

**Insights from the interfaces:
Morphological processes as string transductions**

A Dissertation Presented

by

Andrija Petrovic

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Linguistics

Stony Brook University

August 2023

Stony Brook University

The Graduate School

Andrija Petrovic

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Mark Aronoff – Dissertation Advisor
Distinguished Professor, Department of Linguistics

Jeffrey Heinz - Chairperson of Defense
Professor, Department of Linguistics

Thomas Graf
Associate Professor, Department of Linguistics

John Frederick Bailyn
Professor, Department of Linguistics

Adam Jardine
Associate Professor, Department of Linguistics
Rutgers University

This dissertation is accepted by the Graduate School

Celia Marshik
Dean of the Graduate School

Abstract of the Dissertation

**Insights from the interfaces:
Morphological processes as string transductions**

by

Andrija Petrovic

Doctor of Philosophy

in

Linguistics

Stony Brook University

2023

This dissertation brings together theoretical, empirical, and computational approaches to examine morphology and its interfaces. Processes like morphologically and syntactically conditioned epenthesis present difficulties for theories of morphology that are either syntax-based or fully merged with phonological computation; to overcome them, I provide a principled way to discuss representations and computation across modules using model theory and logical characterizations. I define morphological processes as string transductions over flattened syntactic trees, and offer a formalization of the syntax-morphology and the morphology-phonology interface, all the while bridging the gap between syntax-based and word-and-paradigm-based approaches to morphology.

Much prior work in morphology has been syntax based, and therefore relied on tree-based representations, or did not have a formalization of the syntax-morphology interface. On the other hand, languages of the world abound with paradigms that involve featural and segmental alternations which cannot be explained using the regular phonology of the language. My dissertation addresses these observations, uniting them into a representational solution that retains the computational complexity of morpho(phono)logical mappings, which have been observed to be at most regular (Johnson, 1972; Karttunen et al., 1992; Chandlee, 2017) – i.e. computations with a fixed memory. I apply this insight to a number of case studies like Serbo-Croatian morphological consonant epenthesis, variable morphotactics in Georgian, and total reduplication in Indonesian. Additionally, I show that the inherently hierarchical nature of the proposed architecture allows the integration of well-understood grammatical principles such as inflectional class assignment algorithms and the use of leading forms/principal parts. In the case study of Serbo-Croatian declensions, I show how such predictable processes can be included in a wider, algorithmic interpretation of nominal inflection as logical transductions on strings.

I utilize Boolean Monadic Recursive Schemes (BMRS; Bhaskar et al., 2020; Chandlee & Jardine, 2021), a logical formalism with an intuitive syntax, to provide logical descriptions of

phenomena pertaining to the morphological module. BMRSs are appropriate for modeling morphological processes, as they can intensionally represent morphological substance and generalizations, much like theories of realizational morphology do, while retaining the computationally restrictive nature of such processes.

Table of Contents

List of Tables	vi
List of Figures	vii
Acknowledgments	viii
1 Introduction	1
2 Architecture of the system: basic assumptions.....	4
3 Morphological epenthesis: Serbo-Croatian	8
3.1 Serbo-Croatian neuter noun inflection.....	10
3.1.1 Vowel-final neuter stems as a listed (sub)class	13
3.1.2 <i>bure</i> -type neuter stems as C-final stems with truncation in the nominative singular.....	14
3.1.3 <i>bure</i> -type neuter stems as C-final stems with CV extenders.....	15
3.2 Morphologically conditioned coronal epenthesis	16
3.3 Analysis	20
3.3.1 Boolean Monadic Recursive Schemes (BMRS).....	21
3.3.2 Structure of a BMRS system	22
3.3.3 Serbo-Croatian <i>t</i> -epenthesis: A BMRS analysis.....	23
3.4 Conclusion	37
4 Morphological epenthesis: Romance.....	38
4.1 [u]-epenthesis in Catalan	38
4.2 [z]-epenthesis in Brazilian Portuguese	41
4.3 [s]-epenthesis in Spanish	43
4.4 Conclusion	45
5 Leading forms: cues for inflectional class assignment.....	46
5.1 Class membership assignment in Serbo-Croatian	46
5.2 Serbo-Croatian inflectional classes.....	47
5.3 Leading forms	50
5.4 Deficient entries	51
5.5 Modeling and formalization	52
5.6 Vowel fronting in Serbo-Croatian	67
5.7 Discussion	72
6 Syntax-morphology interface processes	74
6.1 Formalization	74
6.2 Variable morphotactics	76
6.3 Total reduplication	81
6.4 Syntactically conditioned segmental processes	84
7 Conclusion.....	94
References	96

List of Tables

Table 1. Inflection of SC <i>zavod</i> ‘institute’ (MAS.) and <i>selo</i> ‘village’ (NEU.).....	10
Table 2. Inflection of SC <i>ḡekitḡ</i> ‘hammer’ (MAS.) and <i>poḡe</i> ‘field’ (NEU.).....	11
Table 3. Inflection of SC <i>bure</i> ‘barrel’ (NEU.).....	12
Table 4. Values of some input predicates for the segments of SC <i>bure</i> ‘barrel’	26
Table 5. Input and output strings of the dative singular form of SC <i>bure</i> ‘barrel’	27
Table 6. Input and output strings of the instrumental plural form of SC <i>bure</i> ‘barrel’	31
Table 7. Input and output strings of the instrumental singular form of SC <i>selo</i> ‘village’	33
Table 8. Input and output strings of the instrumental plural form of SC <i>poḡe</i> ‘field’	33
Table 9. Input and output strings of Catalan [gosus] ‘dogs’	40
Table 10. Input and output strings of Portuguese [sofaziḡu] ‘sofa.DIM’	42
Table 11. Input and output strings of Spanish [gatito] ‘cat.DIM’	43
Table 12. Input and output strings of Spanish [amorsito] ‘love.DIM’	44
Table 13. Inflectional classes of Serbo-Croatian nouns	48
Table 14. Input and output strings of the genitive singular form of SC <i>zēna</i> ‘woman’	54
Table 15. Input and output strings of the instrumental plural form of <i>zēna</i> ‘woman’	63
Table 16. Input and output strings of the genitive singular form of SC <i>anime</i> ‘anime’	66
Table 17. Input (alternative) and output strings of the dative singular form of SC <i>zēna</i> ‘woman’	67
Table 18. Exponent realization step for the instrumental singular form of SC <i>poḡe</i> ‘field’	69
Table 19. Morphophonological processes step for the instrumental singular form of SC <i>poḡe</i> ‘field’	69
Table 20. Present indicative paradigm for some Georgian verb types.....	76
Table 21. Input and output structures for the first person singular form of Georgian <i>ban</i> ‘bathe’	78
Table 22. Input and output structures for the third person singular form of Georgian <i>ban</i> ‘bathe’	78
Table 23. Input and output structures for the first person plural form of Georgian <i>ban</i> ‘bathe’ ..	79
Table 24. Input and output structures for the third person plural form of Georgian <i>ban</i> ‘bathe’ ..	80
Table 25. Input and output structures for the plural form of Indonesian <i>buku</i> ‘book’	82
Table 26. Input and output structures for Russian <i>u nevo</i> ‘at him’	85
Table 27. Input and output structures for Russian <i>u evo brata</i> ‘at his brother’s’	85
Table 28. Input and output structures for German <i>zum</i> [tsu:m] (< zu dem)	92

List of Figures

Figure 1. Architecture of the proposed model.....	5
Figure 2. Syntactic tree for Russian <i>u evo brata</i> ‘at his brother’s’ and its yield	6
Figure 3. Decision tree for declension assignment of Serbo-Croatian stems.....	51
Figure 4. Syntactic tree for Russian <i>u evo brata</i> ‘at his brother’s’ and its yield as input to morphology	75
Figure 5. Syntactic structure of the prepositional phrase in (124).	90

Acknowledgments

Thank you

Mark
Thomas Jeff
 John
 Adam
 Grev
 Dunstan
 Boban
 Doug Michael
Hossep
 Dakotah
 Sophie
 Grace Sandeep
 Veronica
Bogdan Ivan
 Ana
 Mom and Dad
 Nina

This would not have been possible without your support and guidance. You have all been there along the way to advise, suggest, listen, and understand. Thank you for reading, writing, asking, attending, supporting, sharing, recommending, socializing, playing devil's advocate, planning, improvising, optimizing, moving, sticking around.

This was a journey, and I love to travel.

1 Introduction

This dissertation represents a research project focused on better understanding the nature of the interfaces of morphology with phonology and syntax and the formal nature of the morphological module more generally. In this work I focus on the fact that a number of processes are that are usually thought of as falling in the domain of phonology are in fact conditioned by factors related to morphology and syntax.

In order to account for this, I offer a model that includes morphology as an independent module, taking syntactic structures as input and producing phonological words as output. In this model, morphological processes are represented as string-to-string transductions. The model is based on the observation that morphological mappings can universally, cross-linguistically, be understood as regular functions. In mathematics, a function is regular if the memory required for the computation is bounded by a constant, regardless of the size of the input. One contribution of this dissertation is to show that a morphological system can be generated by a program of only (at most) regular functions, and that variability in such systems follows the same restrictions. Additionally, as predictable processes are preferable to listed allomorphy – a generalization that applies to novel items – I use various case studies to show different processes can be reanalyzed with morphosyntactic and lexical conditioning.

In Chapter 2 I outline and visualize the overall architecture assumed in the dissertation, and briefly discuss what that choice of underlying structures and representations entails for the analyses that are presented in the chapters that follow. Chapter 3 focuses on a morphophonological consonant insertion process in Serbo-Croatian that makes crucial reference to morphological structure, morphosyntactic features, and inflectional class – and, above all, is not necessarily

phonologically optimizing. The fact that we cannot effectively analyze this process in a single-level, purely phonological framework that does not allow for a direct interaction with representations from the morphological module makes it necessary to look deeper into the nature of the morphology-phonology interface, and the formal properties of morphophonological processes. The chapter also introduces the formalism, Boolean Monadic Recursive Schemes, used throughout the dissertation. In Chapter 4 I identify insightful instances of interface processes (morphologically conditioned epenthesis) in Romance languages, proposing ways they can be formalized as logical string transductions. The central idea of this work – that different kinds of morphological processes can be formalized over strings – is further expanded in Chapter 5, which proposes that the intrinsic hierarchical structure of the formalism enables the incorporation of familiar grammatical principles, such as algorithms for assigning inflectional classes, and the utilization of readily available surface forms for different realizations of morphosyntactic features. Furthermore, the input string to the morphological module is understood to be the yield of a syntactic tree structure; benefits of such an approach, where input syntactic trees are flattened pre-morphology, are explored in Chapter 6, where I examine processes from different languages that have crucial syntactic conditions on the insertion, deletion, and transformation of phonological material.

Much of the research presented here was originally centered on exploring the specificities of Serbo-Croatian inflection; however, the broad aims of this dissertation go beyond analyzing how the Serbo-Croatian morphological system works. The proposed analysis, informed by empirical generalizations, and computational and mathematical theory, makes predictions about the treatment of existing and novel words in the systems that were examined. The goal of this work

is to construct a computational account that combines linguistic generalizations and mathematical concepts, delivering a formal characterization of morphological processes and their nature.

2 **Architecture of the system: basic assumptions**

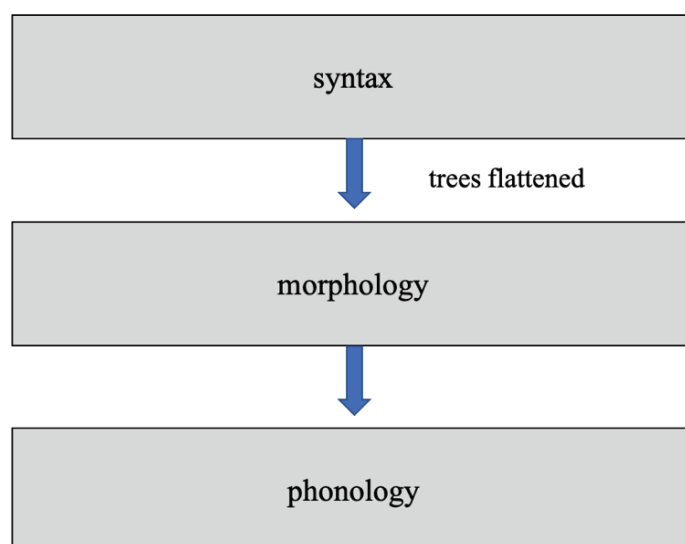
This dissertation focuses on the interfaces of morphology with syntax and phonology – morphological processes can be sensitive to lexical, syntactic, phonological, and purely morphological representations. In theories like Distributed Morphology (DM; Halle & Marantz, 1993) the morphology-syntax interface is directly addressed, as such models typically operate on binary trees which are the output of syntax. More specifically, works like Embick (2010), Arregi & Nevins (2012) and Merchant (2015) examine how syntactic structure affects different morphological processes, seeking to define the representations and contexts that are necessary to capture locality constraints on contextual allomorphy, and spell-out more generally (DM’s Vocabulary Insertion). On the other hand, the nature of the morphosyntax-phonology interface has been extensively explored in, for instance, in Kiparsky (1982), Inkelas & Zec (1990), Scheer (2011); more recently, Sande et al. (2020) and Kalin (2020) have been among those that directly deal with the division of labor between representations and processes, and the order of operations needed to derive the target forms. The overarching generalizations that stem from this line of work is that morphology receives its input from the syntax, and that it has to apply before phonology. In this chapter, I establish the architecture that is assumed throughout this dissertation for the modeling of different morphological processes.

To the generalizations outlined above, I add a computational one – unlike syntax, and like phonology (Johnson, 1972; Kaplan & Kay, 1994; Heinz, 2018), morphology has been observed to be at most REGULAR on the Chomsky hierarchy (Chomsky, 1959; Karttunen et al., 1992; Chandlee, 2017). Historically, this distinction between regular and non-regular grammars is what motivated the need for a different theory of syntax to that of phonology: formalizations of syntactic processes

need to be able to capture, for instance, center embedding, while phonological processes of that kind are unattested. Patterns like iterable nested dependencies are unattested in morphology as well; I thus argue that formalisms like DM, which are defined over binary trees, and in which the syntax-morphology interface is therefore treated as a tree transducer, are overpowered.

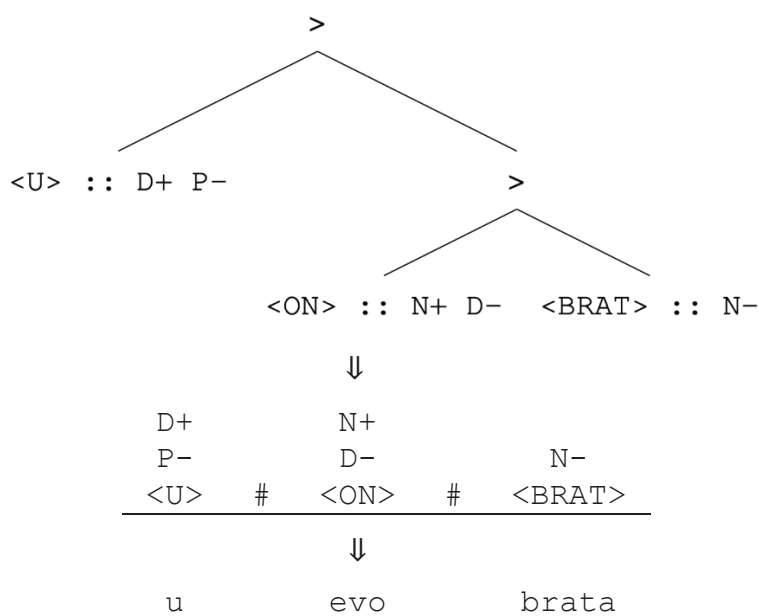
In line with Ermolaeva & Edmiston (2018), I instead propose that morphology be restricted to operating over strings (cf. Dolatian, 2020, where morphology is formalized over tree structures, but note that these are not syntactic binary trees). A morphological module conceptualized in this way correctly predicts that morphological mappings can be modeled with regular string languages, and the morphological component is thus treated as a regular relation. The architecture assumed in this dissertation is visualized in Figure 1. I abstract away from the question of cyclicity (i.e. whether the output of a component can recursively form the input to a derivational component that has already taken place), but see Dolatian (2020) for how an interactionist or cyclic architecture can be established in order to handle phenomena on the morphology-phonology interface.

Figure 1. Architecture of the proposed model.



So, as opposed to the assumptions of tree-based frameworks like DM, the flattening of the derivation is pushed to take place above the morphological module. I further assume that syntactic structures are Minimalist Grammar trees (MG; Stabler, 1997), which are built by the percolation and checking of category lexical features by selector lexical features (Merge), and that of licenser and licensee lexical features (Move). A morphological input string is then the direct yield of a given syntactic tree, formed by a left-to-right traversal and concatenation of the encountered leaf nodes. A Russian example (further examined in Chapter 6) is illustrated in Figure 2:

Figure 2. Syntactic tree for Russian *u evo brata* ‘at his brother’s’ and its yield



The input string to the morphological module is thus made up of unrealized syntactic terminals (<U>, <ON>, and <BRAT> in Figure 2), their corresponding lexical features (P-, D+, D-, N+ and N- in Figure 2), and boundary symbols (e.g. the word boundary #); in the analyses that follow I will also use symbols like the stem boundary, morphosyntactic features (like case, number, and gender), and other diacritics.

So, a string-based account will be developed in this dissertation to capture patterns that require reference to syntactic information and word-internal structure. The architecture I assume restricts the morphological component to working on strings (post-flattening), limiting morphology to regular string languages. In addition, the formalism that is used throughout (Boolean Monadic Recursive Schemes, introduced in Chapter 3.3) brings in a restriction to subsequential functions – a strict sub-class of regular functions (Bhaskar et al., 2020) – in terms of mappings from input to output strings. As shown in Chandlee & Jardine (2021), this formalism describes exactly the subsequential functions, and correctly excludes patterns unattested in phonology. The analyses that follow extend this observation to morphology, allowing for the introduction of further restrictions (e.g. the analysis of variable morphotactics as involving moving affixes tied to a fixed distance from the stem, in Chapter 6.2).

Ultimately, the purpose of this account is to define the necessary conditions on the realization of morphological processes, restricted by their computational complexity. The linguistic contributions include evidence for lexical, syntactic, morphological, and phonological conditions on processes like non-canonical epenthesis, variable morphotactics, and reduplication; through formalizations of such processes, I identify new generalizations about exponent realization, allomorphy, inflectional class membership, and eliminating overreliance on listedness. The architecture outlined in this chapter, the formalism introduced in the following, and the specific analyses developed throughout the dissertation, help facilitate stating these generalizations.

3 Morphological epenthesis: Serbo-Croatian

The purpose of this chapter is to show that the stem allomorphs that appear in Serbo-Croatian¹ neuter noun inflection are predictable. These neuter noun stem allomorphs reflect morphologically conditioned consonant insertion, and their distribution is completely predictable from the phonological shape of the stem and morphosyntactic properties of the lexeme. Specifically, I claim that a *t* is inserted after *e*-final stems to create consonant-final stems before attachment of an overt case suffix. Existing analyses of Serbo-Croatian nominal inflection do not take this insight into account; the formalization of this process, in the form of a system of logical transductions, is a novel contribution of Petrovic (2023), on which this chapter is based.

In Serbo-Croatian, nouns belong to one of three inflectional paradigms – Class I consists of masculine and neuter nouns, Class II comprises feminine (and masculine) nouns that end in *-a* in the nominative singular; feminine nouns that do not receive an overt suffix in the nominative singular belong to Class III. Examples of nouns belonging to each class, with nominative and genitive singular forms, are given in (1).

(1)	Class I:	zavod	zavod-a
		institute.NOM.SG.MAS	institute.GEN.SG.MAS
		sel-o	sel-a
		village.NOM.SG.NEU	village.GEN.SG.NEU
		bure	buret-a
		barrel.NOM.SG.NEU	barrel.GEN.SG.NEU
	Class II:	žen-a	žen-e
		woman.NOM.SG.FEM	woman.GEN.SG.FEM

¹ The language is here referred to, and treated as, one (pluricentric) language, following the Declaration on the Common Language (<http://jezicinalizmi.com/deklaracija/>).

Class III:	refj word.NOM.SG.FEM	refj-i word.GEN.SG.FEM
------------	-------------------------	---------------------------

The forms of interest here are those like *buret-a* ‘barrel.GEN.SG.NEU’ above, where the nominative singular ends in *e*. The claim of this chapter is that this is an effect of morphologically conditioned consonant insertion – as most noun stems in the language are consonant-final, vowel-final stems are repaired by inserting a consonant, namely *t*. The present approach avoids stipulating listed stems (pairs of *e*-final and *t*-final stem allomorphs for each relevant neuter stem), offering instead an account of the pattern as a result of a predictable process. Assuming unpredictable stem allomorphs – one with *t*, the other without – would reduce the phenomenon to an accident.

I demonstrate that this is an algorithmic process, part of a wider algorithmic process of noun inflection. Using Boolean Monadic Recursive Schemes (BMRSs; Bhaskar et al., 2020; Chandlee & Jardine, 2021), Serbo-Croatian nominal inflection is analyzed as a system of logical transductions on strings. Given a stem (which has a phonological form) and a set of morphosyntactic features (which do not) as an input string, the system produces output strings of segments that are fully inflected words. This is achieved through a hierarchical ordering of more specific and less specific blocking and licensing structures, much akin to realization rules that are intrinsically ordered by Pāṇini’s principle (= Elsewhere Condition) in works like Stump (2001). In this way, the process of nominal inflection and the pertinent generalizations are expressed explicitly. With BMRSs, the data can be formally accounted for in a direct and parsimonious way, capturing both intensional (linguistically significant) and extensional (computational) generalizations.

