A formal account of morphological epenthesis in Serbo-Croatian

Version of June 4, 2023

Abstract: This article analyzes stem allomorphy in Serbo-Croatian neuter noun inflection as morphological epenthesis. I demonstrate that consonant insertion in the inflection of Serbo-Croatian neuter nouns is a predictable, morphologically conditioned process, rather than an artifact of listed stem allomorphy. Furthermore, the process is not phonologically optimizing, and does not depend on phonological conditions such as vowel hiatus or illicit phonotactic structure. The present analysis includes the process in a wider, algorithmic interpretation of nominal inflection as logical transductions on strings, using Boolean Monadic Recursive Schemes (BMRSs).

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. A logical description is therefore offered of Serbo-Croatian neuter noun inflection, including the processes of stem-final consonant insertion and suffix-initial vowel fronting. The work presented here bears wider implications about the nature of morphophonological processes, and the interfaces of morphology with phonology and syntax.

Keywords: Morphological epenthesis; Inflectional morphology; Morphology-phonology interface; Serbo-Croatian; Recursive program schemes

1 Introduction

The purpose of this article is to show that the stem allomorphs that appear in Serbo-Croatian[1] 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. Existing analyses of Serbo-Croatian nominal inflection do not take this insight into account; furthermore, the formalization of this process, in the form of a system of logical transductions, is a novel contribution of this article.

In this language, 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 institute.NOM.SG.MAS	zavod-a institute.GEN.SG.MAS
		sel-o village.NOM.SG.NEU	sel-a village.GEN.SG.NEU
		bure barrel.NOM.SG.NEU	bure**t**-a barrel.GEN.SG.NEU
	Class II:	ʒen-a woman.NOM.SG.FEM	ʒen-e woman.GEN.SG.FEM
	Class III:	reʧ word.NOM.SG.FEM	reʧ-i word.GEN.SG.FEM

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

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 article 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 article is organized as follows. Section 2 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 2.1-2.3. In subsection 2.4. I outline the approach adopted in this article, which is then given a BMRS analysis in Section 3.

Section 4 focuses on wider assumptions about the architecture of the grammar, touching on the nature of the syntax-morphology interface, discussing an extension to the internal structure of the morphological module, and referring to further research that uses BMRSs to include an inflectional class assignment algorithm into a system of inflectional morphology.

2 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; in this article, I default to the inanimate pattern, where 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. 1995, Klajn 2005).

The inflection of masculine and neuter nominals is affected by the final consonant of the stem: posterior coronal consonants *j, ʎ, ɲ, ʧ, ʤ, ʨ, ʥ, ʃ, ʒ*, and the dental affricate *ʦ*, 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 *ʧekiʨ* '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 *ʧekiʨ* 'hammer' (MAS.) and *poʎe* 'field' (NEU.)

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

[2] These consonants are termed *palatal* in all traditional Serbo-Croatian literature (e.g. Stevanović 1970), 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, *ʦ* patterns with all other affricates, which are C[cor, -ant] (*ʧ, ʤ, ʨ, ʥ*), and triggers vowel fronting in *o*-initial suffixes.
[3] The attested form is *ʧekiʨ-u*, but this is disregarded for the purposes of this article. The vocative singular suffix allomorphy is underresearched, and is not a focus here.

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', *sr̩tse* '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 article.[4]

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

	SG.	PL.
NOM.	bure	bure**t**-a
GEN.	bure**t**-a	bure**t**-aa
DAT.-LOC.	bure**t**-u	bure**t**-ima
ACC.	bure	bure**t**-a
VOC.	bure	bure**t**-a
INS.	bure**t**-om	bure**t**-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 article; the regular neuter plurals are modeled as they apply to the wider group of neuter nouns in Serbo-Croatian.

(2)	ʒen-a	ʒen-e	ʒen-i	etc.
	woman.NOM.SG	woman.GEN.SG	woman.DAT.SG	

(3)	reʧ	reʧ-i	reʧ-i	etc.
	word.NOM.SG	word.GEN.SG	word.DAT.SG	

I claim that neuter stems such as the ones in Table 3 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. In the literature, these 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. 1995; 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).

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

When discussing Serbo-Croatian noun inflectional classes, most linguists and grammarians refer back to Stevanović (1970), 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. (1995) 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 article 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 article.[5]

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

Brozović (2006: 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 Author (under review), and is briefly discussed in Section 4 below.

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, *reʃeto* 'sieve.NOM.SG' (paradigm in Table 1) and *bure* 'barrel.NOM.SG' (paradigm in Table 3) would both have *t*-final stems: *reʃet* 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 **reʃe* (*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 article avoids this issue and treats all neuter stems in a uniform way.

2.3 *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 *poʎe* is *poʎ*- (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-final[6] 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, ʎ, ɲ, ʧ, ʤ, ʨ, ʥ, ʃ, ʒ*, or *ʦ*-final; see Section 2.

