in linguistics*, vol. 4, 201–225. Philadelphia: University of Pennsylvania.

Marantz, Alec. 2013. Locality Domains for Contextual Allomorphy across the Interfaces. In Ora Matushansky & Alec P. Marantz (eds.), *Distributed Morphology today: Morphemes for Morris Halle*, 95–115. Cambridge, MA: The MIT Press. DOI: 10.7551/mitpress/9701.003.0008OpenthePDFLink.

Markova, Angelina. 2011. On the nature of Bulgarian prefixes: Ordering and modification in multiple prefixation. *Word Structure* 4(2). 244–271. DOI: 10.3366/word.2011.0013.

Marušič, Franc, Petra Mišmaš & Rok Žaucer. 2025. The (un)expectedly stacked prefixes in Slovenian. In Berit Gehrke, Denisa Lenertová, Roland Meyer, Daria Seres, Luka Szucsich & Joanna Zaleska (eds.), *Advances in formal Slavic linguistics 2022*, 311–345. Berlin: Language Science Press.

Marušič, Franc & Rok Žaucer. 2006. Slovenian from a typological perspective. (Sprachtypologie und Universalienforschung, vol. 56, no. 3.) Berlin: Akademie Verlag, 2003. (Language Typology and Universals.) *Journal of Slavic Linguistics* 14(1). 123–160. https://www.jstor.org/stable/24599540.

Marušič, Franc Lanko, Rok Žaucer, Petra Mišmaš, Boban Arsenijević, Marko Simonović, Stefan Milosavljević, Katarina Gomboc Čeh & Jelena Simić. 2022. *Database of the Western South Slavic verb HyperVerb 1.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1683.

Marvin, Tatjana. 2002. *Topics in the stress and syntax of words*. Cambridge, MA: Massachusetts Institute of Technology. (Doctoral dissertation). http://mitwpl.mit.edu/catalog/marv01/.

Matushansky, Ora. 2024. Thematic non-uniformity of Russian vocalic verbal suffixes. *Glossa: a journal of general linguistics* 9(1). DOI: 10.16995/glossa.8571.

Miklosich, Franz. 1874. *Vergleichende grammatik der slavischen sprachen: Formenlehre*. German, vol. 2. Wien: Wilhelm Braumüller.

Milićević, Nataša. 2004. The lexical and superlexical verbal prefix iz-and its role in the stacking of prefixes. *Nordlyd* 32(2). 279–300. DOI: https://doi.org/10.7557/12.70.

Milosavljević, Stefan. 2022. The delimitative prefix *po-*, durative adverbials and Slavic aspectual composition. In Annie Holtz, Iva Kovač & Rasmus Puggaard-Rode (eds.), *Proceedings of console xxx*, 17–42. Leiden: Leiden University Centre for Linguistics.

Milosavljević, Stefan. 2023a. Simple imperfective verbs, the sequence of similar events interpretation, and Slavic aspectual composition. In Petr Biskup, Marcel Börner (né Guhl), Olav Mueller-Reichau & Iuliia Shcherbina (eds.), *Formal advances in Slavic linguistics 2021*. Berlin: Language Science Press. DOI: https://doi.org/10.5281/zenodo.10123645.

Milosavljević, Stefan. 2023b. *Specification of event duration and aspectual composition in Slavic*. University of Graz. (PhD Dissertation).

Milosavljević, Stefan & Boban Arsenijević. 2022. What differentiates Serbo-Croatian verbal theme vowels: content or markedness? *Glossa: a journal of general linguistics* 7(1). DOI: https://doi.org/10.16995/glossa.8535.

Milosavljević, Stefan, Predrag Kovačević & Marko Simonović. 2025. Eliminating redundancy in internal argument selection: the division of labor between prefixes and 'flavors of v'. Submitted to *Linguistics*.

Milosavljević, Stefan, Petra Mišmaš, Marko Simonović, Boban Arsenijević, Katarina Gomboc Čeh, Franc Lanko Marušič, Jelena Simić & Rok Žaucer. 2023. *Database of the western south slavic verb HyperVerb – derivation*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1855.