The chapter is organized as follows. Section 3.1 introduces the relevant Serbo-Croatian data and proposes a novel understanding of the morphological structure of inflected neuter nouns

in the language. Other existing approaches are briefly discussed in subsections 3.1.1–3.1.3. In section 3.2 I outline the approach adopted in this work, which is then formalized in section 3.3. Subsection 3.3.1 introduces the BMRS framework, with more details regarding the structure of a BMRS system provided in 3.3.2; finally, the analysis is provided in 3.3.3.

3.1 Serbo-Croatian neuter noun inflection

Within the Serbo-Croatian (SC) nominal system, a noun paradigm comprises inflected variants of one noun in two numbers and seven cases. In Noun Class I (i.e. masculine and neuter nouns), vowel-initial case endings are added onto consonant-final stems. The inflectional paradigms of two such nouns are given in Table 1. Animacy distinctions are omitted for the sake of simplicity; here I default to the inanimate pattern, where the accusative singular is syncretic with the nominative singular, and the neuter paradigm plural forms are not ineffable.

Table 1. Inflection of SC *zavod* ‘institute’ (MAS.) and *selo* ‘village’ (NEU.)

	MASCULINE			NEUTER	
	SG.	PL.		SG.	PL.
NOM.	zavod	zavod-i		sel-o	sel-a
GEN.	zavod-a	zavod-aa		sel-a	sel-aa
DAT.-LOC.	zavod-u	zavod-ima		sel-u	sel-ima
ACC.	zavod	zavod-e		sel-o	sel-a
VOC.	zavod-e	zavod-i		sel-o	sel-a
INS.	zavod-om	zavod-ima		sel-om	sel-ima

As Table 1 shows, neuter nouns inflect similarly to masculine nouns. Their stems receive a lot of the same case endings (represented in shaded cells) – the difference being the nominative

suffix in both singular and plural, and the accusative and vocative, which are syncretic with the nominative for neuter nouns. As the inflectional paradigms of feminine stems greatly differ from this pattern, masculine and neuter stems are usually seen as belonging to the same inflectional class (Barić et al., 1997; Klajn, 2005).

The inflection of masculine and neuter nominals is affected by the final consonant of the stem: posterior coronal consonants *j, ʎ, ɲ, ʃ, dʒ, tɕ, dʑ, ʒ*, and the dental affricate *ʈ*, in this position yield suffix-initial *e* instead of *o*. These coronal consonants trigger fronting of the suffix-initial vowel, and I will henceforth refer to this class of consonants as FRONTING CONSONANTS (FC).²

Table 2 shows the declension of nouns with FC-final stems, exemplified by the nouns *ʃekite* ‘hammer’ (MAS.) and *poʎe* ‘field’ (NEU.), with respective case endings added to the stems in order to form different inflected forms. The shaded cells contain the forms with the fronted suffix-initial vowel.

Table 2. Inflection of SC *ʃekite* ‘hammer’ (MAS.) and *poʎe* ‘field’ (NEU.)

	MASCULINE			NEUTER	
	SG.	PL.		SG.	PL.
NOM.	ʃekite	ʃekite-i		poʎ-e	poʎ-a
GEN.	ʃekite-a	ʃekite-aa		poʎ-a	poʎ-aa
DAT.-LOC.	ʃekite-u	ʃekite-ima		poʎ-u	poʎ-ima
ACC.	ʃekite	ʃekite-e		poʎ-e	poʎ-a
VOC.	ʃekite-e ³	ʃekite-i		poʎ-e	poʎ-a
INS.	ʃekite-em	ʃekite-ima		poʎ-em	poʎ-ima

² These consonants are termed *palatal* in all traditional Serbo-Croatian literature (e.g. Stevanović, 1969), even though they do not all have a strictly palatal place of articulation. I assume that, in terms of the morphophonological effect of fronting, the category of posterior coronal consonants expands to include all affricates in the language; in other words, *ʈ* patterns with all other affricates, which are C[cor, -ant] (*ʃ, dʒ, tɕ, dʑ*), and triggers vowel fronting in *o*-initial suffixes.

³ The attested form is *ʃekite-u*, but this is disregarded for the purposes of the present work. The vocative singular suffix allomorphy could be captured in the analysis that follows (as backing and raising of *e* after most stem-final FCs, a pattern that is extended to some non-FC-final stems, too – or simply as listed allomorphy), but the phenomenon is not sufficiently well-understood to be accounted for it in an insightful manner.

However, a significant number of neuter nouns inflect by a different pattern; the nominative singular receives no suffix, and there is an additional voiceless dental stop between the stem and the suffix in the cases that are not syncretic with the nominative singular. This is presented in Table 3: the cells in which this sort of consonant insertion occurs are shaded. The inflectional paradigm of *bure* ‘barrel’ is illustrated below, as this is a noun that allows for the default plural forms to surface (along with nouns like *tutse* ‘dozen’, *dugme* ‘button’, *putse* ‘berry’, *srtse* ‘heart’, to name a few). Most nouns of this type use suppletive, collective forms to express plurality – I consider this to be an override of the default pattern laid out in Table 3, and disregard the suppletive plural for the purposes of this work.⁴

Table 3. Inflection of SC *bure* ‘barrel’ (NEU.)

	SG.	PL.
NOM.	bure	buret-a
GEN.	buret-a	buret-aa
DAT.-LOC.	buret-u	buret-ima
ACC.	bure	buret-a
VOC.	bure	buret-a
INS.	buret-om	buret-ima

Serbo-Croatian noun stems are predominantly consonant-final in all inflectional classes, as Tables 1 and 2 show for masculine and neuter (= Class I) stems. The same holds for Class II (2) and Class III (3) stems – both classes of mostly feminine stems, although Class II allows masculine entries as well.

⁴ Collective forms are in competition with the regular plurals, and most often win in such a way that regular plurals rarely appear as output forms. The formalization of this process is outside the bounds of this chapter; the regular neuter plurals are modeled as they apply to the wider group of neuter nouns in Serbo-Croatian.

(2)	žen-a woman.NOM.SG	žen-e woman.GEN.SG	žen-i woman.DAT.SG	etc.
(3)	reč word.NOM.SG	reč-i word.GEN.SG	reč-i word.DAT.SG	etc.

In the literature, neuter stems such as the ones in Table 3 have variously been proposed to be: (a) vowel-final stems with consonant insertion, forming a listed subclass of neuter nouns on their own (Barić et al., 1997; Klajn, 2005); (b) consonant-final stems with truncation in the nominative singular (Brozović, 2006); (c) consonant-final stems with CV stem extenders (Šljivić-Šimšić, 1984). I claim that *bure*-type neuter stems are actually *vowel-final* (as they are in the nominative/accusative/vocative singular), and that these aberrant stems are repaired by consonant insertion between the stem-final vowel and the vowel of the inflectional suffix.

3.1.1 Vowel-final neuter stems as a listed (sub)class

When discussing Serbo-Croatian noun inflectional classes, most linguists and grammarians refer back to Stevanović (1969), who described four distinct classes of stems: consonant-final masculine and neuter stems (such as those in Tables 1 and 2), vowel-final neuter stems (such as those in Table 3), feminine stems that receive *-a* in the nominative singular (like the example in (2)), and feminine stems that do not receive an overt case suffix in the nominative singular (like the example in (3)). More recent works, like Barić et al. (1997) or Klajn (2005), continue this tradition, but note that, with the exception of the nominative singular, the case suffixes are the same for all neuter stems. This is why Barić and colleagues, as well as Klajn, resort to positing a separate, lexically listed

subclass, and define it as consisting of stems that receive an additional consonant in the oblique cases, before the case suffixes are attached.

While the rationale for such a division of lexemes is close to what the present chapter argues for, a broad generalization is thus missed – the stem allomorphy in question is predictable given the phonological form of the underlying stem. In other words, Barić and colleagues, and Klajn, like Stevanović before them, attribute the difference between the paradigms of *selo* (Table 1) and *bure* (Table 3) to the stems being lexically listed as belonging to a certain (sub)class. The difference between the treatments may appear to be minor. However, resorting to listedness should be avoided if the paradigm complexity can be described as resulting from predictable stem allomorphy, which is a central aim of this chapter.⁵

3.1.2 *bure*-type neuter stems as C-final stems with truncation in the nominative singular

Brozović (2006, p. 120), arguing for a greater degree of morphological regularity, proposes an analysis under which the additional *t* that appears in Table 3 is to be regarded as part of the stem. These stems would therefore be regular consonant-final stems, where the final consonant is deleted in the nominative singular (and the syncretic accusative and vocative).

Looking at the paradigm (Table 3) in isolation, Brozović's approach would make sense – final consonant deletion would be a way of avoiding codas in the nominative singular of neuters,

⁵ The problem is avoided altogether if class membership assignment in Serbo-Croatian is understood to be a predictable process, rather than lexically listed information. This is the claim in Petrovic (under review), and is discussed in Chapter 5.

whereas the stem-final consonant in the oblique cases would be resyllabified as the onset of the following syllable, case suffixes being vowel-initial.

The problem of this approach is that it requires restricting consonant deletion only to the pertinent individual lexemes. Under this analysis, *refeto* ‘sieve.NOM.SG’ (paradigm in Table 1) and *bure* ‘barrel.NOM.SG’ (paradigm in Table 3) would both have *t*-final stems: *refet* and *buret*, respectively. In the latter only, the final *-t* would have to be deleted in the nominative singular. The only way to avoid illicit nominative singular forms like **refe* (*t*-deletion) or **bureto* (neuter nominative singular suffix *-o*) would, again, be to lexically list these stems as belonging to separate inflectional classes. The approach outlined in this work avoids this issue and treats all neuter stems in a uniform way.

3.1.3 *bure*-type neuter stems as C-final stems with CV extenders

Šljivić-Šimšić (1984) divides the lexical items quite differently. In her account, all neuter stems are consonant-final, and they take *either -o or -e* in the nominative singular. Therefore, the stem of *selo* is *sel-* (Table 1), the stem of *poće* is *poć-* (Table 2), and the stem of *bure* is *bur-* (Table 3). Stems like *bur-* then receive a “stem extender” morpheme before the case suffixes are attached, while other stems do not.

It is unclear how the distribution of the “stem extenders” can be predicted. Šljivić-Šimšić argues that this morpheme is added only if a non-FC-final⁶ stem takes *-e* in the nominative singular – that is to say, nouns like *bur-e* (Table 3) – otherwise, nouns keep the same, non-extended stem

⁶ i.e. not *j*, *č*, *n*, *g*, *đ*, *te*, *de*, *f*, *z*, or *ts*-final; see section 2.1.

throughout the paradigm. This would assume that the oblique case paradigm is determined based on the inflected form of the nominative singular, not that paradigm membership can be inferred based on the properties of the stem.

Relying on a leading form like the nominative singular is not necessarily problematic in itself (see Petrovic, under review, and Chapter 5 below); under Šljivić-Šimšić’s analysis, however, we would have no way of predicting the distribution of the nominative singular allomorphs (*-o* or *-e*), as the choice would no longer depend on the final consonant of the stem. Here I show that *o*-initial suffixes surface faithfully if the stem-final consonant is not a FC, and that suffix-initial *o* fronts to *e* after stem-final FCs.

Finally, Šljivić-Šimšić assumes that nouns like *sirtȅe* ‘vinegar’, *jagne* ‘lamb’ or *jaje* ‘egg’ would be exceptions, given that they have a FC+*e* sequence in the nominative singular, and yet still extend their stems in the oblique cases (i.e. they follow the pattern in Table 3). These nouns do not present a problem for the present work, as they are analyzed as having vowel-final stems.

3.2 Morphologically conditioned coronal epenthesis

The analysis that follows assumes that *t*-insertion in Table 3 is a case of morphologically conditioned coronal epenthesis (Aronoff & Repetti, 2022), also referred to as non-canonical epenthesis (Moradi, 2017).⁷ As opposed to canonical epenthesis, these terms denote processes of

⁷ Historically, this *t* in Serbo-Croatian is derived from an underlying *t*; in Proto-Slavic, the *t* was present in an entire inflectional paradigm of neuter nouns (denoting young beings, like *tele* ‘calf’), including the nominative singular (Matasović, 2008: 206). Therefore, morphological *t*-epenthesis in modern Serbo-Croatian is a case of diachronic rule inversion (Vennemann, 1972) and generalization to a wider class of nouns (all neuter *e*-final stems, with no reference to the semantic criterion).

insertion of phonological material determined by factors outside of phonology (I specifically focus on morphological/morphosyntactic conditions here). What works like Moradi (2017) and Aronoff & Repetti (2022) show is that a number of different languages exhibit conditioned epenthesis patterns in specific morphological environments, while at the same time having separate strategies for repairing illicit phonological structures more generally. In San Marino, [ɪ] is the default epenthetic vowel, but [ʊ] is used at the end of 3rd person verb forms in both numbers (Aronoff & Repetti, 2022, p. 374). In Brazilian Portuguese, [z] repairs vowel hiatus that is a result of adding an affix (Bachrach & Wagner, 2007, p. 8) – vowel hiatus is otherwise repaired with a [j] or tolerated. The distinction between the two types of processes is also made by Staroverov (2014), who defines (phonological) epenthesis as a result of Splitting (an operation that draws a correspondence between one input segment and multiple output segments), and opposes it to morphologically restricted consonant-zero alternations. What directly follows from these assumptions is the observation that the identity of (canonically) epenthetic segments is predictable and restricted (sharing features with the input segments they correspond to), while that of morphologically conditioned ones is not necessarily so (with historical or other motivations, external to the morphological or phonological system).

In Serbo-Croatian, *t*-insertion occurs when an affix is added to an *e*-final stem, but *t* is not generally used to repair phonologically illicit structures. Vowel hiatus, which can be a result of adding suffixes to vowel-final stems (Table 3), is normally tolerated (4), unless one of the vowels is a front high vowel (5), in which case an epenthetic *j* is found (Marković, 2013, pp. 75-76).

- | | | | |
|-----|-----------|-----------|-----------------|
| (4) | /zaova/ | [zaova] | ‘sister-in-law’ |
| | /beograd/ | [beograd] | ‘Belgrade’ |
| | /pirueta/ | [pirueta] | ‘pirouette’ |

- (5)
- | | | |
|-----------|---------------------|----------|
| /violina/ | [vi j olina] | ‘violin’ |
| /sirius/ | [sir i jus] | ‘Sirius’ |
| /naivan/ | [na j ivan] | ‘naïve’ |

Given that all case suffixes are vowel-initial, as could be seen in Tables 1–3, one could expect *t*-insertion to be a case of syllable structure-driven epenthesis in this language. Note, however, that the language has derivational suffixes that are consonant-initial; the epenthetic *t* is inserted before any suffix (inflectional or derivational), and the C-initial suffixes in (6) show that Serbo-Croatian *t*-insertion is not a phonotactically-triggered process.

- (6)
- | | | | | |
|--------------------------------|-----------------------------------|-----|----------------------------|---------------------------------|
| /sirtee + ni/
vinegar + ADJ | [sirteet n i]
‘acetic’ | vs. | /plod + ni/
fruit + ADJ | [plod n i]
‘fertile’ |
| /tele + ji/
calf + ADJ | [telet e i]
‘calf-like’ | vs. | /koz + ji/
goat + ADJ | [koz j i]
‘goat-like’ |

No other phonologically illicit sequence is observed in the context of *t*-insertion. The process is actually not phonologically optimizing at all: *t* is epenthesized only in a certain morphological context – to repair vowel-final stems. The analysis thus assumes that Serbo-Croatian noun stems must be consonant-final at the point of suffix attachment, as most stems in the language are underlyingly consonant-final. Crucially, however, no consonant insertion occurs if there is no overt suffix attached to the stem, i.e. if the right edge of the stem is the right edge of the word itself.

Serbo-Croatian is not unique in this conditioned behavior: multiple languages are known to exhibit similar processes. A classic, often cited example is *t*-insertion in Ajiñinka Apurucayali

(also pejoratively known as Axininka Campa), which is morphologically restricted: it takes place only in suffixation processes. Consider the examples in (7). *t*-insertion is also found in Odawa, where it takes place at a personal prefix and stem boundary (8).

(7) Ajiyinka Apurucayali

/i-N-koma-i/	[iŋkomati]	‘he will paddle’
/i-N-koma-aa-i/	[iŋkomataati]	‘he will paddle again’

(Payne, 1981, p. 108; Lombardi, 2002, p. 239)

(8) Odawa

/ki-akat-i/	[kitakatʃi]	‘you are shy’
/ni-ompass/	[nito:mpass]	‘you.PL oversleep’

(Piggott, 1980; Žygis, 2010)

It is unclear from these data whether *t*-epenthesis is syllable structure-driven; what is clear, however, is that these are cases of morphologically conditioned epenthesis, as this process only applies in a specific morphological context.

Broselow (1984) argues that Amharic epenthesizes *t* to fill the last C-position in a template in roots like the one in (9a).⁸ This explains the difference from the roots like (9b), which are triconsonantal roots with underlying identical second and third consonants – behaving like the classic triconsonantal roots in (9c).

⁸ Broselow (1984) observes similar phenomena in Hebrew, Temiar, Cree, French, and Maori. In all of these languages, *t* is inserted as a default, unmarked consonant, to repair specific morphological structures.

(9) Amharic

a. /fj/ ‘consume’	b. /wdd/ ‘like’	c. /lbs/ ‘dress’	
fäjjä	wäddädä	läbbäsä	PERF
fäjto	wäddo	läbso	GER
mäfjät	mäwdäd	mälbäs	INF

(Broselow, 1984; Lombardi, 2002)

However, *t* is not generally used for epenthetic purposes in Amharic: depending on the situation, vowels are deleted or glides are inserted in vowel hiatus, and there is optional word-initial [ʔ]-epenthesis (Leslau, 1997). This phenomenon in Amharic is therefore even more similar to the Serbo-Croatian case: *t* is epenthesized to repair a certain morphological context, while the language employs other epenthetic strategies elsewhere.

I conclude that *t* is epenthesized in Serbo-Croatian neuter noun inflection solely for the purpose of repairing aberrant stems. In the following section, I provide a computational analysis of this morphologically conditioned consonant epenthesis process.

3.3 Analysis

The formalism used here is known as *Boolean Monadic Recursive Schemes* (BMRS); it was introduced by Bhaskar et al. (2020) and Chandlee and Jardine (2021) as a framework that captures both linguistic and computational generalizations, mainly about phonology. To my knowledge, BMRS has not yet been applied to morphological processes, but its structure and complexity bound

make this formalism appropriate for analyses of phenomena like the one I focus on in this dissertation.

3.3.1 Boolean Monadic Recursive Schemes (BMRS)

Using BMRSs allows for an incorporation of the observation that phonology is formally at most REGULAR on the Chomsky hierarchy (Chomsky, 1959; Johnson, 1972; Kaplan & Kay, 1994; Heinz, 2018) – i.e., that phonological processes are computations with a fixed memory. The formal nature of phonology is thus crucially different than that of syntax, which has been observed to be more expressive – at least CONTEXT-FREE on the Chomsky hierarchy. From this point of view, a good linguistic theory should be appropriately restrictive, so that it characterizes the kinds of processes that are attested in natural languages, while ruling out the kinds of processes deemed grammatically impossible.

Just like phonology, morphology has also been shown to be largely regular (Karttunen et al., 1992), most often subregular (Chandlee, 2017).⁹ Since the expressivity of BMRSs is appropriately limited to at most regular functions – more specifically, to a strict sub-class called SUBSEQUENTIAL functions (Bhaskar et al., 2020) – we can therefore expect BMRSs to reliably characterize morphological processes.

A generative theory of phonology or morphology has another important goal: to capture linguistically significant generalizations. Extensionally capturing regular patterns can easily be

⁹ The only exception is normally considered to be total reduplication – finite memory is insufficient to model a productive process which assumes copying material of unbounded size. For an account of total reduplication of morphological (unrealized) material, see Chapter 6.3; for a modeling of reduplication with 2-way finite-state transducers, see Dolatian & Heinz (2020).

done with FINITE-STATE TRANSDUCERS (FSTs); these, however, do not intensionally represent linguistic generalizations in the way linguistic theory usually does (Karttunen, 2003). In other words, with finite-state analyses, we can correctly compute the desired output, but not necessarily without losing some of the advantages of theoretical frameworks (such as Pāṇini’s Principle, discussed in the following section). Logical descriptions, on the other hand, can capture the same generalizations on computational complexity with representations that are common in linguistics (Chandlee & Jardine 2021).

3.3.2 Structure of a BMRS system

With BMRSs, morphological processes can be formalized as logical transductions on strings, which are understood to be made up of segments, as well as additional lexical, morphological, and syntactic information. Logical transductions describe functional transformations from input to output strings, where the output is defined by taking a fixed number of copies of the input structure (Filiot, 2015). BMRSs are based on the concept of recursive program schemes (in the sense of Moschovakis, 2018): schemes that recursively define functions that take string indices as input, and return a boolean value. The output value of a segment can then be defined on the basis of predicates that refer to the input *and* output structures local to that segment (Chandlee & Jardine 2021).

Discussing finite-state implementations of realizational morphology, which are more common in computational morphology, Karttunen (2003) notes that, despite the availability of such accounts, they are not evidently endowed with the advantages of a theoretical, inferential-

realizational morphological framework (such as Stump, 2001). For instance, a crucial notion like Pāṇini's Principle (= Elsewhere condition) can be implemented in a finite-state account, but it must be stipulated as an additional mechanism in order to capture *do X unless Y* generalizations. With the 'if... then... else...' structures of BMRSs, on the other hand, elsewhere condition effects directly fall out from how the grammar is evaluated. Pāṇini's Principle has been argued to be crucial in multiple morphological and phonological theories (Network Morphology, Paradigm Function Morphology, Distributed Morphology, Optimality Theory; see Kiparsky, 1973; Aronoff, 1976; Zwicky, 1986; Anderson, 1992; etc.).

Relying on the 'if... then... else...' syntax, BMRSs therefore effectively implement a ranking of predicates that identify particular structures in either the input or output. These structures may act like *licensing* or *blocking* structures, depending on how they affect the output, and how they fit into the overall BMRS template. Furthermore, this ranking is *hierarchical*: this has been described as similar to constraint ranking in OT (Chandlee & Jardine 2021), but can be used to model competition, inheritance, and defaults (normal-case, exceptional-case) in morphology. The analysis below thus makes crucial reference to more specific and less specific conditions on inflectional exponence.

3.3.3 Serbo-Croatian *t*-epenthesis: A BMRS analysis

As argued in section 2.1, *t*-epenthesis in Serbo-Croatian is predictable, and as such should not be analyzed by means of listed stem allomorphy.

The primitives of BMRSs, as the name suggests, are the *boolean values* \top and \perp (True and False), the predecessor and successor function p and s , and a finite set of *monadic predicates* $P(t)$ – predicates that take a single argument t , and return \top or \perp .

The alphabet Σ is a finite set of symbols; it is the union of sets \mathbb{A} , which contains the consonant and vowel segments of Serbo-Croatian, and \mathbb{M} , which is the set of morphosyntactic features like case and number. Σ_{∞} is the union of Σ and all necessary boundary symbols – which, for the purposes of this account, is just the stem boundary symbol $+$.

$$(10) \quad \begin{aligned} \mathbb{A} &= \{a, b, \text{ts}, \text{tʃ}, \text{tʂ}, d, \text{ɟ}, \text{ɟʂ}, e, f, g, x, j, k, l, \text{ʌ}, m, n, \text{ɲ}, o, p, r, s, \text{ʃ}, t, u, v, z, \text{ʒ}\} \\ \mathbb{M} &= \{[\text{nom}], [\text{gen}], [\text{dat}], [\text{acc}], [\text{voc}], [\text{ins}], [\text{loc}], [\text{sg}], [\text{pl}], \\ &\quad [\text{mas}], [\text{neu}]\} \\ \Sigma &= \mathbb{A} \cup \mathbb{M} \\ \Sigma_{\infty} &= \Sigma \cup \{+\} \end{aligned}$$

In the present analysis, set \mathbb{A} is made up of segments, which is an adequate representation for the points being made in the present work. One could, of course, assume a representation of segments as feature bundles, which would be more appropriate for a phonological analysis; a formalization of Serbo-Croatian $[j]$ -epenthesis, illustrated in (5), would benefit from that, as that is a case of homorganic glide epenthesis. Here I abstract away from further details, but see Chandlee & Jardine (2021) for how phonological features can be used instead of segments in a BMRS system.

For all symbols in Σ_{∞} , there is a set \mathcal{I} of input predicates (marked with a subscript i), and a set \mathcal{O} of output predicates (marked with a subscript o , and a superscript output copy number) for each copy of the output string. Multiple output copies are needed to capture processes that involve

outputting multiple characters in a position occupied by a single character in the input string. As the longest exponent in Serbo-Croatian Noun Class I is three segments long, we will need three copies of the output string for the present analysis; this is further explained and illustrated later in the section, when the copy set C is introduced in (21).

$$\begin{aligned}
 (11) \quad I &= \{a_i(x), \dots, z_i(x), [\text{nom}]_i(x), \dots, [\text{pl}]_i(x), +_i(x)\} \\
 \mathcal{O}^1 &= \{a_o^1(x), \dots, z_o^1(x), [\text{nom}]_o^1(x), \dots, [\text{pl}]_o^1(x), +_o^1(x)\} \\
 \mathcal{O}^2 &= \{a_o^2(x), \dots, z_o^2(x), [\text{nom}]_o^2(x), \dots, [\text{pl}]_o^2(x), +_o^2(x)\} \\
 \mathcal{O}^3 &= \{a_o^3(x), \dots, z_o^3(x), [\text{nom}]_o^3(x), \dots, [\text{pl}]_o^3(x), +_o^3(x)\}
 \end{aligned}$$

The argument x ranges over domain elements. Strings in $\Sigma_{\mathfrak{M}}$ are identified with structures of the form in (12), where the domain D is the set of indices, and each character $\sigma \in \Sigma_{\mathfrak{M}}$ has σ_n as the unary relation $\sigma_n \subseteq D$ selecting the indices that that segment occupies. $p(x)$ is a term referring to the predecessor of x , and $s(x)$ is a term referring to the successor of x .

$$(12) \quad \mathbf{S} = \langle D; \sigma_1, \sigma_2, \dots, \sigma_n, p, s \rangle$$

To illustrate how input predicates are evaluated, I use the vowel-final stem of *bure* ‘barrel’ in Table 4. Note that, in a given analysis, all BMRS predicates are evaluated in parallel (rather than sequentially), much akin to how rewrite rules are applied in *two-level morphology* (Koskenniemi, 1983). Some example input predicates are given in Table 4, with various terms as arguments. These are listed with their truth values for each segment in the word.

Table 4. Values of some input predicates for the segments of SC *bure* ‘barrel’

	1	2	3	4	5
	[neu] b	[neu] u	[neu] r	[neu] e	+
$[neu]_i(x)$	\top	\top	\top	\top	\perp
$e_i(x)$	\perp	\perp	\perp	\top	\perp
$+_i(x)$	\perp	\perp	\perp	\perp	\top
$r_i(p(x))$	\perp	\perp	\perp	\top	\perp
$r_i(s(s(x)))$	\top	\perp	\perp	\perp	\perp

I assume that all Serbo-Croatian stems come lexically marked with a gender feature; as Table 4 shows, every segment of the stem *bure* is associated with the neuter gender feature $[neu]$. The predicate $[neu]_i(x)$ therefore evaluates to true for positions 1-4, which are occupied by the stem segments, but not position 5, which is occupied by the stem boundary symbol +.

Moving down Table 4, we can see that $e_i(x)$ is true for the input position occupied by the stem-final vowel of *bure* (position 4), and no other position. $+_i(x)$ evaluates to true only in position 5 – the position marking the morpheme boundary directly following the stem. The last two predicates have arguments that refer to successors and predecessors. $r_i(p(x))$ is true for *e* in position 4, as $r_i(p(x))$ can be understood as “*r* is the predecessor of *x*”. Similarly, $r_i(s(s(x)))$ is true for *b* in position 1, as the interpretation is “*r* is the successor of the successor of *x*”.

Consider now the realization of the dative (and locative) singular form of *bure* ‘barrel’ in Table 5. The input string consists of the stem *bure*+, directly followed by the dative case feature $[dat]$, and the singular number feature $[sg]$.¹⁰

¹⁰ The input structures are assumed to come from the syntax; this is further explained in Chapters 5 and 6.

Table 5. Input and output strings of the dative singular form of SC *bure* ‘barrel’

<u>input</u>	1	2	3	4	5	6	7
	[neu]	[neu]	[neu]	[neu]			
	b	u	r	e	+	[dat]	[sg]
<u>output</u>							
Copy 1:	b	u	r	e	t	u	

In order to get the desired output string *buretu* as the dative singular form of *bure* ‘barrel’, while ensuring that the locative singular output form is the same, we need to define the dative-locative syncretism, as well as the dative-locative singular feature bundle. Given that the dative and locative are always syncretic, I define the predicate $[\text{dat}/\text{loc}]_i(x)$ in (13) to have it return \top for any instance of the dative $[\text{dat}]$ or locative $[\text{loc}]$ feature in the input.

The conditions under which an input case feature will be marked as the dative-locative syncretism are laid out using an ‘if... then... else...’ statement. Let us examine the definition of $[\text{dat}/\text{loc}]_i(x)$ in (13): if, for a given input position x , the predicate $[\text{dat}]_i(x)$ evaluates to \top , $[\text{dat}/\text{loc}]_i(x)$ will also return \top ; if the predicate $[\text{dat}]_i(x)$ returns \perp , the predicate $[\text{loc}]_i(x)$ is evaluated, and the boolean value it returns will be the evaluation of $[\text{dat}/\text{loc}]_i(x)$.

$$(13) \quad [\text{dat}/\text{loc}]_i(x) = \text{if } [\text{dat}]_i(x) \text{ then } \top \text{ else } [\text{loc}]_i(x)$$

The predicate $[\text{nom}/\text{acc}/\text{voc}]_i(x)$ is similarly defined in (14), with two ‘if... then... else...’ statements: the predicate returns \top for any instance of $[\text{nom}]$, $[\text{acc}]$ or $[\text{voc}]$ in the input structure, as the nominative, accusative and vocative are syncretic for neuter nouns.

$$(14) \quad [\text{nom}/\text{acc}/\text{voc}]_i(x) = \text{if } [\text{nom}]_i(x) \text{ then } \top \text{ else} \\ \text{if } [\text{acc}]_i(x) \text{ then } \top \text{ else } [\text{voc}]_i(x)$$

The necessary feature bundles are defined in (15). I start with $[\text{dat/loc sg}]_i(x)$, which is needed for the realization of the u suffix invoked in Table 5. The predicate $[\text{dat/loc sg}]_i(x)$ evaluates to \top *iff* the user-defined input predicate $[\text{dat/loc}]_i(x)$ returns \top in the same position, and the position that directly succeeds is associated with the input number feature $[\text{sg}]$ (the “else \perp ” clause is implied, and is left out for clarity). Similarly, other necessary combinations of case and number features available for neuter nouns are defined below.¹¹

- (15) $[\text{dat/loc sg}]_i(x) = \text{if } [\text{dat/loc}]_i(x) \text{ then } [\text{sg}]_i(s(x))$
 $[\text{nom/acc/voc sg}]_i(x) = \text{if } [\text{nom/acc/voc}]_i(x) \text{ then } [\text{sg}]_i(s(x))$
 $[\text{ins sg}]_i(x) = \text{if } [\text{ins}]_i(x) \text{ then } [\text{sg}]_i(s(x))$
 $[\text{nom/acc/voc pl}]_i(x) = \text{if } [\text{nom/acc/voc}]_i(x) \text{ then } [\text{pl}]_i(s(x))$
 $[\text{gen pl}]_i(x) = \text{if } [\text{gen}]_i(x) \text{ then } [\text{pl}]_i(s(x))$
 $[\text{dat/loc pl}]_i(x) = \text{if } [\text{dat/loc}]_i(x) \text{ then } [\text{pl}]_i(s(x))$
 $[\text{ins pl}]_i(x) = \text{if } [\text{ins}]_i(x) \text{ then } [\text{pl}]_i(s(x))$

Feature bundles can also be defined so as to capture syncretic patterns at the same time: consider the definition of $[\text{dat/loc/ins pl}]_i(x)$ in (16), which is responsible for ensuring that the dative, locative and instrumental plural suffixes of neuter nouns in Serbo-Croatian will be realized by one exponent. Similarly, $[\text{gen/ nom/acc/voc pl}]_i(x)$ will evaluate to \top if a

¹¹ Standard logical conjunction could also be used (i.e., $[\text{dat/loc sg}]_i(x) \wedge [\text{sg}]_i(s(x))$), as it is equally expressive (Chandlee and Jardine 2021, Moschovakis 2018); here I choose to follow the ‘if... then... else...’ syntax consistently.

position is associated with the genitive input feature $[\text{gen}]$ (regardless of number), or the nominative-accusative-vocative plural feature bundle $[\text{nom/acc/voc pl}]$.

$$(16) \quad [\text{dat/loc/ins pl}]_i(x) = \text{if } [\text{dat/loc pl}]_i(x) \text{ then } T \text{ else } [\text{ins pl}]_i(x) \\ [\text{gen/ nom/acc/voc pl}]_i(x) = \text{if } [\text{gen}]_i(x) \text{ then } T \text{ else } [\text{nom/acc/voc pl}]_i(x)$$

Let us now turn to output forms: the predicates in (17–18) define the conditions under which elements in the output alphabet appear in the first copy of the output structure. The predicate in (17) models inflectional exponence: as illustrated in Table 5, the phonological realization of the dative (and locative) singular is u . If the input position under consideration does not return true for the dative-locative singular predicate $[\text{dat/loc sg}]_i(x)$, u can only be output faithfully as a counterpart of an input u .

$$(17) \quad u_o^1(x) = \text{if } [\text{dat/loc sg}]_i(x) \text{ then } T \text{ else } u_i(x)$$

The definition of $a_o^1(x)$ in (18) states that a is the exponent of the genitive case (in both singular and plural), as well as the nominative, accusative and vocative plural. If this condition for a is not met, the only other way to get an a in the output is for it to be a realization of an input a . Similarly, i can be the (first segment of the) instrumental plural exponent, or a faithful output of an input i – otherwise it does not appear in the (first copy of the) output string.

$$(18) \quad a_o^1(x) = \text{if } [\text{gen/ nom/acc/voc pl}]_i(x) \text{ then } T \text{ else } a_i(x) \\ i_o^1(x) = \text{if } [\text{ins pl}]_i(x) \text{ then } T \text{ else } i_i(x)$$

Turning back to the example in Table 5, in which the input string *bure*+ [dat] [sg] gets output as *buretu*, we can see that, under certain conditions (e-final stem followed by an overt suffix), the stem boundary symbol + can be output as *ɫ*. In order to adequately capture that, in (19) I first define the predicate *outseg(x)*, which returns *T* for all instances of segments realized in the first copy of the output string:

$$(19) \quad \text{outseg}(x) = \begin{array}{l} \text{if } a_o^1(x) \text{ then } T \text{ else} \\ \text{if } b_o^1(x) \text{ then } T \text{ else} \\ (...) \\ \text{if } z_o^1(x) \text{ then } T \text{ else } z_o^1(x) \end{array}$$

This user-defined predicate is then referred to in the definition of $t_o^1(x)$ in (20), which specifies that the stem boundary symbol + will be output as *ɫ* *iff* *e* (the only stem-final vowel in the native lexicon) directly precedes, *and* there is an output segment directly following the stem boundary (as shown in Table 5). Otherwise, *ɫ* will only be output as a faithful counterpart of an input *ɫ*.

$$(20) \quad t_o^1(x) = \begin{array}{l} \text{if } +_i(x) \text{ then} \\ \quad \text{if } e_i(p(x)) \text{ then outseg}(s(x)) \\ \quad \text{else } \tau_i(x) \end{array}$$

This predicate effectively models morphological *t*-epenthesis. A few details remain to be ironed out, most notably the treatment of exponents longer than a single segment. Consider the example in Table 6, which shows the input and output strings of the instrumental plural form of *bure* ‘barrel’.

Table 6. Input and output strings of the instrumental plural form of SC *bure* ‘barrel’

<u>input</u>	1	2	3	4	5	6	7
	[neu]	[neu]	[neu]	[neu]			
	b	u	r	e	+	[ins]	[pl]
<u>output</u>							
Copy 1:	b	u	r	e	t	i	
Copy 2:						m	
Copy 3:						a	

As shown in Tables 1-3, the Serbo-Croatian noun inflection system consists of case-number exponents; these feature bundles (and syncretisms) are defined in (13-16), and they occupy the same position as the case feature (position 6 in Table 6). For any exponent longer than one segment, all output segments must still occupy the same position; in order to make that possible, we can define a larger COPY SET of output structures. As the longest case-number exponent in our system is three segments long (*-ima*, dative-locative-instrumental plural), a copy set of size 3 will suffice (21):

$$(21) \quad C = \{1, 2, 3\}$$

For the purpose of retaining the desired computational complexity of the system, the order of the output copies is fixed: it is derived from the order on C (21), and the order on the indices in S (12). This means that segments are output in a top-to-bottom, left-to-right fashion: given an input string of length n , for any single index i_m , output copy 1 is ordered first, directly followed by copies 2 and 3, respectively, for the same index. These are then followed by the respective output copies of the immediately following indices i_{m+1} , i_{m+2} , ..., i_n , for all output characters whose functions return \top at each string index. This ensures order preservation; for a more formal

definition, refer to Bhaskar et al. (2020). For more about the relation between order-preserving logical transductions and one-tape finite-state transducers, see Filiot (2015).

Output copy 2 and copy 3 predicates can now be defined. As shown in Tables 1-3, *m* is the second segment of the instrumental singular exponent, or it can surface as the second output segment of the dative-locative-instrumental plural syncretism (exemplified in Table 6); otherwise, *m* does not appear in Copy 2 of the output. This is captured in the definition of $m_o^2(x)$ in (22).

$$(22) \quad m_o^2(x) = \text{if } [\text{ins sg}]_i(x) \text{ then } \top \text{ else } [\text{dat/loc/ins pl}]_i(x)$$

As per the definition of $a_o^2(x)$ in (23), *a* surfaces in output Copy 2 as the second segment of the genitive plural exponent, essentially realizing the *-aa* suffix (see Tables 1 and 2 in Section 2). If this condition is not met, *a* does not surface as a Copy 2 output segment.

$$(23) \quad a_o^2(x) = [\text{gen pl}]_i(x)$$

Finally, only one output Copy 3 predicate is needed – for the realization of the final segment of the dative-locative-instrumental plural suffix *-ima* (Table 6). This is defined in (24):

$$(24) \quad a_o^3(x) = [\text{dat/loc/ins pl}]_i(x)$$

At this point, the only exponents laid out in Section 2 that have not yet been defined are *o*- and *e*-initial suffixes. Remember that the examples in Table 2 illustrate the effect of stem-final fronting consonants (FC) on the choice of surface form of certain inflectional suffixes: specifically, neuter FC-final stems additionally take *-e* as the nominative-accusative-vocative singular suffix

(which is otherwise *-o* for neuter nouns), while all FC-final Class I stems take *-em* in the instrumental singular (*-om* elsewhere). Consider the input and output structures of the instrumental singular forms of *selo* ‘village’ in Table 7, and those of *ποῖε* ‘field’ in Table 8.

Table 7. Input and output strings of the instrumental singular form of SC *selo* ‘village’

<u>input</u>	1	2	3	4	5	6
	[neu]	[neu]	[neu]			
	s	e	l	+	[ins]	[sg]
<u>output</u>						
Copy 1:	s	e	l		o	
Copy 2:					m	

Table 8. Input and output strings of the instrumental plural form of SC *ποῖε* ‘field’

<u>input</u>	1	2	3	4	5	6
	[neu]	[neu]	[neu]			
	p	o	λ	+	[ins]	[sg]
<u>output</u>						
Copy 1:	p	o	λ		e	
Copy 2:					m	

In order to capture the pattern illustrated in Tables 7 and 8, I utilize the user-defined input predicate $FC_i(x)$ (25), as well as output Copy 1 predicates $e_o^1(x)$ (26) and $o_o^1(x)$ (27). Let us first look at $FC_i(x)$: it returns T for all members of the set of fronting consonants, which was defined in Section 2.¹²

¹² The terms in (25) can be reordered in any way; this is equally expressive as standard logical disjunction.

(25) $FC_i(x) =$ if $\mathfrak{t}_i(x)$ then \top else
 if $\mathfrak{f}_i(x)$ then \top else
 if $\mathfrak{b}_i(x)$ then \top else
 if $\mathfrak{g}_i(x)$ then \top else
 if $\mathfrak{d}_i(x)$ then \top else
 if $\mathfrak{j}_i(x)$ then \top else
 if $\mathfrak{A}_i(x)$ then \top else
 if $\mathfrak{n}_i(x)$ then \top else
 if $\mathfrak{j}_i(x)$ then \top else $\mathfrak{z}_i(x)$

Then, in (26), the definition of $e_o^1(x)$ states that e is output in Copy 1 as the instrumental singular exponent after an FC-final stem (line 1). For cases other than the instrumental singular, the gender feature of the stem needs to be checked; a stem bearing the neuter gender feature $[\text{neu}]$ (line 2) can have e following in output Copy 1 as a realization of the nominative-accusative-vocative singular syncretism, provided that the stem is FC-final (line 3). If $[\text{neu}]_i(p(p(x)))$ in line 3 returns \perp , however, the ‘else’ part of that statement is evaluated (line 4). Given that Inflectional Class I in Serbo-Croatian consists only of masculine and neuter nouns, the statements in lines 5 and 6 describe e ’s as suffixes attaching to masculine stems – specifically as the vocative singular exponent, or the accusative plural exponent (regardless of the identity of the stem-final consonant). Finally, if none of the conditions above hold, e can only surface as a faithful output of an input e (line 6).