Mišmaš, Petra, Marko Simonović, Boban Arsenijević, Stefan Milosavljević, Agniszka Bedkowska-Kopczyk & Petya Rogić. 2020. Plavanje da, zaplavanje ne? Obrazilo -je v glagolnikih in drugih okoljih v vseslovanskem kontekstu. In Matej Šekli & Lidija Rezoničnik (eds.), *Slovenski jezik in književnost v srednjeevropskem prostoru*, 221–234. Ljubljana: Zveza društev Slavistično društvo Slovenije. https://zdsds.si/tiskovina/2695/.

Nikolić, Miroslav. 2000. *Obratni rečnik srpskoga jezika [Serbian reverse dictionary]*. Beograd/Novi Sad: Institut za srpski jezik SANU/Matica srpska.

Oltra-Massuet, Isabel. 1999. *On the notion of theme vowel: a new approach to Catalan verbal morphology*. MIT. (MA thesis). https://dspace.mit.edu/bitstream/handle/1721.1/9532/43893506-MIT.pdf.

Oltra-Massuet, Isabel. 2020. Conjugation class. In *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. DOI: 10.1093/acrefore/9780199384655.013.545.

Panagiotidis, Phoevos. 2020. On the nature of roots: content, form, identification. *Evolutionary Linguistic Theory* 2(1). 56–83. DOI: 10.1075/elt.00015.pan.

Polančec, Jurica. 2018. Orphan secondary imperfectives in Croatian. *Suvremena lingvistika* 44(85). 113–138. DOI: 10.22210/suvlin.2018.085.06.

Quaglia, Stefano, Marko Simonović, Svitlana Antonyuk-Yudina & Boban Arsenijević. 2022. Allomorphy, morphological operations and the order of Slavic verb-prefixes. *Journal of Slavic Linguistics* 30(FASL 29 extra issue). 1–15. DOI: 10.1353/jsl.2022.a923080.

Ramchand, Gillian. 2004. Time and the event: The semantics of Russian prefixes. *Nordlyd* 32(2). 323–361. DOI: 10.7557/12.72.

Ramchand, Gillian Catriona. 2008. Perfectivity as aspectual definiteness: time and the event in Russian. *Lingua* 118(11). 1690–1715. DOI: 10.1016/j.lingua.2007.03.008.

Rappaport Hovav, Malka. 2014. Building scalar changes. In Artemis Alexiadou, Hagit Borer & Florian Schäfer (eds.), *The syntax of roots and the roots of syntax*, 259–281. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199665266.003.0012.

Rappaport Hovav, Malka. 2017. Grammatically relevant ontological categories underlie manner/result complementarity. In Noa Brandel (ed.), *Proceedings of iatl 32*, 77–98. Cambridge, MA: MITWPL.

Rappaport Hovav, Malka. 2021. Uncovering the scale. Presented at WCCFL 39, April 11, 2021.

Rappaport Hovav, Malka & Beth Levin. 2010. Reflections on manner/result complementarity. In Edit Doron, Malka Rappaport Hovav & Ivy Sichel (eds.), *Lexical semantics, syntax, and event structure*, 21–38. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199544325.003.0002.

Romanova, Eugenia. 2004. Superlexical versus lexical prefixes. *Nordlyd* 32(2). DOI: 10.7557/12.69.

Salvadori, Justine & Richard Huyghe. 2023. Affix polyfunctionality in French deverbal nominalizations. *Morphology* 33. 1–39. DOI: 10.1007/s11525-022-09401-4.

Schütze, Carson T. 1996. *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. Chicago, IL: University of Chicago Press.

Šekli, Matej. 2010. Oblikotvorje in jakostni naglas glagola v knjižni slovenščini [Morphology and dynamic stress in literary Slovenian]. In Vera Smole (ed.), *Slovanstvo v slovenskem jeziku, literaturi in kulturi*, 139–160. Ljubljana: Znanstvena založba Filozofske fakultete.

Simonović, Marko. 2020. Categories, root complexes and default stress. *Linguistica* 60(1). 103–117. DOI: 10.4312/linguistica.60.1.103-117.