(26) $e_o^1(x) =$ if $[\text{ins sg}]_i(x)$ then $FC_i(p(p(x)))$ else
 if $[\text{neu}]_i(p(p(x)))$ then
 if $[\text{nom/acc/voc sg}]_i(x)$ then $FC_i(p(p(x)))$
 else
 if $[\text{voc sg}]_i(x)$ then \top else
 if $[\text{acc pl}]_i(x)$ then \top else $e_i(x)$

In (27), the definition of $\circ_o^1(x)$ states that \circ is the exponent of the instrumental singular (line 1), provided that the conditions that would produce an e in the same position in output Copy 1 do not hold (line 2). This makes \circ the default exponent of the instrumental singular in Class I – it gets output in the absence of more specific conditions. \circ is also the default realization of the nominative-accusative-vocative singular syncretism for neuter nouns (lines 4-5), subject to two blocking structures: the stem cannot be e -final (line 6), and, once again, the conditions that would give rise to an e in the same position in output Copy 1 must not hold (line 7). Note how the structure of the definition of $\circ_o^1(x)$ (the hierarchical order of predicates) is parallel to that of $e_o^1(x)$: this captures the generalization that, in this inflectional class, suffix-initial default \circ is overwritten by e in a specific context (after FC-final stems).¹³

(27) $\circ_o^1(x) =$ if $[ins \ sg]_i(x)$ then
 if $e_o^1(x)$ then \perp else \top
 else
 if $[neu]_i(p(p(x)))$ then
 if $[nom/acc/voc \ sg]_i(x)$ then
 if $e_i(p(p(x)))$ then \perp else
 if $e_o^1(x)$ then \perp else \top
 else $\circ_i(x)$

Finally, in (28) I list the identity functions over the remainder of the members of the set \mathbb{A} of segments; in output Copy 1, these are all output faithfully, without further stipulations. No equivalent predicates are defined for output Copies 2 and 3.

¹³ In an alternative approach, the initial vowel of the $[ins \ sg]$ and the $[nom/acc/voc \ sg]$ suffixes is always \circ , which then gets fronted to e after FCs. I develop that account in Chapter 5.

$$\begin{aligned}
(28) \quad r_o^1(x) &= r_i(x) \\
m_o^1(x) &= m_i(x) \\
b_o^1(x) &= b_i(x) \\
t_o^1(x) &= t_i(x) \\
ʃ_o^1(x) &= ʃ_i(x) \\
tʂ_o^1(x) &= tʂ_i(x) \\
&(\text{etc., for } \{d, ɖ, ɗ, f, g, x, j, k, l, \Lambda, n, ɲ, p, s, ʃ, v, z, ʒ\})
\end{aligned}$$

No other output predicate needs to be defined; characters that are not defined by output predicates will not appear in any of the output copies. This includes non-segmental characters (morphosyntactic features, feature bundles, boundary symbols), as well as null case endings (e.g. masculine nominative singular).

The analysis above accounts for τ -epenthesis as rewriting of an input stem boundary symbol with an output consonant. This is not in itself a practice that is unheard of – see Bhaskar et al. (2020) for an example of word-final consonant insertion (where the right word boundary input symbol \propto is similarly output as the desired consonant), and Koskeniemi (1983) for how, more generally, strings are padded out with special symbols in two-level morphology (given that it fundamentally relies on equal length relations). The approach in this chapter specifically captures the understanding that the Serbo-Croatian t -insertion process occurs at a morphologically salient position, while being sensitive to the phonological form of the stem, and the presence of a suffix that follows.

3.4 Conclusion

This chapter provided an analysis of morphologically conditioned consonant insertion in Serbo-Croatian; as shown in section 2.1, native stems in this language are required to be consonant-final on the surface. Those that are not end up with an epenthesized *t* at the end, whenever an overt suffix is attached to a stem – and that happens regardless of whether that suffix is vowel- or consonant-initial. In other words, while sensitive to the phonological form of the stem, the insertion process is not triggered by a need for resyllabification, or to break up vowel hiatus.

The BMRS analysis illustrated how Boolean Monadic Recursive Schemes can directly capture morphological and phonological generalizations, retaining the computationally restrictive nature of such processes. This formalism can therefore easily account for morphologically conditioned epenthesis, which is sensitive both to phonological form and morphological properties. Based on the generalization that morphological and phonological phenomena, unlike syntactic ones, are restricted in much the same way in terms of computational complexity and locality, the morphological module can be characterized more generally, formalizing the insights of both morpheme-based and paradigm-based approaches.

4 Morphological epenthesis: Romance

In this chapter I turn to allomorphy found in several Romance languages. I demonstrate that these processes are predictable, and lie on the morphology-phonology interface. The allomorphs under consideration reflect morphologically conditioned consonant insertion, and their distribution is completely predictable from the phonological shape of the stem and morphosyntactic properties of the lexeme.

It would be reasonable to assume that insertion processes such as the one outlined in the previous chapter should not be limited to coronal stops, or even consonants; this is indeed what we find in a number of different languages. Epenthesis is usually thought of as a purely phonological process that improves marked structures; however, the phenomena presented here cannot be accounted for in exclusively phonological terms, and are morphosyntactically conditioned. In this chapter I examine more instances of morphological epenthesis, and conclude that a morphological module that operates over strings correctly predicts the existence of such processes and properly handles them.

4.1 [u]-epenthesis in Catalan

In Catalan, [e] is used as the default epenthetic vowel to syllabify an illicit sequence of consonants (29) **Error! Reference source not found..** However, within the nominal domain, [u] is epenthesized to repair syllable structure if the masculine gender is involved. Specifically, masculine adjective stems ending in /s/ are realized with an [o] before the plural suffix /s/ (30a)

Error! Reference source not found.. In the same context, with a feminine stem, [o]-insertion is not triggered (30b). The same generalization applies to noun stems (30c). Finally, if there is no disallowed */ss/ sequence, no epenthesis occurs (30d).

(29) [ɛspilber] ‘Spielberg (name)’

- (30) a. /felis + s/ → [felisus] ‘happy.MAS.PL’
 b. /felis + s/ → [felises] ‘happy.FEM.PL’
 c. /gos + s/ → [gosus] ‘dog.MAS.PL’
 d. /gat + s/ → [gats] ‘cat.FEM.PL’

The Catalan data are analyzed by Zimmermann (2019) as involving a ghost segment /u/; in this analysis, masculine nominal stems underlyingly contain a latent segment that only surfaces to avoid a marked structure. The generalization that masculine stems trigger [u]-epenthesis is thus lost – a ghost segment will surface to repair a marked structure, and masculine nominal lexical entries just happen to have them. Catalan [u]-epenthesis is analyzed by others as morphologically conditioned epenthesis (Loporcaro, 1997; Artés Cuenca, 2016; Moradi, 2017; Aronoff & Repetti, 2022); this is the approach I pursue here.¹⁴

Consider the input and output strings of [gosus] ‘dogs’ in Table 9:

¹⁴ Loporcaro (1997) suggests that these alternations in Catalan constitute an example of a suffix (masculine gender marker) that was reanalyzed as epenthetic. Similarly to Serbo-Croatian [t], this would then be a case of diachronic rule inversion (Vennemann, 1972).

Table 9. Input and output strings of Catalan [gosus] ‘dogs’

<u>input</u>	<hr/>				
	1	2	3	4	5
	[m]	[m]	[m]		[pl]
	g	o	s	+	s
<u>output</u>	<hr/>				
	<hr/>				
Copy 1:	g	o	s	u	s

The insertion of [u] in Catalan /gos + s/ \rightarrow [gosus] ‘dogs’ can be formalized by specifying the conditions under which the morpheme boundary symbol + can be rewritten as u. Assuming that every segment of /gos/ is associated with the masculine gender feature [m], in the definition of $u_o(x)$ in (31) I provide the conditions under which u appears in the output structure. If the position under consideration is occupied by a morpheme boundary + (line 1), the position that directly precedes needs to contain a segment that is marked for masculine gender (line 2), and what needs to follow is a suffix bearing the plural number diacritic [pl] (line 3). The statement in line 4 is a formalization of the */ss/ constraint: if all of the conditions in lines 1-3 evaluate to T, there must also be an s in the input structure directly preceding the morpheme boundary +, and an input s directly following it. Finally, if the position under consideration is not occupied by a morpheme boundary + in the input, the only condition that can trigger an occurrence of an output u is the presence of an input u in the same position (line 5).

$$\begin{aligned}
 (31) \quad u_o^1(x) = & \text{if } +_i(x) \text{ then} \\
 & \text{if } [m]_i(p(x)) \text{ then} \\
 & \quad \text{if } [pl]_i(s(x)) \text{ then} \\
 & \quad \quad \text{if } s_i(p(x)) \text{ then } s_i(s(x)) \\
 & \text{else } u_i(x)
 \end{aligned}$$

The faithful output of any input \cup is thus ensured, and this is the most general, default condition on any such output symbol – parallel to low-ranked Ident constraints in OT, for instance.

4.2 [z]-epenthesis in Brazilian Portuguese

Aronoff and Repetti (2022) discuss a case in Brazilian Portuguese in which the choice between two epenthetic segments to repair a phonological violation is made on the basis of morphological structure: [j] is the default epenthetic segment used to repair hiatus (32). However, if the hiatus is formed between morphemes, [z] is used instead (32a-c) (Garcia, 2017; Bachrach & Wagner, 2007).

(32) [koʁɛjɐ] ‘Correa/Corrêa/Correia (name)’

- (33) a. /sofa + iɲu/ → [sofaʒiɲu] ‘sofa.DIM’
 b. /kafe + aw/ → [kafezaw] ‘coffee grove’
 c. /kafe + ejɾu/ → [kafezejɾu] ‘coffee producer’

No allomorphic solution can express the fact that Brazilian Portuguese epenthesizes [z] rather than [j] only between morphemes. A conditioned epenthesis account is fairly straightforward, however: as the morpheme boundary symbol + can only appear at morpheme edges, we can use that representation to define the locus of morphological epenthesis.

Table 10. Input and output strings of Portuguese [sofazĩnu] ‘sofa.DIM’

<u>input</u>								
	1	2	3	4	5	6	7	8
	s	o	f	a	+	i	n	u
<u>output</u>								
Copy 1:	s	o	f	a	z	i	n	u

The morpheme boundary symbol + gets rewritten as z only if a certain condition is met – there needs to be a vowel directly on each side of the morpheme boundary. The predicate $VOW_i(x)$ (34) is thus defined to return T for all the members of the Brazilian Portuguese vowel inventory. [z]-insertion is then modeled by $z_o^l(x)$ in (35).

$$(34) \quad VOW_i(x) = \begin{array}{l} \text{if } a_i(x) \text{ then T else} \\ \text{if } i_i(x) \text{ then T else} \\ (\dots) \\ \text{if } o_i(x) \text{ then T else } u_i(x) \end{array}$$

$$(35) \quad z_o^l(x) = \begin{array}{l} \text{if } +_i(x) \text{ then} \\ \quad \text{if } VOW_i(p(x)) \text{ then } VOW_i(s(x)) \\ \text{else } z_i(x) \end{array}$$

Morphological epenthesis in Brazilian Portuguese is therefore indeed phonologically optimizing – z appears in the output to satisfy a markedness constraint, encoded as “if $VOW_i(p(x))$ then $VOW_i(s(x))$ ” in (35). However, the process is crucially restricted to a morphological context (it can only happen at a morpheme juncture), and the strategy used to resolve hiatus in that specific context differs from how the structure is repaired more generally in the language (via glide insertion).

4.3 [s]-epenthesis in Spanish

In Spanish, diminutives are formed by removing the masculine gender marker *-o* or the feminine gender marker *-a* (if present) from the base form, and then adding the diminutive suffix *-ito/-ita* (36a) (Crowhurst, 1992; Norrmann-Vigil, 2012). As seen in (36b), vowel hiatus contexts created through this process are not resolved. [s] is inserted between the stem and the diminutive suffix if the base form ends in a vowel that is not the *-o/-a* gender marker (36c-d), [n] (36e) or [r] (36f).

- (36) a. /gato + ito/ → [gatito] ‘cat.DIM’
 b. /bakalao + ito/ → [bakalaito] ‘cod.DIM’
 c. /kable + ito/ → [kablesito] ‘cable.DIM’
 d. /sofá + ito/ → [sofasito] ‘sofa.DIM’
 e. /kamjon + ito/ → [kamjonsito] ‘truck.DIM’
 f. /amor + ito/ → [amorsito] ‘love.DIM’

I therefore assume the input strings to have a “base form + diminutive suffix” structure, where each segment of the suffix *ito* bears a diminutive feature [dim]. This is illustrated in Tables 11 and 12 below.

Table 11. Input and output strings of Spanish [gatito] ‘cat.DIM’

<u>input</u>							
1	2	3	4	5	6	7	8
					[dim]	[dim]	[dim]
g	a	t	o	+	i	t	o

<u>output</u>								
Copy 1:	g	a	t		i	t	o	

Table 12. Input and output strings of Spanish [amorsito] ‘love.DIM’

<u>input</u>								
	1	2	3	4	5	6	7	8
						[dim]	[dim]	[dim]
	a	m	o	r	+	i	t	o
<u>output</u>								
Copy 1:	a	m	o	r	s	i	t	o

Once again, we can think about morphologically conditioned epenthesis as rewriting of the morpheme boundary symbol $+$; in Spanish diminutives, it is output as s unless the base form ends in an (unstressed) o or a (gender marker), or any consonant other than n or r . Therefore, a user-defined predicate such as $\text{CONS-}n,r_i(x)$ – which returns \top for all Spanish consonants except for n and r -- needs to be defined (37), and referred to in the definition of $s_o^1(x)$ (38). Additionally, the o/a gender marker needs to be dropped from the output; this is reflected in the definitions of $o_o^1(x)$ and $a_o^1(x)$ in (39) and (40), respectively.

$$(37) \quad \text{CONS-}n,r_i(x) = \text{if } b_i(x) \text{ then } \top \text{ else} \\ \text{if } c_i(x) \text{ then } \top \text{ else } \dots \\ (\text{etc., for all Spanish consonants except for } n \text{ and } r)$$

$$(38) \quad s_o^1(x) = \text{if } +_i(x) \text{ then} \\ \text{if } [\text{dim}]_i(s(x)) \text{ then} \\ \text{if } o_i(p(x)) \text{ then } \perp \text{ else} \\ \text{if } a_i(p(x)) \text{ then } \perp \text{ else} \\ \text{if } \text{CONS-}n,r_i(x) \text{ then } \perp \text{ else } \top \\ \text{else } s_i(x)$$

$$(39) \quad o_o^1(x) = \text{if } o_i(x) \text{ then} \\ \text{if } +_i(s(x)) \text{ then } \perp \text{ else } \top$$

$$(40) \quad a_o^1(x) = \text{if } a_i(x) \text{ then} \\ \text{if } +_i(s(x)) \text{ then } \perp \text{ else } \top$$

4.4 Conclusion

The focus of this chapter has been the generalization that insertion of phonological material can be governed by morphological criteria alongside or independently of phonological ones, and that insertion processes such as those outlined in this section are not necessarily phonologically optimizing. A system formalized in this way is not directly dependent on naturalness or phonological substance – its purpose is rather to define the necessary conditions on the realization of processes, restricted by their computational complexity. The motivations behind such processes can be varied, but are crucially external to the system that generates them, not central to formalizing them. With BMRSs, we can model processes like canonical and non-canonical epenthesis with no additional impositions on the input structure (like abstract segments or listed allomorphs), other than assuming representations like morpheme boundaries and morphosyntactic features as part of the input string.

5 Leading forms: cues for inflectional class assignment

As can be seen from the analysis offered in the previous chapters, the ‘if... then... else...’ syntax of BMRS necessarily imposes a hierarchical structure of predicates. This ordering of licensing structures and blocking structures can be crucial for the analysis of morphological systems, characterized by phenomena like inheritance or Pāṇini’s principle. Recursive schemes are an abstract way of studying algorithms (Moschovakis, 2018), and algorithms are crucial to linguistic theory; this chapter focuses precisely on the algorithms that play a major role in the inflectional system of Serbo-Croatian. Petrovic (under review) analyzes the inflectional system of Serbo-Croatian nouns in its entirety, investigating how inflectional class membership can be characterized as a result of a predictable process, rather than lexically listed information; this chapter is based on that work.

5.1 Class membership assignment in Serbo-Croatian

Deciding what declension class a Serbo-Croatian noun belongs to is a relatively complex task; the phonological form of the stem and the morphosyntactic properties (specifically, gender information), which we suppose are listed in the lexical entry, are not enough to place a stem into the correct paradigm of case endings (inflectional class). In fact, multiple inflectional classes take consonant-final stems, and the identity of the stem-final consonant is irrelevant; moreover, nouns

of different genders can belong to the same class. Cross-linguistically, inflectional class membership is often listed information (Aronoff, 1994, p. 65), but an analysis that characterizes it as a result of a predictable process is always preferable, if available. Furthermore, in the specific case of Serbo-Croatian, such an analysis would also account for how new words (loanwords, nonce words) get assigned to a declension class.

The approach that I adopt takes into account a *leading form* (Wurzel, 1990; McCarthy, 2005), also referred to as *base* (Albright, 2008b), for every lexical entry, in addition to the lexical stem. I argue that Serbo-Croatian is sensitive to the nominative singular as the leading form, as it provides sufficient information for correct declension placement, in combination with the lexically listed gender and stem properties. This notion is here incorporated into an algorithmic interpretation of nominal inflection as logical transductions on strings, using Boolean Monadic Recursive Schemes (BMRSs; Chandlee & Jardine, 2021).

The chapter is organized as follows. In section 5.2, I provide an overview of the noun inflection paradigms in Serbo-Croatian, outlining the patterns that the analysis is meant to capture. In section 5.3 I argue for the status of leading forms as part of lexical entries, and introduce the inflectional class assignment algorithm. Section 5.4 discusses some additional implications of the algorithm, and Section 5.5 offers a BMRS analysis.

5.2 Serbo-Croatian inflectional classes

Chapter 3 was mainly devoted to the inflection of neuter nouns in Serbo-Croatian; here I will expand and update the paradigm of noun inflection in this language. Serbo-Croatian nouns can fall

into one of the three inflectional paradigms – Class I consists of masculine and neuter nouns, Class II comprises feminine and masculine nouns that end in *-a* in the nominative singular; feminine nouns that do not receive an overt suffix in the nominative singular belong to Class III. The (still slightly simplified) noun inflectional paradigms with example stems are shown in Table 13.

Table 13. Inflectional classes of Serbo-Croatian nouns

	CLASS I						CLASS II		CLASS III	
	masculine		neuter, masculine		neuter, masculine <i>e</i> -final		feminine, masculine		feminine	
	<i>zavod-</i> 'institute'		<i>sel-</i> 'village'		<i>bure-</i> 'barrel'		<i>žen-</i> 'woman'		<i>reč-</i> 'word'	
	SG.	PL.	SG.	PL.	SG.	PL.	SG.	PL.	SG.	PL.
NOM.	Ø	i	o	a	Ø	ta	a	e	Ø	i
GEN.	a	aa	a	aa	ta	taa	e	aa	i	ii
DAT.-LOC.	u	ima	u	ima	tu	tima	i	ama	i	ima
ACC.	Ø	e	o	a	Ø	ta	u	e	Ø	i
VOC.	e	i	o	a	Ø	ta	o	e	i	i
INS.	om	ima	om	ima	tom	tima	om	ama	i	ima

Neuter nouns are considered to belong in Class I together with masculine nouns because they inflect very similarly: *zavod*-type stems, *sel*-type stems and *bure*-type stems receive largely the same case endings, the only difference being the nominative suffixes in both numbers, and the cases patterning after them. Additionally, masculine nouns can follow any of the patterns grouped under the CLASS I label, including those that are canonically considered neuter – *auto* 'car' and *marko* 'Marko (proper name)' inflect like *selo* 'village', *mare* 'Mare (proper name, nickname)' inflects like *bure* 'barrel'. *e*-final stems receive an additional *t* before a suffix is added – this is not an exponent of a morphosyntactic feature (or feature bundle), but rather a case of morphologically conditioned consonant insertion (cf. Chapter 3). As will be shown in sections 4.3-4.5, all of these

patterns are directly derived with reference to the leading form and the phonological form of the stem.

Not represented in Table 13 is an alternation that affects the entirety of Class I: in this inflectional class, every suffix-initial *o* fronts to *e* if the stem-final consonant is a posterior coronal or an affricate, i.e. $\{t_s, t_c, d_s, f, d_j, j, ʒ, n, j, \lambda\}$. This set of segments has been introduced in Chapter 3 as the set of *fronting consonants* (FC). The surface form of the nominative singular suffix for the neuter stem *poʎ-* ‘field’ is therefore *e*, not *o* (as could be deduced from Table 13; but see Table 2). In addition to the identity of the stem-final consonant, there are other conditions that affect the choice of suffix-initial vowel (*o* or *e*), such as the identity of the stem-final vowel or the status of the lexeme as native or borrowed. Right now I will abstract away from these conditions, but see Section 5.5 for a more detailed analysis of the phenomenon.

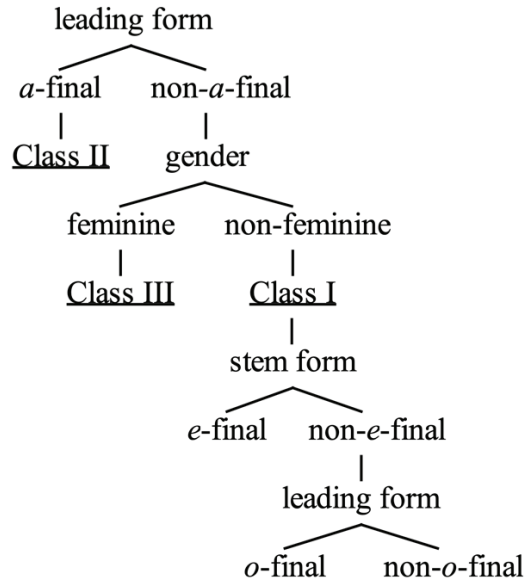
For the purposes of this work, I disregard a few additional generalizations that can be made about the inflectional paradigms of nouns in Serbo-Croatian. First, borrowed feminine nouns whose nominative singular does not end in *-a* (proper names like *karen* ‘Karen’ or *beti* ‘Betty’) are indeclinable. Second, as discussed in Chapter 3, most *bure*-type nouns do not follow the default plural pattern presented in Tables 3 and 14, and use collective forms to express plurality. Finally, for a subset of *e*-final neuter stems that end in *me*, the consonant that is added between the stem and the case ending is *n*, not *t*. The first two generalizations are robust and can easily be integrated into the analysis that follows; they are omitted for reasons of simplicity and straightforwardness of presentation. As for the third, the issue of non-*t*-stem extenders is debated (cf. Šljivić-Šimšić, 1984; Brozović, 2006, pp. 119–121), and it is not clear to which extent it can be characterized as a predictable process, which is why it is not discussed here.

5.3 Leading forms

Which paradigm a lexeme belongs to is determined by three factors: the form of the stem (*e*-final or non-*e*-final), gender information (masculine, feminine or neuter), and the leading form itself (*a*-final or non-*a*-final). If the only available information were the form of the stem and the morphosyntactic property of gender, we would not be able to know whether feminine stems (all consonant-final) would belong to Class II or Class III. On the other hand, if we only had the leading form to work with, we could not differentiate between *ποῖε* ‘field’, which contains the nominative singular suffix *-e* (given that its stem *ποῖ-* ends in a fronting consonant), and *ῥε* ‘rope’, which takes no overt suffix in the nominative singular, and has *t*-insertion in the oblique cases.

There are good reasons why the nominative singular is to be taken as the leading form: it is the most frequent form, given that it is the subject case, and it ranks highest in the Case Hierarchy (Blake, 2001). Even though there may be more informative forms (e.g. genitive singular), they would neither rank as well in the Hierarchy, nor be as frequent in use. Crucially, however, loanwords (especially from non-inflecting languages) do not have an oblique form one could refer to, but they are still regularly incorporated into the declension system on the basis of their input form – which is borrowed as nominative singular. All nouns whose nominative singular ends in *-a* (both native and borrowed, regardless of gender) belong to Class II; all non-*a*-final feminine nouns are assigned to Class III. Finally, all other nouns end up in Class I (broadly construed, including the canonically masculine and neuter patterns), which is the default declension for non-gendered loanwords. The declension class assignment algorithm is shown in Figure 3 below.

Figure 3. Decision tree for declension assignment of Serbo-Croatian stems



Animacy distinctions are omitted from Figure 3, as it is understood that they apply across the board in Class I. In this work I abstract away from accounting for the animate syncretic pattern (genitive singular-accusative singular) and default to the non-animate syncretisms for Class I (nominative singular-accusative singular, as shown in Figure 3); however, assuming $[\pm\text{animate}]$ features would allow for capturing both.

5.4 Deficient entries

An implication of the algorithm in Figure 3 is that it works with a deficient lexical entry. A complete lexical entry consists of a stem, a leading form, and morphosyntactic gender information; however, not all lexical entries – especially ones corresponding to loanwords – will be complete.

We can assume that most loanwords are incorporated into the system only with a leading form, i.e. they are neither semantically prespecified for gender, nor do they have a predetermined, listed stem form. Entries like these fall under the “non-feminine” gender and “non-*e*-final” stem form labels in Figure 3. Thus, consonant-final borrowed words (e.g. *kompjuter* ‘computer’) inflect like native masculine nouns, and *a*-final borrowed words (e.g. *opera* ‘opera’) inflect like native feminine nouns. Crucially, Figure 3 correctly predicts that *o*-final loanwords, like *torpedo* ‘torpedo’ or *kino* ‘cinema’ will inflect like native *o*-final nouns (a category that includes “fronted *o*-final” nouns, i.e. leading forms ending in an FC+*e* sequence), and that non-*o*-final loanwords (e.g. *sufi* ‘sushi’, *anime* ‘anime’) will inflect like native masculine nouns – *e*-final loanwords do not follow the pattern of *e*-final native neuter nouns (i.e. they inflect like *zavod* in Table 13, not like *bure*).

In addition, the existence of different patterns for certain nouns – such as, for instance, *jaje* ‘egg’, whose genitive singular form can be *jajeta* or *jaja* – straightforwardly comes out of the assumption that nominative singulars as leading forms have a special status. As explained in Chapter 3 (Section 3.1.3), a leading form like *jaje* can plausibly be analyzed by the speaker or learner of the language as a *j*-final stem followed by the nominative singular suffix (given that stems ending in a fronting consonant take *-e* instead of *-o*), or a leading form consisting of an *e*-final stem with no overt case suffix. Both patterns are available and attested.

5.5 Modeling and formalization

The decision tree in Figure 3 can easily be understood as a schematic representation of the evaluation of ‘if... then... else...’ statements, which necessarily impose a hierarchical structure of predicates. This ordering of structures can be crucial for the analysis of morphological systems, characterized by phenomena like inheritance or Pāṇini’s principle. This chapter demonstrates that they can be used to account for the complete inflectional system of Serbo-Croatian, including a process of assigning nominal lexical entries to inflectional classes.

Recursive schemes are an abstract way of studying algorithms (Moschovakis, 2018), which are crucial to linguistic theory. Capturing the generalizations on computational complexity of morphological and phonological processes is an important aspect that BMRs successfully fulfill. As pointed out in Chapter 3, the expressivity of BMRs is appropriately limited to subsequential functions (Bhaskar et al., 2020), while at the same time it easily expresses linguistically significant generalizations, featuring a direct capture of Pāṇini’s Principle (i.e. Elsewhere condition).

For the purposes of the analysis presented in this chapter, I update the alphabet Σ for Serbo-Croatian, initially presented in (10). Once again, I assume a set \mathbb{A} of (phonological) segments (like a or \mathfrak{z} ; (41)) and a set \mathbb{M} of morphosyntactic features (like $[\text{mas}]$ for the masculine gender, $[\text{nom}]$ for the nominative case, or $[\text{sg}]$ for the singular number; (42)), and I add a set \mathbb{L} of symbols denoting the leading form categories relevant for the target inflectional patterns ($\{a\}$ for a -final, $\{o\}$ for o -final, $\{e\}$ for e -final; (43)). The alphabet Σ represents the union of these sets (44), while $\Sigma_{\mathfrak{m}}$ is the union of Σ with the set of necessary boundary symbols (45) – here we need the morpheme boundary symbol $+$ and the inferred stem marker $*$.

$$(41) \quad \mathbb{A} = \{a, b, \mathfrak{t}, \mathfrak{f}, \mathfrak{t}\mathfrak{e}, d, \mathfrak{g}, \mathfrak{d}\mathfrak{z}, e, f, g, x, i, j, k, l, \mathfrak{A}, m, n, \mathfrak{n}, o, p, r, s, \mathfrak{f}, t, u, v, z, \mathfrak{z}\}$$

$$(42) \quad \mathbb{M} = \{[\text{mas}], [\text{fem}], [\text{neu}], [\text{nom}], [\text{gen}], [\text{dat}], [\text{acc}], [\text{voc}], [\text{ins}], \\ [\text{loc}], [\text{sg}], [\text{pl}]\}$$

$$(43) \quad \mathbb{L} = \{\{a\}, \{o\}, \{e\}\}$$

$$(44) \quad \Sigma = \mathbb{A} \cup \mathbb{M} \cup \mathbb{L}$$

$$(45) \quad \Sigma_{\infty} = \Sigma \cup \{+, *\}$$

I once again define predicates that identify particular structures in either the input or output. I assume that leading form categories, like gender features, are lexical information marked on every segment of the stem.

To take a concrete example, I represent the input and output strings of *zene* ‘woman.GEN.SG’ in Table 14. The input string begins with the stem, where each segment occupies a single position in the string. Each is also marked with the feminine gender feature $[\text{fem}]$, and a symbol denoting the final segment of the leading form (nominative singular), which for *zene* ‘woman’ is $\{a\}$. In position 4 we find the morpheme boundary symbol $+$. Finally, in positions 5 and 6 respectively, we have the case marker $[\text{gen}]$ (genitive), and the number marker $[\text{sg}]$ (singular).

Table 14. Input and output strings of the genitive singular form of SC *zene* ‘woman’

input _____

	1	2	3	4	5	6
	{ a }	{ a }	{ a }			
	[fem]	[fem]	[fem]			
	3	e	n	+	[gen]	[sg]
<u>output</u>						
Copy 1:	3	e	n		e	

With these representations in place, we can proceed to the formalization of the Serbo-Croatian noun inflection class assignment algorithm. To do this, three user-defined input-oriented predicates will suffice. As outlined in section 4.2, any lexical entry containing an *a*-final leading form will automatically be assigned to Inflectional Class II. Therefore, the predicate $II_i(x)$ (46) adds a Class II diacritic to every morpheme boundary + that is directly preceded by the { a } leading form marker in the input structure.

$$(46) \quad II_i(x) = \text{if } +_i(x) \text{ then } \{ a \}_i(p(x))$$

Similarly, the predicate $III_i(x)$ (47) also targets the morpheme boundary +. Two additional conditions are imposed: the morpheme boundary + cannot be associated with the Class II diacritic, and its precedent in the input string must bear the feminine gender feature [fem].

$$(47) \quad III_i(x) = \text{if } +_i(x) \text{ then} \\ \text{if } II_i(x) \text{ then } \perp \text{ else } [fem]_i(p(x))$$

Finally, Class I is the default inflectional class – if the leading form of a lexical entry is not *a*-final, or if it does not bear the feminine gender feature, that lexical entry will follow the Class I pattern in inflection. This is modeled in (48): the predicate $I_i(x)$ will return T for any morpheme boundary + not associated with the Class II or Class III diacritic.

$$(48) \quad I_i(x) = \text{if } +_i(x) \text{ then} \\ \quad \quad \quad \text{if } II_i(x) \text{ then } \perp \text{ else} \\ \quad \quad \quad \text{if } III_i(x) \text{ then } \perp \text{ else } \top$$

The rest of the conditions laid out in Figure 3, as well as the patterns shown in Table 13, can be captured by predicates defining exponent realization. Predicates that capture syncretic patterns in the inflectional paradigms have already been defined in (13) and (14) – I repeat them below as (49) and (50). Therefore, given that the dative and locative are always syncretic, the predicate $[dat/loc]_i(x)$ (49) returns \top for any instance of the dative $[dat]$ or locative $[loc]$ feature in the input. Similarly, $[nom/acc/voc]_i(x)$ (50) returns \top for any instance of $[nom]$, $[acc]$ or $[voc]$ in the input structure, given that the nominative, accusative and vocative are syncretic in some of the paradigms (neuters, Class II plural forms, Class III plural forms).

$$(49) \quad [dat/loc]_i(x) = \text{if } [dat]_i(x) \text{ then } \top \text{ else } [loc]_i(x)$$

$$(50) \quad [nom/acc/voc]_i(x) = \text{if } [nom]_i(x) \text{ then } \top \text{ else} \\ \quad \quad \quad \text{if } [acc]_i(x) \text{ then } \top \text{ else } [voc]_i(x)$$

Some of the necessary feature bundles – combinations of case and number – have already been modeled in (15) and (16); I repeat them below, and define the additional input predicates needed to capture all of the patterns in Table 13. In each of these, two conditions must hold for the predicate to return \top : the position under evaluation must be occupied by a certain case feature, and its direct successor must be a specific number feature.

(51) $[\text{nom sg}]_i(x) = \text{if } [\text{nom}]_i(x) \text{ then } [\text{sg}]_i(s(x))$

(52) $[\text{gen sg}]_i(x) = \text{if } [\text{gen}]_i(x) \text{ then } [\text{sg}]_i(s(x))$

(53) $[\text{acc sg}]_i(x) = \text{if } [\text{acc}]_i(x) \text{ then } [\text{sg}]_i(s(x))$

(54) $[\text{ins sg}]_i(x) = \text{if } [\text{ins}]_i(x) \text{ then } [\text{sg}]_i(s(x))$

(55) $[\text{voc sg}]_i(x) = \text{if } [\text{voc}]_i(x) \text{ then } [\text{sg}]_i(s(x))$

(56) $[\text{dat/loc sg}]_i(x) = \text{if } [\text{dat/loc}]_i(x) \text{ then } [\text{sg}]_i(s(x))$

(57) $[\text{nom/acc/voc sg}]_i(x) = \text{if } [\text{nom/acc/voc}]_i(x) \text{ then } [\text{sg}]_i(s(x))$

(58) $[\text{nom pl}]_i(x) = \text{if } [\text{nom}]_i(x) \text{ then } [\text{pl}]_i(s(x))$

(59) $[\text{gen pl}]_i(x) = \text{if } [\text{gen}]_i(x) \text{ then } [\text{sg}]_i(s(x))$

(60) $[\text{acc pl}]_i(x) = \text{if } [\text{acc}]_i(x) \text{ then } [\text{pl}]_i(s(x))$

(61) $[\text{voc pl}]_i(x) = \text{if } [\text{voc}]_i(x) \text{ then } [\text{pl}]_i(s(x))$

(62) $[\text{gen sg}]_i(x) = \text{if } [\text{gen}]_i(x) \text{ then } [\text{sg}]_i(s(x))$

$$(63) \quad [\text{nom/acc/voc pl}]_i(x) = \text{if } [\text{nom/acc/voc}]_i(x) \text{ then } [\text{sg}]_i(s(x))$$

The definition of a feature bundle predicate can also be written to capture the syncretic patterns that exist in the paradigm. As can be seen in Table 13, instrumental plural forms and dative-locative plural forms are always syncretic, regardless of the inflectional class. This is captured in (64), where $[\text{ins pl}]_i(x)$ returns T whenever the instrumental feature $[\text{ins}]$ is directly followed by the plural feature $[\text{pl}]$, or when the dative-locative feature $[\text{dat/loc}]$ is directly followed by the plural feature $[\text{pl}]$.

$$(64) \quad [\text{ins pl}]_i(x) = \text{if } [\text{ins}]_i(x) \text{ then } [\text{pl}]_i(s(x)) \text{ else} \\ \text{if } [\text{dat/loc}]_i(x) \text{ then } [\text{pl}]_i(s(x))$$

The predicates that follow model exponent realization. These are output-oriented predicates, and represent redefined versions of their counterparts in Chapter 3 – they are here updated to capture Serbo-Croatian inflectional processes beyond Inflectional Class I.

The definition of $a_o^1(x)$ in (65) is interpreted as follows. If a position on the input string is directly preceded by the Class I diacritic (line 1 in (65)), the position under evaluation will have an *a* in (the first copy of) the output string provided that (a) it is occupied by the genitive case feature $[\text{gen}]$ in the input (line 2), or (b) it is occupied by the $[\text{nom/acc/voc pl}]$ feature bundle, on an input string that also contains the neuter gender feature $[\text{neu}]$ two positions to the left (line 3). On the other hand, if the position under evaluation is directly preceded by the Class II diacritic (line 5 in (33)), it must also be associated with the $[\text{nom sg}]$ feature bundle (line 6),

or the [gen pl] feature bundle, or the [ins pl] feature bundle (line 7). Finally, if none of these conditions hold, a will be output only if there is an input a in the same position (line 8).

(65) $a_o^1(x) =$ if $I_i(p(x))$ then
 if [gen]_i(x) then T else
 if [nom/acc/voc pl]_i(x) then [neu]_i(p(p(x)))
 else
 if $II_i(p(x))$ then
 if [nom sg]_i(x) then T else
 if [gen pl]_i(x) then T else [ins pl]_i(x)
 else $a_i(x)$

In (66), the definition of $i_o^1(x)$ states that, if the position under consideration is occupied by the instrumental plural [ins pl] feature bundle in the input structure (line 1), i will surface in Copy 1 of the output string provided that [ins pl] is not preceded by the Class II diacritic in the input (line 2). If [ins pl]_i(x) returns ⊥, however, the evaluation moves on to line 4, which houses the condition that the position under consideration must be directly preceded by the Class I diacritic. This condition, if it holds, allows for the evaluation of line 5, which states that i will be output in Copy 1 if the input string contains the masculine gender feature [mas] two positions to the left, and the position under evaluation is occupied by either the nominative plural feature bundle [nom pl], or the vocative plural feature bundle [voc pl]. If the condition in line 4 does not hold, line 7 is considered: this condition states that the position must be occupied by the [dat/loc sg] feature bundle, and directly preceded by the Class II diacritic. If that condition evaluates to ⊥, the predicate considers whether the position on the input string is directly preceded by the Class III diacritic (line 8). If it is, $i_o^1(x)$ will return T provided that the position is not

occupied by the nominative singular feature bundle $[\text{nom sg}]$ (line 9), or the accusative singular feature bundle $[\text{acc sg}]$ (line 10). Finally, if the condition in line 8 does not hold, the only thing that could trigger an occurrence of an output i is the evaluation of $i_i(x)$ – i.e. the presence of an input i in the same position (line 11).

(66) $i_o^1(x) = \text{if } [\text{ins pl}]_i(x) \text{ then}$
 $\text{if } \Pi_i(p(x)) \text{ then } \perp \text{ else } \top$
 else
 $\text{if } I_i(p(x)) \text{ then}$
 $\text{if } [\text{mas}]_i(p(p(x))) \text{ then } [\text{nom pl}]_i(x) \text{ else } [\text{voc pl}]_i(x)$
 else
 $\text{if } \Pi_i(p(x)) \text{ then } [\text{dat/loc sg}]_i(x) \text{ else}$
 $\text{if } \text{III}_i(p(x)) \text{ then}$
 $\text{if } [\text{nom sg}]_i(x) \text{ then } \perp \text{ else}$
 $\text{if } [\text{acc sg}]_i(x) \text{ then } \perp \text{ else } \top$
 $\text{else } i_i(x)$

Table 13 shows that e is an exponent found in Classes I and II, but not in Class III; the evaluation of the output predicate $e_o^1(x)$ will therefore depend on conditions that reflect that. Furthermore, as explained in Chapter 3, Class I stems that end in a fronting consonant (FC) take e -initial instead of o -initial suffixes. The definition of $e_o^1(x)$ will have to capture that, too; to make that possible, I reuse the input predicate $\text{FC}_i(x)$, first defined in (25), and repeated below in (67).

(67) $\text{FC}_i(x) = \text{if } \text{ts}_i(x) \text{ then } \top \text{ else}$
 $\text{if } \text{t\textsubscript{s}}_i(x) \text{ then } \top \text{ else}$
 $\text{if } \text{dz}_i(x) \text{ then } \top \text{ else}$
 $\text{if } \text{tj}_i(x) \text{ then } \top \text{ else}$
 $\text{if } \text{d\textsubscript{z}}_i(x) \text{ then } \top \text{ else}$
 $\text{if } \text{f}_i(x) \text{ then } \top \text{ else}$
 $\text{if } \text{z}_i(x) \text{ then } \top \text{ else}$
 $\text{if } \text{n}_i(x) \text{ then } \top \text{ else}$
 $\text{if } \text{j}_i(x) \text{ then } \top \text{ else } \text{\textasciitilde{A}}_i(x)$

We can now move on to the definition of $e_o^1(x)$ in (68). If the position under consideration is directly preceded by the Class I diacritic in the input (line 1), and there is a masculine gender feature symbol $[mas]$ two positions to the left (line 2), the corresponding output Copy 1 position will contain an e as the surface form of either the vocative singular feature bundle $[voc\ sg]$ (line 3), or the accusative plural $[acc\ pl]$ (line 4), or the instrumental singular suffix $[ins\ sg]$ that follows a stem-final fronting consonant (FC, two positions to the left; line 5). If the position under consideration does not contain a masculine gender feature symbol $[mas]$ two positions to the left (line 6), the evaluation moves on to line 7: in this case, a (non-masculine) lexical entry with an FC-final stem will receive an e -initial instrumental singular suffix (line 8), and an e -final leading form (meaning that there is an $\{e\}$ diacritic two positions to the left) will have e as the exponent of the nominative-accusative-vocative singular syncretism (line 9).

If none of the above holds (line 10), and what directly precedes the position under evaluation in the input is rather the Class II diacritic (line 11), then the feature bundle that occupies the input position must be either the genitive singular $[gen\ sg]$, or the nominative-accusative-vocative plural syncretism $[nom/acc/voc\ pl]$ (line 12), in order to yield an e in output Copy 1. If neither $I_i(p(x))$ nor $II_i(p(x))$ return \top , the default condition is once again the presence of the segment in the input; the fully faithful output of an input e is thus ensured in line 13 of (68).