Simonović, Marko. 2022a. Derivational affixes as roots across categories. *Journal of Slavic linguistics* 30(2). 195–233. DOI: 10.1353/jsl.2022.a909904.

Simonović, Marko. 2022b. Neo-Štokavian deverbal je-nominalisations contain passive participles. *Journal of Slavic Linguistics* 30(FASL 29 extra issue). 1–13. http://ojs.ung.si/index.php/JSL/article/view/91.

Simonović, Marko & Boban Arsenijević. 2014. Regular and honorary membership: On two kinds of deverbal nouns in Serbo-Croatian. *Lingue e Linguaggio* 13(2). 185–210. DOI: 10.1418/78407.

Simonović, Marko & Boban Arsenijević. 2020. Syntax predicts prosody: multipurpose morphemes in Serbo-Croatian. In Franc Marušič, Petra Mišmaš & Rok Žaucer (eds.), *Advances in formal Slavic linguistics 2017*, 277–304. Berlin: Language Science Press. DOI: 10.5281/zenodo.3764865.

Simonović, Marko & René Kager. 2023. Avoiding stress on non-lexical material in nouns and verbs: predictable verb prosody in Serbo-Croatian stress standard varieties. *Linguistics* 61(5). 1105–1130. DOI: 10.1515/ling-2020-0111.

Simonović, Marko & Predrag Kovačević. 2022. Possessive, kind and not so kind: the different uses of the adjectival -ov in Serbo-Croatian. *Godišnjak Filozofskog fakulteta u Novom Sadu* 47(3). 87–109. DOI: 10.19090/gff.2022.3.87-109.

Simonović, Marko, Stefan Milosavljević & Boban Arsenijević. 2023. Serbo-Croatian secondary imperfectivisers consist of theme vowels. *Journal of Slavic Linguistics* 31(FASL 30 extra issue). http://ojs.ung.si/index.php/JSL/article/view/178.

Simonović, Marko & Petra Mišmaš. 2020. √ov is in the air: The extreme multifunctionality of the Slovenian affix *ov*. *Linguistica* 60(1). 83–102. DOI: 10.4312/linguistica.60.1.83-102.

Simonović, Marko & Petra Mišmaš. 2022. Lowest theme vowels or highest roots? An 'unaccusative' theme-vowel class in Slovenian. *Glossa: a journal of general linguistics* 7(1). 1–31. DOI: 10.16995/glossa.5809.

Simonović, Marko & Petra Mišmaš. 2023a. The interaction between theme vowels and secondary imperfectives in Slovenian: the curious case of *ava* and *eva*. *Journal of Slavic linguistics* 30(FASL extra issue). 1–21. https://ojs.ung.si/index.php/JSL/article/view/173.

Simonović, Marko & Petra Mišmaš. 2023b. Theme-vowel class indeterminacy and root allomorphy in Slovenian. *Glossa: a journal of general linguistics* 47(1). 1–38. DOI: 10.16995/glossa.8550.

Simonović, Marko, Petra Mišmaš, Stefan Milosavljević, Boban Arsenijević, Katarina Gomboc Čeh, Franc Marušič & Rok Žaucer. 2024. How departicipial are "ʟ-participle" nominalisations in Western South Slavic. In Berit Gehrke, Denisa Lenertová, Roland Meyer, Daria Seres, Luka Szucsich & Joanna Zaleska (eds.), *Advances in formal Slavic linguistics 2022*. Berlin: Language Science Press. DOI: XX.

Snoj, Marko. 2009. *Slovenski etimološki slovar [Slovenian Etymological Dictionary]*. Ljubljana: Modrijan.

SSKJ. 2014. *Slovar slovenskega knjižnega jezika, druga, dopolnjena in deloma prenovljena izdaja [the dictionary of standard Slovenian language, second edition]*. Ljubljana: Inštitut za slovenski jezik Frana Ramovša ZRC SAZU. www.fran.si.

Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. *Nordlyd* 36(1). 1–6. DOI: 10.7557/12.213. https://septentrio.uit.no/index.php/nordlyd/article/view/213.