(68) $e_o^1(x) =$ if $I_i(p(x))$ then
 if $[mas]_i(p(p(x)))$ then
 if $[voc\ sg]_i(x)$ then \top else
 if $[acc\ pl]_i(x)$ then \top else
 if $FC_i(p(p(x)))$ then $[ins\ sg]_i(x)$
 else
 if $FC_i(p(p(x)))$ then

```

        if [ins sg]i(x) then T else
        if {e}i(p(p(x))) then [nom/acc/voc sg]i(x)
    else
        if IIi(p(x)) then
            if [gen sg]i(x) then T else [nom/acc/voc pl]i(x)
        else ei(x)

```

In the definition of $\circ_o^1(x)$ in (69), we can now define the generalization that \circ is the first segment of the output realization of the instrumental singular feature bundle [ins sg] (line 1), provided that the lexical entry has been assigned to Class I (line 2), and that [ins sg] cannot be output as e in that position (line 3). Outside of Class I, [ins sg] surfaces as \circ in output Copy 1 as long as the Class III diacritic does not occupy the position that directly precedes (line 5).

\circ is also the vocative singular suffix in Class II (line 7), and it is also the exponent of the nominative, accusative and vocative singular for all lexical entries which contain an *o*-final leading form (line 8). As in all Copy 1 predicates, if none of the more specific conditions hold, the most general, default condition on an output segment is the presence of the corresponding input segment in the same position ($\circ_i(x)$ in the final line of (69)).

```

(69)  $\circ_o^1(x)$  = if [ins sg]i(x) then
        if Ii(p(x)) then
            if  $e_o^1(x)$  then  $\perp$  else T
        else
            if IIIi(p(x)) then  $\perp$  else T
    else
        if IIi(p(x)) then [voc sg]i(x) else
        if {o}i(p(p(x))) then [nom/acc/voc sg]i(x) else  $\circ_i(x)$ 

```

The predicate $u_o^1(x)$ is defined in (70). u will be output as the dative-locative singular exponent for Class I lexemes (line 1), or as the accusative singular exponent for entries assigned

to Class II (line 2). If none of these conditions are met, u can be found in the first copy of the output string only as a faithful surface counterpart of an input u in the same position.

$$(70) \quad u_o^1(x) = \text{if } I_i(p(x)) \text{ then } [\text{dat/loc sg}]_i(x) \text{ else} \\ \text{if } II_i(p(x)) \text{ then } [\text{acc sg}]_i(x) \text{ else } u_i(x)$$

Like in Chapter 3, a copy set of size 3 (71) suffices to output the rest of the segmental content of the instrumental singular suffix, or any other exponent longer than a single character. The instrumental plural form of *zena* ‘woman’ is illustrated as an example in Table 15. The order of the output copies remains fixed as before.

$$(71) \quad C = \{1, 2, 3\}$$

Table 15. Input and output strings of the instrumental plural form of *zena* ‘woman’

	<u>input</u>					
	1	2	3	4	5	6
	{a}	{a}	{a}			
	[fem]	[fem]	[fem]			
	3	e	n	+	[ins]	[sg]
<u>output</u>						
Copy 1:	3	e	n		a	
Copy 2:					m	
Copy 3:					a	

The necessary Copy 2 output predicates are (re)defined in (72)-(74). As can be seen in Table 13, m shows up as the second segment in all instrumental plural and dative-locative plural suffixes, as well as the instrumental singular suffix in all inflectional classes but Class III. This is expressed in (72); remembering that we have already captured the instrumental plural and dative-

locative plural syncretism in (64), we do not need to specify again the presence of [dat/loc pl] in the input for *m* to surface in Copy 2 of the output.

$$(72) \quad m_o^2(x) = \text{if } [\text{ins pl}]_i(x) \text{ then } \top \text{ else} \\ \quad \quad \quad \text{if } [\text{ins sg}]_i(x) \text{ then} \\ \quad \quad \quad \text{if } \text{III}_i(p(x)) \text{ then } \perp \text{ else } \top$$

An *i* or an *a* in Copy 2 of the output string can only surface to ensure vowel length in the genitive plural exponent (see Table 13): *i* only shows up in Class III, and *a* everywhere else. This is modeled in (73) and (74), respectively.

$$(73) \quad i_o^2(x) = \text{if } [\text{gen pl}]_i(x) \text{ then } \text{III}_i(p(x))$$

$$(74) \quad a_o^2(x) = \text{if } [\text{gen pl}]_i(x) \text{ then} \\ \quad \quad \quad \text{if } \text{III}_i(p(x)) \text{ then } \perp \text{ else } \top$$

Finally, the Copy 3 output predicate we need to derive *zenama* in Table 15 – and any other instrumental plural form – is defined in (75). The instrumental plural exponents in Table 13 are the only ones requiring a Copy 3 output predicate, and they all end in *a*, which makes the definition of $a_o^3(x)$ straightforward: the only condition that needs to be met is that there is an [ins pl] feature in this position in the input.

$$(75) \quad a_o^3(x) = [\text{ins pl}]_i(x)$$

These are all the predicates that are needed to model the exponent realization patterns that are laid out in Table 13. Finally, as the morphologically conditioned *t*-insertion process, on which I focused in Chapter 3 (/bure + a/ → [bureta] ‘barrel.GEN.SG.NEU’), only occurs in Inflectional Class I, the definition of $\tau_o^1(x)$ is slightly updated in (77): this predicate now ensures that *t*-insertion will only apply at a morpheme boundary (i.e. between a stem and a suffix), and specifically one that bears a Class I diacritic (line 1 in (77)). Again, for the process to be triggered, the last segment of the stem (directly preceding the morpheme boundary) must be *e*, and at the same time there must be an output segment in the position directly following on the output string (line 2 in (77)). The latter condition is captured by the user-defined predicate *outseg*(*x*), defined in (19) and repeated below in (78), which returns T for any segment realized in output Copy 1.

$$(77) \quad \tau_o^1(x) = \begin{array}{l} \text{if } I_i(x) \text{ then} \\ \quad \text{if } e(p(x)) \text{ then } \text{outseg}(s(x)) \\ \quad \text{else } \tau_i(x) \end{array}$$

$$(78) \quad \text{outseg}(x) = \begin{array}{l} \text{if } a_o^1(x) \text{ then T else} \\ \text{if } b_o^1(x) \text{ then T else} \\ \quad (\dots) \\ \text{if } z_o^1(x) \text{ then T else } \mathfrak{z}_o^1(x) \end{array}$$

The only thing left to ensure at this point is that *t*-insertion does not kick in with *e*-final loanwords. As discussed in section 4.4, nouns like *anime* ‘anime’ inflect like native masculine nouns, i.e. following the pattern of *zavod* in Table 13. The assumption here is that such borrowings do not come with a lexically listed stem form (or listed gender information) – case endings, if applicable, are added to the leading form taken in its entirety. Therefore, to capture the fact that these lexical entries contain an inferred stem rather than a listed one, I use an additional boundary

symbol, the inferred stem marker *. Inferred stems end in *, which blocks *t*-insertion, as well as vowel fronting after such FC-final stems (Petrovic, 2022). The condition in line 2 of the predicate $\tau_o^1(x)$ in (77) is therefore not satisfied, as the predicate $e(p(x))$ does not return T.

Table 16. Input and output strings of the genitive singular form of SC *anime* ‘anime’

<u>input</u>									
1	2	3	4	5	6	7	8	9	
{e}	{e}	{e}	{e}	{e}					
a	n	i	m	e	*	+	[gen]	[sg]	
<u>output</u>									
Copy 1:									
a	n	i	m	e			a		

Finally, all other output segments are defined by Copy 1 output predicates such as those in (79)-(81). They contain no further conditions other than the requirement for the presence of the corresponding segment in the input, which will be output faithfully.

$$(79) \quad b_o^1(x) = b_i(x)$$

$$(80) \quad \tau_o^1(x) = b_i(x)$$

$$(81) \quad d_o^1(x) = d_i(x)$$

(etc., for {ʃ, ʈ, ɕ, ɗ, f, g, x, j, k, l, ʎ, n, ɲ, p, r, s, ʂ, t, v, z, ʒ})

Nothing that was not defined by output predicates will appear in the output structure. This includes -Ø case endings, as well as the undefined output counterparts of the many non-segmental input predicates. This system of logical string transductions, as defined in this section, derives the

Serbo-Croatian inflectional paradigms outlined in Table 13, incorporating the inflectional class assignment process illustrated in Figure 3.

5.6 Vowel fronting in Serbo-Croatian

The work presented here conceptualizes morphology as an independent module of transductions on strings that lies between syntax and (pure) phonology. The input to the morphological module, therefore, is understood to be the yield of a (syntactic) tree structure, while the output is a (phonological) string. For instance, in all Serbo-Croatian examples I assume the input string to have a “stem – case – number” structure, which is what these components get rearranged to prior to tree flattening. Furthermore, the morphosyntactic features map to a single exponent in the analysis above, but could each have an exponent of their own. Alternatively, morphosyntactic features could also be treated like phonological features, in which case multiple features (i.e. input predicates) could occupy (i.e. return \top for) the same input position. Such a representation is illustrated in Table 17, which is an alternative version of what was shown in Table 14:

Table 17. Input (alternative) and output strings of the dative singular form of SC *žena* ‘woman’

<u>input</u>	1	2	3	4	5
	{a}	{a}	{a}		
	[fem]	[fem]	[fem]		[sg]
	3	e	n	+	[gen]
<u>output</u>					
Copy 1:	3	e	n		e

Here I will not make any claims as to whether it would be optimal to treat morphosyntactic features as sets (unordered) or strings (ordered); I simply acknowledge that both worldviews can be accommodated. Regardless of the motivation behind the choice of the structure of the input string, however, this approach brings closer together approaches like DM to those like Paradigm Function Morphology (PFM). Works like Karttunen (2003) and Roark & Sproat (2007) cover the topic of extensional, computational equivalence of the two kinds of frameworks; however, the theories have had different intensional, theoretical assumptions and foci. With a model such as the one outlined in this chapter, the interface with syntax is overtly addressed – operations that occur over trees, pre-flattening, result in the necessary arrangement of morphosyntactic items, which will yield the input string for the morphological module of string transductions. On the other hand, BMRS can be used as a formalism for describing functions like those that lie in the core of PFM, where rules of exponence are taken to be functions from pairings of strings (stems) and morphosyntactic property sets to output strings. Proponents of PFM have argued that such an approach is particularly suitable for accounting for phenomena that are not concatenative, one-to-one mappings between input representations and surface realizations (Stump, 2001, pp. 3–12; Bonami & Stump, 2016). These insights are directly translatable into a BMRS formalization as post-flattening processes over strings.

Furthermore, the architecture of the morphological module can be expanded on, compared to what was presented in the previous sections, by assuming that the realization of a word form occurs in a step-wise fashion, within the morphological module. Such an approach makes it possible to straightforwardly model phenomena like contextual stem and affix allomorphy, as well as phonological processes that are dependent on morphological information.

Consider once again the process of suffix-initial vowel fronting in Serbo-Croatian, illustrated in Tables 7 and 8, and modeled by the predicates $e_o^1(x)$ and $o_o^1(x)$ in (26-27) and their updated versions in (68-69). With these, the realization of the initial vowel of certain suffixes (instrumental singular, neuter nominative-accusative-vocative singular syncretism) is defined as [e] if a specific condition holds (the stem is FC-final), otherwise the suffix-initial vowel surfaces as [o]. One might want to define the process differently – all suffix-initial [o]’s (in Inflectional Class I) front to [e] after FC-final stems. In other words, the instrumental singular suffix and the neuter nominative-accusative-vocative singular syncretism are realized as *-om* and *-o*, respectively, in what we could call the *exponent realization step*, while fronting of the mid back vowel happens subsequently, in the *morphophonological processes step*, taking as input the output of exponent realization. I illustrate below in Tables 18 and 19:

Table 18. Exponent realization step for the instrumental singular form of SC *poľe* ‘field’

	<u>input</u>				
	1	2	3	4	5
	{o}	{o}	{o}		
	[neu]	[neu]	[neu]	I	
	p	o	ʎ	+	[ins sg]
	<u>output</u>				
	{o}	{o}	{o}		
	[neu]	[neu]	[neu]		[ins sg]
Copy 1:	p	o	ʎ	+	o
					[ins sg]
Copy 2:					m

Table 19. Morphophonological processes step for the instrumental singular form of SC *poľe* ‘field’

	<u>input</u>					
	1	2	3	4	5	6
	{o}	{o}	{o}			
	[neu]	[neu]	[neu]	I	[ins sg]	[ins sg]

	p	o	Λ	+	o	m
<u>output</u>						
Copy 1:	p	o	Λ		e	m

The process is thus altered in several ways compared to what was outlined in the previous sections. Therefore, in comparison to $o_o^1(x)$ in (69), the predicate $o_{Eo}^1(x)$ in (82) is simplified – it now surfaces as the (first segment of the) exponent of $[ins\ sg]$ in all inflectional classes but Class III (lines 1-2), as the $[voc\ sg]$ exponent in Class II (line 4), and as the $[nom/acc/voc\ sg]$ exponent for all lexical entries that bear the $\{o\}$ leading form diacritic (line 5).¹⁵ For clarity, the exponent realization output predicates are henceforth marked with a subscript E.

$$(82) \quad o_{Eo}^1(x) = \text{if } [ins\ sg]_i(x) \text{ then} \\
\quad \quad \text{if } III_i(p(x)) \text{ then } \perp \text{ else } \top \\
\quad \quad \text{else} \\
\quad \quad \text{if } II_i(p(x)) \text{ then } [voc\ sg]_i(x) \text{ else} \\
\quad \quad \text{if } \{o\}_i(p(p(x))) \text{ then } [nom/acc/voc\ sg]_i(x) \text{ else } o_i(x)$$

Similarly, compared to $e_o^1(x)$ in (68), $e_{Eo}^1(x)$ is defined more succinctly in (83): e can now only be realized as the exponent of either the $[voc\ sg]$ or the $[acc\ pl]$ feature bundle in Class I (lines 1-3), or as the surface form of an underlying $[voc\ sg]$ or $[acc\ pl]$ feature bundle after Class II stems (lines 4-6).

$$(83) \quad e_{Eo}^1(x) = \text{if } I_i(p(x)) \text{ then} \\
\quad \quad \text{if } [mas]_i(p(p(x))) \text{ then} \\
\quad \quad \quad \text{if } [voc\ sg]_i(x) \text{ then } \top \text{ else } [acc\ pl]_i(x) \\
\quad \quad \text{else} \\
\quad \quad \text{if } II_i(p(x)) \text{ then} \\
\quad \quad \quad \text{if } [gen\ sg]_i(x) \text{ then } \top \text{ else } [nom/acc/voc\ pl]_i(x)$$

¹⁵ As foreshadowed in section 4.4, “fronted *o*-final” leading forms like *po.ŋe* ‘field.NOM.SG’ – i.e. leading forms ending in an $FC+e$ sequence – are included in the category of *o*-final leading forms, denoted by $\{o\}$.

else $e_i(x)$

Additionally, output predicates for boundary symbols, morphosyntactic features and feature bundles also need to be defined – this information needs to be retained in the output of the exponent realization step, so that predicates in the morphophonological processes step can refer to it as present in the input structure. Listing identity functions like the ones in (84) will suffice:

$$(84) \quad +_{Eo}^1(x) = +_i(x)$$

$$I_{Eo}^1(x) = I_i(x)$$

$$[neu]_{Eo}^1(x) = [neu]_i(x)$$

$$\{\circ\}_{Eo}^1(x) = \{\circ\}_i(x)$$

$$[ins \ sg]_{Eo}^1(x) = [ins \ sg]_i(x)$$

$$[ins \ sg]_{Eo}^2(x) = [ins \ sg]_i(x)$$

(etc., for all morphosyntactic features and feature bundles)

This consequently means that, in the stem selection and exponent realization steps, multiple output predicates can return T for a position x in the same output copy (e.g., both $\circ_{Eo}^1(x)$ and $[ins \ sg]_{Eo}^1(x)$ evaluate to T for position 5 in Table 18). Such transductions by themselves do not fall under the definition of well-defined transductions (Bhaskar et al., 2020, p. 163). However, the output of the exponent realization step serves as input to the morphophonological processes step, which consists solely of well-defined transductions – this means that, in the final output, only one output predicate can evaluate to T for a position x in the same output copy, i.e. exactly one character per output copy position will get printed in the end. Furthermore, as Bhaskar et al.

confirm, non-well-defined transductions do not increase the computational complexity of the system.

Finally, in the morphophonological processes step, new predicates (now marked with a subscript M) are defined. To capture the process illustrated in Table 19, I define $e_{Mo}(x)$ in (85): the predicate targets input \circ 's (line 1) that are directly preceded by the stem boundary $+$, and have a fronting consonant two positions to the left (line 2). Mid back vowel fronting after FC-final stems is thus modeled; if these conditions are not met, e 's are only output faithfully (line 3). No special conditions are imposed on $\circ_{Mo}^1(x)$ other than the presence of an \circ in the input position (86), and the definition of $FC_i(x)$ is the one given in (25) and (67).

$$(85) \quad e_{Mo}^1(x) = \begin{array}{l} \text{if } \circ_i(x) \text{ then} \\ \quad \text{if } +_i(p(x)) \text{ then } FC_i(p(p(x))) \\ \quad \text{else } e_i(x) \end{array}$$

$$(86) \quad \circ_{Mo}^1(x) = \circ_i(x)$$

Predicates like the one in (86) are defined for all other characters that we want to be output by the system (i.e. vowels and consonants); nothing else can appear in the output structure (i.e. non-segmental characters, $-\emptyset$ case endings). Any other morphophonological process would be defined via predicates similar to the ones in (85) and (86).

5.7 Discussion

The purpose of this chapter was to show that inflectional class membership does not have to be listed information, and can be a result of a predictable process. In Serbo-Croatian noun inflection, a crucial concept that makes such a process possible is that of a leading form. As leading forms have been proposed to play an important role in other languages (Albright, 2008a), the approach adopted in this chapter can be extended to, and tested on, those data, further probing the hierarchical, algorithmic nature of inflectional systems. Additionally, further work on class membership assignment will focus on how such generalizations can be learned; learning such systems would involve adding and re(de)fining parameters, evaluating productivity at every step using the Tolerance Principle (as in Belth et al., 2021).

The ‘if... then... else...’ syntax of BMRS makes a class assignment algorithm like the one presented in this chapter easy to embed in a wider system of nominal inflection. BMRSs are intuitive, implementable, and translatable to extensionally equivalent finite-state accounts. They are also easily extendable to a wide range of observed phenomena (Dolatian, 2022; Jardine & Oakden, 2022; Oakden, 2020).

More generally, this research project aims to bring theories like Distributed Morphology closer to Word-and-Paradigm approaches. Following Ermolaeva & Edmiston (2018), and compatible with the work presented here, morphological processes can be defined as string transductions over flattened binary trees. With rule ordering, this allows for a more direct comparison between different frameworks, and opens doors for taking the best of both worlds. It is then possible to abstract away from specific frameworks that are traditionally used, which gives us an opportunity to directly consider what representations are necessary and sufficient.

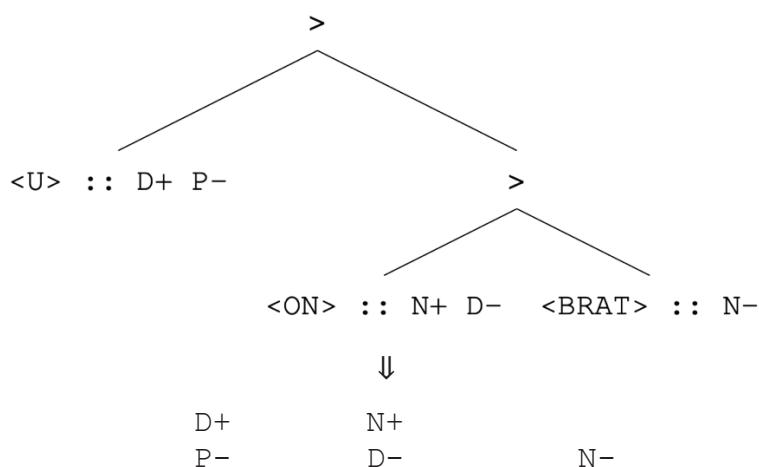
6 Syntax-morphology interface processes

In Chapter 3 and throughout the dissertation I have discussed the observation that morphological mappings are computationally at most regular (Karttunen et al., 1992) – and normally below (Chandlee, 2017) – i.e., that they are computations with a fixed memory, not requiring unbounded counting. However, in linguistic theory, mainstream formalisms like Distributed Morphology (DM; Halle & Marantz, 1993) operate over binary trees, assuming structures parallel to what one finds in syntax. Given that the strong generative capacity of such a system is at least context-free, and that such a system predicts patterns which are unattested in morphology (e.g. center embedding), the goal of this work is to outline an alternative: a model of morphology that operates over strings, featuring a transparent interface with both syntax and phonology (receiving input from the former, and feeding the latter). In this chapter, I extend the approach to model processes on the syntax-morphology interface, such as variable morphotactics, syntactically conditioned segment insertion and deletion processes, as well as total reduplication, which is often cited as a major exception to the generalization that morphological and phonological mappings are regular.

6.1 Formalization

As highlighted in section 4.6, this work conceptualizes morphology as an independent module of transductions on strings that lies between syntax and (pure) phonology. Following Ermolaeva & Edmiston (2018), I propose that the flattening of syntactic tree structure happens above the morphological module, not post-morphology as normally assumed in frameworks like DM (cf. Embick, 2010 on linearization and Vocabulary Insertion). I assume that structures in the syntactic module are trees such as those featured in Minimalist Grammars (MG; Stabler, 1997), where lexical features (selector/category, licensor/licensee) drive structure building operations (like Merge and Move). The flattening of a given syntactic derivation for the morphological module therefore yields a string formed by concatenation of the terminal symbols labeling the tree's leaf nodes, as encountered in a left-to-right traversal. Like in the analyses presented so far, multi-character symbols are allowed in these strings, as are diacritics indicating features. An example tree and its yield serving as input to the morphological module are shown in Figure 4:

Figure 4. Syntactic tree for Russian *u evo brata* ‘at his brother’s’ and its yield as input to morphology



<U> # <ON> # <BRAT>

The input string is thus made up of syntactic terminals (which have not been spelled out yet), as well as boundary symbols (morpheme boundaries are denoted by +, word boundaries by #), lexical features (structure-building, MG features) and morphosyntactic features (gender, number, etc.). The output of the morphological module is a phonological string. Like deterministic finite-state transducers (FSTs), BMRSs can be expected to reliably characterize morphological processes, without overgenerating; unlike finite-state methods, however, a logical description can capture generalizations with representations that are common in linguistics (Chandlee & Jardine, 2021).

6.2 Variable morphotactics

Syntactic terminals have no phonological form in syntax – they are spelled out in the morphological component. In a canonical morphological system, there is a constant, one-to-one mapping between a morpheme and its realization (Brown et al., 2012). More often than not, however, realizational processes are conditioned by their context; contextual allomorphy, stem formation, and variable morphotactics are just some of them. In this section I focus on modeling variable morphotactics – the phenomenon of variable surface order of morphs, resulting from a fixed order of underlying elements in the input structure.

Let us take Georgian verb inflection as an example: in this language, for a certain class of verbal stems, person and number features can surface as prefixes, or suffixes, or a combination of both. This is illustrated in Table 20:

Table 20. Present indicative paradigm for some Georgian verb types
(Harris & Samuel, 2021: 151)

	SINGULAR	PLURAL
1	<i>v-stem</i>	<i>v-stem-t</i>
2	<i>stem</i>	<i>stem-t</i>
3	<i>stem-s</i>	<i>stem-en</i>

In this paradigm, first person is marked with the prefix *v-* in both numbers, third person singular is marked with the suffix *-s*, in the first and second person plural the number exponent is a suffixal *-t*, and in the third person plural the stem takes a single suffix, *-en*. Second person is not overtly marked, and the second person singular form takes no affix (Harris & Samuel, 2021). These variable surface patterns are presumably derived from a uniform order in the underlying structure; I assume that the input string to the Georgian morphological module has a structure like “number – person – stem” (with the morpheme boundary symbol present on both sides of the stem), following the syntactic analysis proposed in McGinnis (2013). The input and output strings for different inflected forms of Georgian *ban* ‘bathe’ are laid out in (87).

(87) Georgian input:output pairs for present indicative forms of *ban* ‘bathe’

[sg] [1] +<BAN>+ → *vban*

[sg] [2] +<BAN>+ → *ban*

[sg] [3] +<BAN>+ → *bans*

[pl] [1] +<BAN>+ → *vbant*

[p1] [2]+<BAN>+ → bant

[p1] [3]+<BAN>+ → banen

Consider now what the input-output correspondences look like in the realization of the first person and third person forms of Georgian *ban* ‘bathe’ in Tables 21 and 22. The multi-character symbol <BAN> represents the present indicative stem for the Georgian verb ‘bathe’, which gets spelled out as *ban* (Harris & Samuel, 2021: 164).

Table 21. Input and output structures for the first person singular form of Georgian *ban* ‘bathe’

<u>input</u>					
	1	2	3	4	5
	[sg]	[1]	+	<BAN>	+
<u>output</u>					
Copy 1:	v		b		
Copy 2:				a	
Copy 3:				n	

Table 22. Input and output structures for the third person singular form of Georgian *ban* ‘bathe’

<u>input</u>					
	1	2	3	4	5
	[sg]	[3]	+	<BAN>	+
<u>output</u>					
Copy 1:				b	
Copy 2:				a	s
Copy 3:				n	

The necessary output predicates are defined below. As shown in Table 20, the first person feature is invariably realized as the prefix *v-*; this is defined in (93). BMRS predicates are defined using ‘if... then... else...’ clauses, which directly capture the linguistic intuition that output forms are produced via a set of hierarchically ordered, embedded, violable conditions.

$$(93) \quad v_o^1(x) = \text{if } [1]_i(x) \text{ then } \top \text{ else } \perp$$

Given that the person features and the stem node occupy constant positions in the input string, defining the transductions for variable morphotactic processes can be done with finite counting, by resorting to the predecessor and successor functions. So, in order to model the realization of third person and singular number as a single suffix, *-s*, I first define the third person singular feature bundle $[3sg]$ with the input predicate in (94). $[3sg]_i(x)$ returns \top if the position under evaluation is occupied by the stem boundary symbol $+$ in the input, and there is a third person feature $[3]$ three positions to the left on the input string, and a singular number feature $[sg]$ four positions to the left.

$$(94) \quad [3sg]_i(x) = \text{if } +_i(x) \text{ then} \\ \text{if } [3]_i(p(p(p(x)))) \text{ then } [sg]_i(p(p(p(p(x))))))$$

The output copy 2 predicate $s_o^2(x)$ is then defined in (95): *s* will be consistently output as the exponent of the third person singular feature bundle $[3sg]$ in the second copy of the output string.

$$(95) \quad s_o^2(x) = [3sg]_i(x)$$

Referring to Table 20 once again, we can see that the second person normally remains unrealized. The plural feature $[pl]$, on the other hand, is realized as a suffixal *-t*, except in the

third person plural form, where the suffix is *-en*. I illustrate the input and output strings for the first and third person plural forms of *ban* ‘bathe’ in Tables 23 and 24 below.

Table 23. Input and output structures for the first person plural form of Georgian *ban* ‘bathe’

<u>input</u>					
	1	2	3	4	5
	[p1]	[1]	+	<BAN>	+
<u>output</u>					
Copy 1:	v		b		
Copy 2:				a	t
Copy 3:	n				

Table 24. Input and output structures for the third person plural form of Georgian *ban* ‘bathe’

<u>input</u>					
	1	2	3	4	5
	[p1]	[3]	+	<BAN>	+
<u>output</u>					
Copy 1:				b	
Copy 2:				a	e
Copy 3:				n	n

I assume that the plural feature surfaces as a suffixal *t* whenever possible – i.e. whenever the [3] symbol is not directly preceded by [p1] in the input string, in which case an *e* is output in the target position (directly following the stem). I first define the third person plural bundle predicate $[3p1]_i(x)$ in (96), similarly to $[3sg]_i(x)$ in (94). The definition of $e_o^2(x)$ in (97) then states that *e* surfaces in the second copy of the output string if a string index is occupied by the third person plural feature bundle [3p1] in the input.

$$(96) \quad [3p1]_i(x) = \text{if } +_i(x) \text{ then} \\ \text{if } [3]_i(p(p(p(x)))) \text{ then } [p1]_i(p(p(p(p(x))))))$$

$$(97) \quad e_o^2(x) = [3p1]_i(x)$$

The target position for $t_o^2(x)$ (98) is the same (stem boundary symbol + with a plural feature $[p1]$ four positions to the left in the input); this predicate can only return T if $e_o^2(x)$ does not evaluate to T for the same index.

$$(98) \quad t_o^2(x) = \text{if } +_i(x) \text{ then} \\ \quad \quad \text{if } [p1]_i(p(p(p(x)))) \text{ then} \\ \quad \quad \text{if } e_o^2(x) \text{ then } \perp \text{ else T}$$

Finally, I define the segments that comprise the realization of $\langle \text{BAN} \rangle$ in (13), (14), and (15). Note in the definition of $n_o^3(x)$ in (101) that n can surface in output copy 3 as part of the $[3p1]$ feature bundle, as well as a part of the present indicative stem *ban* ‘bathe’.

$$(99) \quad b_o^1(x) = \langle \text{BAN} \rangle_i(x)$$

$$(100) \quad a_o^2(x) = \langle \text{BAN} \rangle_i(x)$$

$$(101) \quad n_o^3(x) = \text{if } \langle \text{BAN} \rangle_i(x) \text{ then T else } [3p1]_i(x)$$

Here I do not show the intermediate stages (in which the boundary symbols and morphosyntactic features are retained) between inputs consisting of syntactic terminals and phonological outputs. As established in section 4.6, however, any morphophonological process

that applies over strings generated in the exponent realization step would be sensitive to such elements, and therefore the corresponding output predicates would need to be defined.

6.3 Total reduplication

Total reduplication is normally understood to be a major exception to the generalization that morphology is computationally at most regular (cf. Dolatian & Heinz, 2020). This stems from the observation that, if we are dealing with strings (of segments) of arbitrary length, the copying involved is unbounded. However, assuming an input structure that is made up of unrealized morphemes (such that they occupy a single position each in the input string) is a way to model total reduplication as a (sub)regular pattern – such copying does not exceed the observed computational complexity of morphological processes. I illustrate this with an example from Indonesian, where plurality is marked by total reduplication (Cohn, 1989):

Table 25. Input and output structures for the plural form of Indonesian *buku* ‘book’

<u>input</u>			
	1	2	3
	<BUKU>	+	[p l]
<u>output</u>			
Copy 1:	b		b
Copy 2:	u		u
Copy 3:	k		k
Copy 4:	u		u

Note that this account predicts that unbounded reduplication can only involve a finitely bounded number of syntactic terminals. In addition, while the copying of a stem, as illustrated in

Table 25 (i.e. identifying the input structure $\langle \text{BUKU} \rangle + [\text{p1}]$ as $\langle \text{BUKU} \rangle + \langle \text{BUKU} \rangle$), is not unbounded copying, unbounded phonological exponence cannot be captured in a BMRS analysis. The assumption here is that the maximum number of output copies is determined by the longest word in the (by definition, finite) lexicon. Any addition of a longer word would mean updating the lexicon, which would entail updating the size of the copy set – as such, this is a learning question, and falls outside of the scope of the present work. Furthermore, assuming finitely many copies rules out unbounded total reduplication of forms with morphological recursion (like, for instance, *great great great (...) grandfather* \rightarrow *great great great (...) grandfather great great great (...) grandfather*), which is unattested.

The pattern is captured with the input predicate $\langle \text{BUKU} \rangle_i(x)$ in (102): it is defined to return T for any position occupied by the plural feature $[\text{p1}]$ in the input string, which also contains the stem $\langle \text{BUKU} \rangle$ two positions to the left.

$$(102) \quad \langle \text{BUKU} \rangle_i(x) = \text{if } [\text{p1}]_i(x) \text{ then } \langle \text{BUKU} \rangle_i(p(p(x)))$$

The segment output predicates in (103–106) then ensure that all the segments of *buku* ‘book’ are output in the correct order, for each index that evaluates to T for $\langle \text{BUKU} \rangle_i(x)$.

$$(103) \quad b_o^1(x) = \langle \text{BUKU} \rangle_i(x)$$

$$(104) \quad u_o^2(x) = \langle \text{BUKU} \rangle_i(x)$$

$$(105) \quad k_o^3(x) = \langle \text{BUKU} \rangle_i(x)$$

$$(106) \quad u_o^4(x) = \langle \text{BUKU} \rangle_i(x)$$

Similar predicates can then be defined for any lexical entry that derives its plural form via total reduplication. Specifically, we would need predicates like $\langle \text{BUKU} \rangle_i(x)$ (102) for every member of the fixed collection of Indonesian stems. So, given a finite set of lexical stems \mathbb{S} in (107), containing entries like *buku* ‘book’, *wanita* ‘woman’, *masyarakat* ‘society’, *hak* ‘right’, *kera* ‘monkey’ etc., we can combine all elements of \mathbb{S} via the logical conjunction operator \wedge (108):

$$(107) \quad \mathbb{S} = \{ \langle \text{BUKU} \rangle, \langle \text{WANITA} \rangle, \langle \text{MASYARAKAT} \rangle, \langle \text{HAK} \rangle, \langle \text{KERA} \rangle, \dots \}$$

$$(108) \quad \bigwedge_{q \in \mathbb{S}} (q_i(x) = \text{if } [p1]_i(x) \text{ then } q_i(p(p(x))))$$

Additional length and/or structure constraints may also be placed on the reduplicant in order to model partial reduplication (Moravcsik, 1978), or reduplication with fixed segmentism (Alderete et al., 1999). In a system that relies on an alphabet of phonological features rather than segments, capturing ablaut reduplication (Minkova, 2002) would also be straightforward. For a BMRS analysis that uses phonological features instead of segments, see Chandlee & Jardine (2021).

6.4 Syntactically conditioned segmental processes

Processes involving segment insertion or deletion can be even further constrained by syntactic conditions. Consider Russian, where *n* is inserted before third person possessive pronouns, but not third person personal pronouns (Philippova, 2018):

(109) (a) *u evo* → *u nevo* ‘at him’ vs. (b) *u evo brata* ‘at his brother’s’

In both cases, *evo* is a third person pronoun, the genitive form of *on* ‘he’; in (109a), however, *evo* does not take a nominal complement, whereas in (109b) it necessarily modifies a noun. In cases like (109b), the syntactic terminal <ON> is thus assumed to bear a lexical selector feature N+, which checks the succeeding noun’s category feature N– in the syntax, resulting in the structure building operation Merge. The hierarchical (constituent) structure is lost after linearization, but the selector/category features are retained, and can be referred to. The input and output structures of strings in (109a) and (109b) are illustrated in Tables 26 and 27, respectively. In Russian, the preposition *u* ‘at’ assigns genitive case to its complement; *evo* ‘him/his’ and *brata* ‘brother’s’ therefore bear the genitive feature [gen].

Table 26. Input and output structures for Russian *u nevo* ‘at him’

	<u>input</u>		
	1	2	3
			D–
	D+		[3]
	P–		[gen]
	<U>	#	<ON>
<u>output</u>			
	Copy 1:	u	n e
	Copy 2:		v
	Copy 3:		o

Table 27. Input and output structures for Russian *u evo brata* ‘at his brother’s’

<u>input</u>					
1	2	3	4	5	
		N+			
		D-			
D+		[3]		N-	
P-		[gen]		[gen]	
<U>	#	<ON>	#	<BRAT>	
<u>output</u>					
Copy 1:	u	e		b	
Copy 2:		v		r	
Copy 3:		o		a	
Copy 4:				t	
Copy 5:				a	

The conditions under which an *n* is epenthesized in Table 26 (and not in Table 27) are captured in (110): *n*-insertion occurs only at a word boundary (#; line 1), directly following a preposition (P-; line 2) in a prepositional phrase (i.e. the preceding preposition also selects a *D* element; line 3). The position that immediately follows needs to be occupied by a pronoun in the input (D-; line 4), specifically a third person pronoun ([3]; line 5). Crucially, the third person pronoun cannot be one that takes an *N* complement (line 6) – *n*-insertion fails in that case, otherwise it occurs (provided that all the other conditions are met).

$$\begin{aligned}
 (110) \quad n_o^1(x) = & \text{if } \#_i(x) \text{ then} \\
 & \text{if } P-_i(p(x)) \text{ then} \\
 & \quad \text{if } D+_i(p(x)) \text{ then} \\
 & \quad \quad \text{if } D-_i(s(x)) \text{ then} \\
 & \quad \quad \quad \text{if } [3]_i(s(x)) \text{ then} \\
 & \quad \quad \quad \quad \text{if } N+_i(s(x)) \text{ then } \perp \text{ else } T
 \end{aligned}$$

Given the high degree of embedding, one might prefer to define $n_o^1(x)$ with the binary logical conjunction operator \wedge , as in (110'). The two definitions are equivalent:

$$(110') \quad n_o^1(x) = \text{if } \#_i(x) \wedge P_{-i}(p(x)) \wedge D_{+i}(p(x)) \wedge D_{-i}(s(x)) \wedge [\exists]_i(s(x)) \text{ then} \\ \text{if } N_{+i}(s(x)) \text{ then } \perp \text{ else } \top$$

The form *evo* is the genitive form of the third person singular masculine pronoun *on* ‘he’ in Russian; the realization of <ON> thus relies on the presence of the [gen] feature, as shown in (111–113) below:

$$(111) \quad e_o^1(x) = \text{if } \langle \text{ON} \rangle_i(x) \text{ then } [\text{gen}]_i(x)$$

$$(112) \quad v_o^2(x) = \text{if } \langle \text{ON} \rangle_i(x) \text{ then } [\text{gen}]_i(x)$$

$$(113) \quad o_o^3(x) = \text{if } \langle \text{ON} \rangle_i(x) \text{ then } [\text{gen}]_i(x)$$

Similarly, *brata* is the genitive form of Russian *brat* ‘brother’; the predicates in (114–117) therefore model the realization of the nominal stem for this lexeme, whereas $a_o^5(x)$ in (118) defines its exponent of the genitive case. Finally, Russian <U> is always output as *u* (119).

$$(114) \quad b_o^1(x) = \langle \text{BRAT} \rangle_i(x)$$

$$(115) \quad r_o^2(x) = \langle \text{BRAT} \rangle_i(x)$$

$$(116) \quad a_o^3(x) = \langle \text{BRAT} \rangle_i(x)$$

$$(117) \quad t_o^4(x) = \langle \text{BRAT} \rangle_i(x)$$

(118) $a_o^5(x) = \text{if } \langle \text{BRAT} \rangle_i(x) \text{ then } [\text{gen}]_i(x)$

(119) $u_o^1(x) = \langle U \rangle_i(x)$

A different syntactically conditioned process occurs in German: the preposition *zu* [tsu:] and the dative singular masculine determiner *dem* [dem], when adjacent, can be contracted to *zum* [tsu:m] if *dem* is in the complement of *zu*.¹⁶ Consider the examples below (Thomas Graf, personal communication):

(120) Ich ging [PP **zu** [DP **dem** Bahnhof]]
 I walked to the.DAT.SG.MASC train-station.DAT.SG.MASC
 ‘I went to the train station.’

(120') Ich ging **zum** Bahnhof

Both (120) and (120') are grammatical. In (120), the dative case is assigned to the DP by the preposition *zu*. Interestingly, there is a possessive construction in German in which the possessor also carries dative case: the form in (121) reflects a masculine possessor, and the possessee inflects for whatever case it is assigned from beyond the DP.

(121) [DP [DP dem NP] [D' sein(Ø/e/er/em/en) NP]]
 the.DAT.SG.MASC his

¹⁶ Article contraction in German is optional; I disregard this for the purposes of the present work, and model the process with BMRS, which are equivalent to subsequential transductions, and are thus deterministic. Depending on the nature of the optionality, variability in the output would be generated by different speaker grammars (inter-speaker variability), or a special marker in a single system that would distinguish the representations which trigger contraction from those which do not (intra-speaker variability).

So, *zu* and *dem* can surface as adjacent words in different syntactic structures – while *dem* is the head of the complement of *zu* in (120), it is the head of the specifier of the complement of *zu* in (122). Contraction is possible in both cases:

(122) Ich ging [PP **zu** [DP [DP **dem** Künstler] [D' seiner Bildergalerie]]]
 I walked to the.DAT.SG.MASC artist.DAT.SG.MASC his.DAT.SG gallery.DAT.SG
 'I went to the artist's gallery.'

(122') Ich ging **zum** Künstler seiner Bildergalerie

Even bigger structures can be constructed recursively. Consider the examples in (123-124); the structure of the PP in the example in (124) is illustrated in Figure 5.

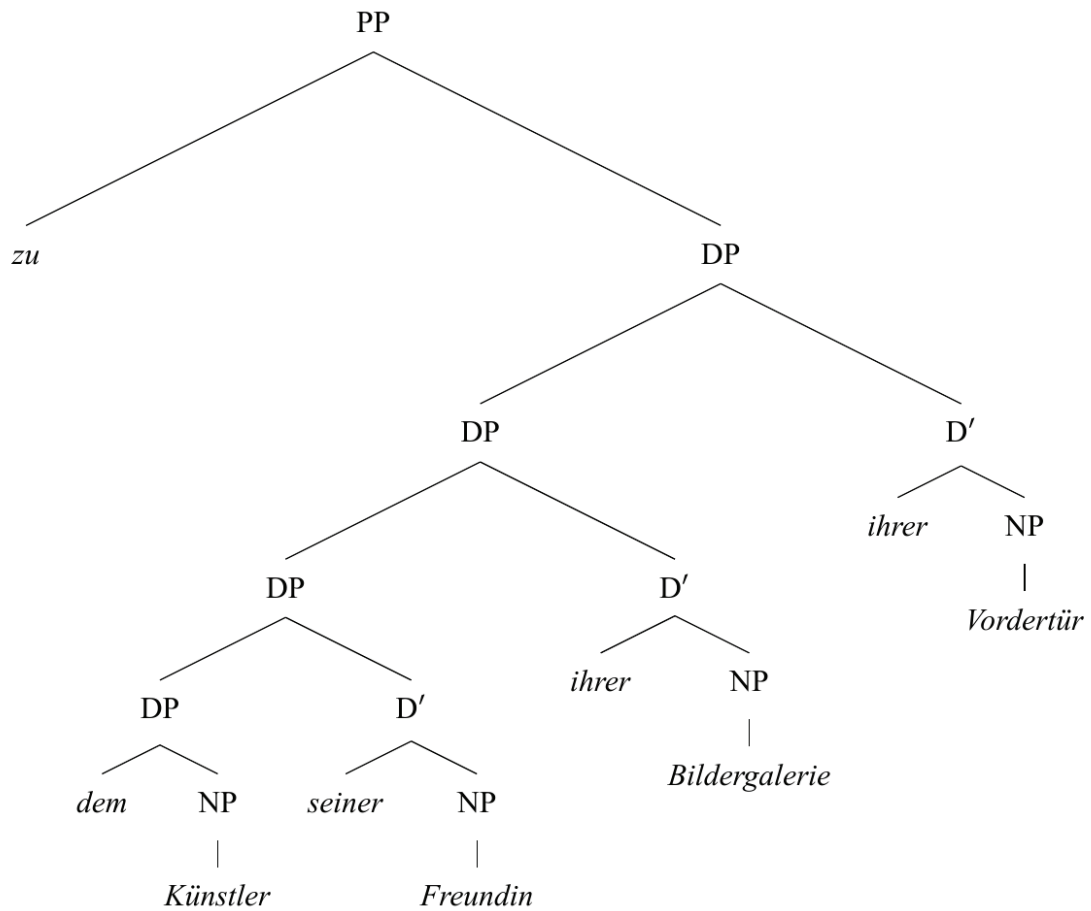
(123) Ich ging **zu dem** Künstler seiner Freundin ihrer Bildergalerie
 I walked to the artist his girlfriend her gallery
 'I went to the artist's girlfriend's gallery'

(123') Ich ging **zum** Künstler seiner Freundin ihrer Bildergalerie

(124) Ich ging **zu dem** Künstler seiner Freundin ihrer Bildergalerie ihrer Vordertür
 I walked to the artist his girlfriend her gallery.FEM her front-door
 'I went to the artist's girlfriend's gallery's front door'

(124') Ich ging **zum** Künstler seiner Freundin ihrer Bildergalerie ihrer Vordertür

Figure 5. Syntactic structure of the prepositional phrase in (124).