Štarkl, Ema, Marko Simonović, Stefan Milosavljević & Boban Arsenijević. 2024. The Western South Slavic verbal suffix nV/ne is a diminutive affix with a theme vowel. In Berit Gehrke, Denisa Lenertová, Roland Meyer, Daria Seres, Luka Szucsich & Joanna Zaleska (eds.), *Formal advances in Slavic linguistics 2022*. Berlin: Language Science Press.

Stojanović, Milica. 2016. *Poliprefiksirani glagoli u srpskom jeziku*. University of Belgrade. (Doctoral dissertation).

Stojković, Jelena & Marko Simonović. Forthcoming. Allomorphy reveals the trace of a (different) theme vowel. In Daniel Gleim, Andrew Nevins, Sören E. Tebay & Eva Zimmermann (eds.), *Traces in morphology and phonology* (Linguistische Arbeiten Berichte 9X), 1–20. Leipzig: Universität Leipzig.

Stump, Gregory T. & Raphael A. Finkel. 2013. Principal parts. In Balthasar Bickel & Johanna Nichols (eds.), *Morphological typology: from word to paradigm* (Cambridge Studies in Linguistics), 9–39. Cambridge: Cambridge University Press.

Svenonius, Peter. 2004a. Slavic prefixes and morphology: an introduction to the Nordlyd volume. *Nordlyd* 32(2). 177–204. DOI: 10.7557/12.67.

Svenonius, Peter. 2004b. Slavic prefixes inside and outside VP. *Nordlyd* 32(2). 206–253. DOI: 10.7557/12.68.

Taraldsen Medová, Lucie & Bartosz Wiland. 2019. Semelfactives are bigger than degree achievements: the nanosyntax of Czech and Polish semelfactive and degree achievement verb stems. *Natural Language & Linguistic Theory* 37(4). 1463–1513. DOI: 10.1007/s11049-018-9434-z. https://doi.org/10.1007/s11049-018-9434-z.

Tatevosov, Sergei. 2008. Intermediate prefixes in Russian. In Andrei Antonenko, Christina Bethin & John Baylin (eds.), *Proceedings of the annual workshop on Formal Approaches to Slavic Linguistics 16. The Stony Brook meeting*, vol. 16, 423–445. Ann Arbor, MI: Michigan Slavic Publications.

Toporišič, Jože. 2000. *Slovenska slovnica*. Maribor: Založba Obzorja.

Toporišič, Jože. 2004. *Slovenska slovnica [Slovenian grammar], 4th edition*. Maribor: Obzorja.

Vanden Wyngaerd, Guido, Michal Starke, Karen De Clercq & Pavel Caha. 2020. How to be positive. *Glossa: a journal of general linguistics* 5. DOI: 10.5334/gjgl.1114.

Wiltschko, Martina. 2014. *The universal structure of categories: Towards a formal typology*. Cambridge: Cambridge University Press. DOI: 10.1353/lan.2016.0065.

Wood, Jim & Alec Marantz. 2017. The interpretation of external arguments. In Roberta D'Alessandro, I Franco & Ángel Gallego (eds.), *The verbal domain*, 255–278. Oxford: Oxford University Press. https://lingbuzz.net/lingbuzz/002487/v2.pdf.

Žaucer, Rok. 2002. *The role of verbal prefixes in Slavic: evidence from Slovenian locative denominal verbs*. University of Ottawa. (MA thesis). http://hdl.handle.net/10393/6153.

Žaucer, Rok. 2009. *A VP-internal/resultative analysis of 4 "VP-external" uses of Slavic verbal prefixes*. Ottawa, ON: University of Ottawa. (Doctoral dissertation). http://hdl.handle.net/10393/29850.

Žaucer, Rok. 2013. A case for result-modifying prefixes. In Alexander Podobryaev (ed.), *Formal Approaches to Slavic Linguistics 20: the Second MIT Meeting*, 290–304. Ann Arbor, MI: Michigan Slavic Publications.

Zec, Draga. 2018. *Sonority constraints on prosodic structure*. Routledge.

Zinova, Yulia. 2021. *Russian verbal prefixation: a frame semantic analysis*. Berlin: Language Science Press. DOI: 10.5281/zenodo.4446717.

Hyperspacing the Verb

Set blurb on back with \BackBody{my blurb}