Motivating and modeling German *zum* contraction to apply over syntactic trees would be difficult; as shown in the examples above, this can prove to be an extremely non-local process.

After linearization, however, *zu* and *dem* are always adjacent.¹⁷ Note that knowing the syntactic category of *zu* is crucial – contraction to *zum* is impossible with homophonous words. This is illustrated in (125) with the verb *zugeben* ‘admit’, which is made up of *geben* ‘give’ and the verbal particle *zu*. Here, (125') is ungrammatical:

(125) Ich gab zu dem Verbrecher geholfen zu haben
 I gave *zu* the.DAT.SG criminal.DAT.SG helped to have
 ‘I admitted that I had helped the criminal’

(125') *Ich gab zum Verbrecher geholfen zu haben

Therefore, *zu* needs to be a P-head for contraction to occur. Similarly, *dem* can be used as both an article and a demonstrative: the process is not triggered by the latter. This is shown in (126) and (126'):

(126) Ich ging zu dem.
 I walked to DEM
 ‘I went to that one.’

(126') *Ich ging zum.

With this in mind, we can formalize the process using representations illustrated in Table 28. The lexical item <ZU>, being a preposition, bears the category feature P–, as well as a selector

¹⁷ This is, technically, a simplification – if empty functional heads are retained after linearization (which would be required for syntactically conditioned processes more generally), there would be a C-head or T-head intervening in (125) below, for instance. In a model that requires this, these syntactic heads would be blocking the contraction process instead of the features on lexical items. For the purposes of this chapter, the simplified analysis is adequate.

feature $D+$, as it forms prepositional phrases with DP complements. Similarly, the article $\langle DEM \rangle$ bears the category feature $D-$, and the selector feature $N+$. Additionally, to distinguish it from the demonstrative (which would be marked with a $[Dem]$ feature), I assume the article also has an $[Art]$ lexical feature.

Table 28. Input and output structures for German *zum* [tsu:m] (*< zu dem*)

<u>input</u>			
	1	2	3
			$N+$
	$D+$		$D-$
	$P-$		$[Art]$
	$\langle ZU \rangle$	$\#$	$\langle DEM \rangle$
<u>output</u>			
Copy 1:	ts		
Copy 2:	$\text{u} :$		
Copy 3:	m		

Article contraction is formalized here as non-realization of the two initial segments of *dem* under the conditions outlined above. Those conditions are captured in the definition of $d_o^1(x)$ in (127): a d surfaces in output copy 1 if the position under evaluation is occupied by the lexical item $\langle DEM \rangle$ in the input (line 1), which also bears the article feature $[Art]$ (line 2). At the same time, the preceding word – two positions to the left – must be a preposition (bearing the lexical feature $P-$; line 3) that selects a D element ($D+$; line 4). These are blocking conditions: if they all hold, a d will not be output – otherwise it will (lines 4-6).

(127) $d_o^1(x) = \text{if } \langle DEM \rangle_i(x) \text{ then}$

if $[Art]_i(x)$ then
 if $P_{-i}(p(p(x)))$ then
 if $D_{+i}(p(p(x)))$ then \perp else T
 else T
 else T

The vowel e appears in output copy 2 as part of the realization of $\langle DEM \rangle$ only if a d is output in the same position in output copy 1 (128). Therefore, the same conditions hold as in (127).

$$(128) \quad e_o^2(x) = \text{if } \langle DEM \rangle_i(x) \text{ then } d_o^1(x)$$

An m , on the other hand, always surfaces in output copy 3 for an input $\langle DEM \rangle$ (129), regardless of whether *zu* and *dem* are contracted or not. Similarly, \mathfrak{s} (130) and $u:$ (131) are constant realizations of the lexical item $\langle ZU \rangle$, as illustrated in Table 28.

$$(129) \quad m_o^3(x) = \langle DEM \rangle_i(x)$$

$$(130) \quad \mathfrak{s}_o^1(x) = \langle ZU \rangle_i(x)$$

$$(131) \quad u:o^2(x) = \langle ZU \rangle_i(x)$$

In this way, syntactically conditioned morpho(phono)logical processes can be formalized as logical transductions on strings, which are understood to be augmented with additional lexical, morphological, and syntactic information. This work treats such computations as transformations between syntactic structures and phonological material, determining what is necessary and

sufficient for their modeling under various conditions, and formalizing the relevant theoretical and empirical generalizations. I show that phenomena like variable morphotactics, total reduplication, syntactically conditioned segment insertion and article contraction are (a) sensitive to lexical and syntactic information, and (b) definable as operating over strings, involving only bounded counting.

7 Conclusion

The work presented in this dissertation is mainly about asking what is necessary and sufficient for modeling different morphological processes under various conditions, addressing the implications of different assumptions, and providing adequate accounts. My intention was to show that examining processes on the interfaces of morphology with phonology and syntax can be understood as a formal study of computations. Therefore, the broad objective of this dissertation was to examine the relationships between structures and representations in morphosyntax and morphophonology, with a focus on formalizing the interfaces.

Previous work in computational linguistics has demonstrated that a vast majority of morphological and phonological phenomena are restricted in a way that is dependent on the notions of computational complexity and locality. I apply this insight to a case study of Serbo-Croatian morphological consonant epenthesis, and use the findings as a point of departure to characterize the morphological module more generally, as well as to detect related phenomena in different languages. Additionally, I show that the inherently hierarchical nature of the proposed

architecture allows the integration of well-understood grammatical principles such as inflectional class assignment algorithms and the use of leading forms.

The formal relationship between representations and computation sheds light on the nature of morphological processes on the interfaces. As exhibited in the dissertation, “non-canonical” processes like morphologically and syntactically conditioned epenthesis present difficulties to theories of morphology that are either syntax-based or fully merged with phonological computation, resulting in the morphological interfaces lacking unified formal integration. I have worked to remedy this by providing a principled way to discuss representations and computation across modules using model theory and logical characterizations. In this way, my work derives the observation that morphophonology is not necessarily phonologically optimizing, and that it has to apply before phonology. Syntactically conditioned processes are similarly handled with the representational capacities of the morphological module that I explore, meaning morphology is sensitive to syntactic, phonological, and purely morphological representations.

References

- Albright, A. (2008a). Explaining universal tendencies and language particulars in analogical change. *Linguistic Universals and Language Change*, 144–181.
- Albright, A. (2008b). Inflectional paradigms have bases too: Arguments from Yiddish. In A. Bachrach & A. Nevins (Eds.), *Inflectional Identity* (pp. 271–312). Oxford University Press.
- Alderete, J., Beckman, J., Benua, L., Gnanadesikan, A., McCarthy, J., & Urbanczyk, S. (1999). Reduplication with fixed segmentism. *Linguistic Inquiry*, 30(3), 327–364.
- Anderson, S. R. (1992). *A-morphous morphology* (Issue 62). Cambridge University Press.
- Aronoff, M. (1976). *Word formation in generative grammar*. MIT Press.
- Aronoff, M. (1994). *Morphology by itself: Stems and inflectional classes* (Vol. 22). MIT Press. <http://mitpress.mit.edu/9780262510721/>
- Aronoff, M., & Repetti, L. (2022). Conditioned epenthesis in Romance. In A. Ledgeway, J. C. Smith, & N. Vincent (Eds.), *Periphrasis and Inflexion in Diachrony: A View from Romance* (pp. 362–380). Oxford University Press. <https://doi.org/10.1093/oso/9780198870807.003.0014>

- Arregi, K., & Nevins, A. (2012). *Morphotactics: Basque auxiliaries and the structure of spellout* (Vol. 86). Springer Science & Business Media. DOI: <https://doi.org/10.1007/978-94-007-3889-8>
- Artés Cuenca, E. (2016). *The influence of phonology on inflection: The interplay between syllabification and lexical insertion in Pallarese Catalan* [PhD dissertation, Universitat Autònoma de Barcelona]. <https://ddd.uab.cat/record/167840>
- Bachrach, A., & Wagner, M. (2007). *Syntactically driven cyclicity vs. Output-output correspondence: The case of adjunction in diminutive morphology*. <https://ling.auf.net/lingbuzz/000383/v1.pdf>
- Barić, E., Lončarić, M., Malić, D., Pavešić, S., Peti, M., Zečević, V., & Znika, M. (1997). *Hrvatska gramatika*. Školska knjiga.
- Belth, C. A., Payne, S. R., Beser, D., Kodner, J., & Yang, C. (2021). The Greedy and Recursive Search for Morphological Productivity. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43, 2869–2875.
- Bhaskar, S., Chandlee, J., Jardine, A., & Oakden, C. (2020). Boolean Monadic Recursive Schemes as a Logical Characterization of the Subsequential Functions. In A. Leporati, C. Martín-Vide, D. Shapira, & C. Zandron (Eds.), *Language and Automata Theory and Applications* (pp. 157–169). Springer International Publishing.
- Blake, B. J. (2001). *Case*. Cambridge University Press.
- Bonami, O., & Stump, G. T. (2016). Paradigm Function Morphology. In A. Hippisley & G. T. Stump (Eds.), *Cambridge Handbook of Morphology* (pp. 449–481). Cambridge University Press. <https://shs.hal.science/halshs-01428292>
- Broselow, E. (1984). Default consonants in Amharic morphology. *MIT Working Papers in Linguistics*, 7, 15–31.
- Brown, D., Chumakina, M., & Corbett, G. G. (Eds.). (2012). *Canonical Morphology and Syntax*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199604326.001.0001>
- Brozović, D. (2006). *Neka bitna pitanja hrvatskoga jezičnog standarda*. Školska knjiga.
- Chandlee, J. (2017). Computational locality in morphological maps. *Morphology*, 27(4), 599–641.
- Chandlee, J., & Jardine, A. (2021). Computational universals in linguistic theory: Using recursive programs for phonological analysis. *Language*, 97(3), 485–519.
- Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, 2(2), 137–167.

- Cohn, A. C. (1989). Stress in Indonesian and bracketing paradoxes. *Natural Language & Linguistic Theory*, 7(2), 167–216.
- Crowhurst, M. J. (1992). Diminutives and augmentatives in Mexican Spanish: A prosodic analysis. *Phonology*, 9(2), 221–253.
- Dolatian, H. (2020). *Computational locality of cyclic phonology in Armenian* [PhD dissertation, Stony Brook University].
<https://you.stonybrook.edu/deovlet/files/2020/09/Dissertation.pdf>
- Dolatian, H. (2022, September 22). *Implementing logic-based approaches to phonology*. Workshop on Model Theoretic Representations in Phonology, Stony Brook University.
<https://www.jeffreyheinz.net/events/WMTRPprogram.html>
- Dolatian, H., & Heinz, J. (2020). Computing and classifying reduplication with 2-way finite-state transducers. *Journal of Language Modelling*, 8(1), 179–250.
- Embick, D. (2010). *Localism versus globalism in morphology and phonology* (Vol. 60). MIT Press.
- Ermolaeva, M., & Edmiston, D. (2018). Distributed morphology as a regular relation. *Proceedings of the Society for Computation in Linguistics*, 1(1), 178–181.
- Filiot, E. (2015). Logic-automata connections for transformations. *Indian Conference on Logic and Its Applications*, 30–57.
- Garcia, G. (2017). Weight gradience and stress in Portuguese. *Phonology*, 34, 41–79.
- Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In *The view from building 20* (pp. 111–176). The MIT Press.
- Harris, A. C., & Samuel, A. G. (2021). The suffixing preference. In S. Moradi, M. Haag, J. Rees-Miller, & A. Petrovic (Eds.), *All Things Morphology: Its independence and its interfaces* (Vol. 353, pp. 147–168). John Benjamins Publishing Company.
<https://doi.org/10.1075/cilt.353.09har>
- Heinz, J. (2018). The computational nature of phonological generalizations. In L. M. Hyman & F. Plank (Eds.), *Phonological Typology* (pp. 126–195).
- Inkelas, S., & Zec, D. (1990). *The phonology-syntax connection*. University of Chicago Press.
- Jardine, A., & Oakden, C. (2022). *Process-specific constraint effects in BMRS* [Unpublished manuscript]. <https://adamjardine.net/files/jardineoakdenBMRSPSC.pdf>
- Johnson, C. D. (1972). *Formal aspects of phonological description*. Mouton.
- Kaplan, R. M., & Kay, M. (1994). Regular models of phonological rule systems. *Computational Linguistics*, 20(3), 331–378.

- Karttunen, L. (2003). *Computing with realizational morphology*. 203–214.
- Karttunen, L., Kaplan, R. M., & Zaenen, A. (1992). Two-level morphology with composition. *Proceedings of COLING 1992: The 15th International Conference on Computational Linguistics, 1*, 141–148.
- Kiparsky, P. (1982). Lexical morphology and phonology. In I.-S. Yang (Ed.), *Linguistics in the morning calm: Selected papers from SICOL-1981* (pp. 3–91). Hansin.
- Kiparsky, R. P. V. (1973). “Elsewhere” in phonology. In S. R. Anderson & P. Kiparsky (Eds.), *Festschrift for Morris Halle* (pp. 93–106). Holt, Rinehart, and Winston.
- Klajn, I. (2005). *Gramatika srpskog jezika*. Zavod za udžbenike i nastavna sredstva.
- Koskeniemi, K. (1983). *Two-level morphology: A general computational model for word-form recognition and production* [PhD dissertation, University of Helsinki].
<http://www.ling.helsinki.fi/~koskenni/doc/Two-LevelMorphology.pdf>
- Leslau, W. (1997). *Ethiopic documents: Argobba: Grammar and dictionary* (Vol. 47). Otto Harrassowitz Verlag.
- Lombardi, L. (2002). Coronal epenthesis and markedness. *Phonology*, 219–251.
- Loporcaro, M. (1997). On vowel epenthesis in Alguer Catalan. In P. M. Bertinetto, L. Gaeta, G. Jetchev, & D. Michaels (Eds.), *Certamen Phonologicum III: Papers from the Third Cortona Phonology Meeting* (pp. 205–227). Rosenberg and Sellier.
- Marković, I. (2013). *Hrvatska morfonologija*. Disput.
- Matasović, R. (2008). *Poredbenopovijesna gramatika hrvatskoga jezika*. Matica hrvatska.
- McCarthy, J. J. (2005). Optimal paradigms. In L. J. Downing, T. A. Hall, & R. Raffelsiefen (Eds.), *Paradigms in Phonological Theory* (pp. 170–210).
- McGinnis, M. (2013). Agree and Fission in Georgian Plurals. In O. Matushansky & A. Marantz (Eds.), *Distributed Morphology Today: Morphemes for Morris Halle* (pp. 39–58). The MIT Press. <https://doi.org/10.7551/mitpress/9780262019675.003.0003>
- Merchant, J. (2015). How much context is enough? Two cases of span-conditioned stem allomorphy. *Linguistic Inquiry*, 46(2), 273–303. https://doi.org/10.1162/LING_a_00182
- Minkova, D. (2002). Ablaut reduplication in English: The criss-crossing of prosody and verbal art. *English Language & Linguistics*, 6(1), 133–169.
- Moradi, S. (2017). *Non-canonical Epenthesis: Epenthetic quality and the role of morphonology* [Unpublished manuscript].
https://www.academia.edu/100511567/Non_canonical_Epenthesis_Epenthetic_quality_and_the_role_of_morphonology

- Moravcsik, E. A. (1978). Reduplicative constructions. In J. H. Greenberg (Ed.), *Universals of Human Language* (Vol. 1, pp. 297–334). Stanford University Press.
- Moschovakis, Y. N. (2018). *Abstract recursion and intrinsic complexity* (Vol. 48). Cambridge University Press.
- Norrmann-Vigil, I. (2012). Accounting for variation of diminutive formation in Porteño Spanish. *Mester*, 41(1). <https://doi.org/10.5070/M3411011449>
- Oakden, C. (2020). *The Ticuna nominalizer tone circle and input strict-locality* [Unpublished manuscript]. https://chrisoakden.com/files/ticuna_manuscript.pdf
- Payne, D. L. (1981). *Phonology and morphology of Axininca Campa*. Summer Institute of Linguistics and the University of Texas at Arlington. <https://peru.sil.org/resources/archives/8521>
- Petrovic, A. (2022). Vowel fronting in Serbian noun inflection. In A. Aitha, S. Castro, & B. Wilson (Eds.), *Proceedings of the Fifty-seventh Annual Meeting of the Chicago Linguistic Society (CLS 57)*. The Chicago Linguistic Society. <http://chicagolinguisticsociety.org/public/cls57-toc.pdf>
- Petrovic, A. (2023). A formal account of morphological epenthesis in Serbo-Croatian. *Morphology*, 33(3). <https://doi.org/10.1007/s11525-023-09412-9>
- Petrovic, A. (under review). Where do they belong: Serbo-Croatian noun stems need leading forms. *Journal of Slavic Linguistics (FASL 31 Extra Issue)*.
- Philippova, T. (2018). *Prepositional repercussions in Russian: Pronouns, comparatives and ellipsis* [Doctoral dissertation, Ben-Gurion University of the Negev]. <https://ling.auf.net/lingbuzz/004016/current.pdf>
- Piggott, G. L. (1980). *Aspects of Odawa morphophonemics*. Garland.
- Roark, B., & Sproat, R. (2007). *Computational approaches to morphology and syntax* (Vol. 4). OUP Oxford.
- Scheer, T. (2011). *A guide to morphosyntax-phonology interface theories: How extra-phonological information is treated in phonology since Trubetzkoy's Grenzsignale*. Mouton de Gruyter.
- Šljivić-Šimšić, B. (1984). Neuter nouns in -Ø or neuter nouns in -e with extended stems in Standard Serbo-Croatian. *Folia Slavica*, 6(3), 372–388.
- Stabler, E. (1997). Derivational Minimalism. *Logical Aspects of Computational Linguistics: First International Conference, LACL '96, Nancy, France, September 23-25, 1996. Selected Papers*, 1328, 68.

- Staroverov, P. (2014). *Splitting theory and consonant epenthesis* [Rutgers, The State University of New Jersey].
<https://www.proquest.com/openview/76a6537dfac768cbbc0ea645a83605dc/>
- Stevanović, M. (1969). *Savremeni srpskohrvatski jezik: Gramatički sistemi i književnojezička norma* (Vol. 1). Naučna knjiga.
- Stump, G. T. (2001). *Inflectional morphology: A theory of paradigm structure* (Vol. 93). Cambridge University Press.
- Vennemann, T. (1972). Rule inversion. *Lingua*, 29(3–4), 209–242.
- Wurzel, W. U. (1990). The Mechanism of Inflection: Lexicon Representations, Rules, and Irregularities. In W. U. Dressler, H. C. Luschützky, O. E. Pfeiffer, & J. R. Rennison (Eds.), *Contemporary morphology* (pp. 203–216). Mouton de Gruyter.
- Zimmermann, E. (2019). Gradient Symbolic Representations and the Typology of Ghost Segments. *Proceedings of the Annual Meetings on Phonology*, 6.
- Zwicky, A. M. (1986). The general case: Basic form versus default form. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 12, 305–314.
- Žygis, M. (2010). Typology of consonantal insertions. *ZAS Papers in Linguistics*, 52, 111–140.

ProQuest Number: 30637200

INFORMATION TO ALL USERS

The quality and completeness of this reproduction is dependent on the quality and completeness of the copy made available to ProQuest.



Distributed by ProQuest LLC (2023).

Copyright of the Dissertation is held by the Author unless otherwise noted.

This work may be used in accordance with the terms of the Creative Commons license or other rights statement, as indicated in the copyright statement or in the metadata associated with this work. Unless otherwise specified in the copyright statement or the metadata, all rights are reserved by the copyright holder.

This work is protected against unauthorized copying under Title 17,
United States Code and other applicable copyright laws.

Microform Edition where available © ProQuest LLC. No reproduction or digitization of the Microform Edition is authorized without permission of ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346 USA