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 Author (under review) and Section 4 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 *sirtɕe* 'vinegar', *jagɲe* '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 article, as they are analyzed as having vowel-final stems.

2.4 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 [ɐ] is used at the end of 3rd person verb forms in both numbers (Aronoff & Repetti 2022: 374). In Brazilian Portuguese, [z] repairs vowel hiatus that is a result of adding an affix (Bachrach & Wagner 2007: 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: 75-76).

(4)	/zaova/	[zaova]	'sister-in-law'
	/beograd/	[beograd]	'Belgrade'
	/pirueta/	[pirueta]	'pirouette'

(5)	/violina/	[vi**j**olina]	‘violin’
	/sirius/	[siri**j**us]	‘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)	/sirtɕe + ni/ vinegar + ADJ	[sirtɕe**t**ni] ‘acetic’	vs.	/plod + ni/ fruit + ADJ	[plodni] ‘fertile’
	/tele + ji/ calf + ADJ	[tele**t**ɕi] ‘calf-like’	vs.	/koz + ji/ goat + ADJ	[kozji] ‘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 Ajyíninka 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) Ajyíninka Apurucayali

/i-N-koma-i/	[iŋkoma**t**i]	'he will paddle'
/i-N-koma-aa-i/	[iŋkoma**t**aa**t**i]	'he will paddle again'

(Payne 1981: 108; Lombardi 2002: 239)

(8) Odawa

/ki-akat-i/	[ki**t**akaʧi]	'you are shy'
/ni-ompass/	[ni**t**o:mpass]	'you (pl) oversleep'

(Pigott 1980, in Żygis 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 Analysis

The formalism used here is known as *Boolean Monadic Recursive Schemes* (BMRS); it was introduced by Bhaskar et al. (2020) and Chandlee & 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 article.

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. 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 center embedding, while phonological processes of that kind are unattested. Therefore, 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

[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 a modeling of reduplication with 2-way finite-state transducers, see Dolatian & Heinz (2020).

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 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).

3.2 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 2019): 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, crucial notions like Pāṇini's Principle (= Elsewhere condition) have no obvious finite-state implementation. The 'if... then... else...' structures of BMRSs, on the other hand, directly capture this *do X unless Y*-type behavior, which 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 inflectional rules. Further possibilities for a morphological system that is founded on the notion of hierarchical structures are addressed after the analysis.

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

As argued in Section 2, *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 $\bot$ (True and False), and a finite set of *monadic predicates* $P(t)$ – predicates that take a single argument t, and return $\top$ or $\bot$.

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. $\Sigma_{\bowtie}$ is the union of Σ and all necessary boundary symbols – which, for the purposes of this account, is just the stem boundary symbol +.

(10) $\mathbb{A}$ = {a, b, ʦ, ʧ, ʨ, d, ʤ, ʥ, e, f, g, h, j, k, l, ʎ, m, n, ɲ, o, p, r, s, ʃ, t, u, v, z, ʒ}
$\mathbb{M}$ = {[nom], [gen], [dat], [acc], [voc], [ins], [loc], [sg], [pl], [mas], [neu]}
$\Sigma = \mathbb{A} \cup \mathbb{M}$
$\Sigma_{\bowtie} = \Sigma \cup \{+\}$

In the present analysis, set $\mathbb{A}$ is made up of segments, which is an adequate representation for the points being made in the article. 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 $\Sigma_{\bowtie}$, there is a set $\mathcal{I}$ of input predicates, and a set $\mathcal{O}$ of output predicates 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).

(11) $\mathcal{I} = \{\mathtt{a}_i(x), \ldots, \mathtt{ʒ}_i(x), \mathtt{[nom]}_i(x), \ldots, \mathtt{[pl]}_i(x), +_i(x)\}$
$\mathcal{O}^1 = \{\mathtt{a}_o{}^1(x), \ldots, \mathtt{ʒ}_o{}^1(x), \mathtt{[nom]}_o{}^1(x), \ldots, \mathtt{[pl]}_o{}^1(x), +_o{}^1(x)\}$
$\mathcal{O}^2 = \{\mathtt{a}_o{}^2(x), \ldots, \mathtt{ʒ}_o{}^2(x), \mathtt{[nom]}_o{}^2(x), \ldots, \mathtt{[pl]}_o{}^2(x), +_o{}^2(x)\}$
$\mathcal{O}^3 = \{\mathtt{a}_o{}^3(x), \ldots, \mathtt{ʒ}_o{}^3(x), \mathtt{[nom]}_o{}^3(x), \ldots, \mathtt{[pl]}_o{}^3(x), +_o{}^3(x)\}$

The argument x ranges over domain elements. Strings in $\Sigma_{\bowtie}$ are identified with structures of the form in (12), where the domain D is a finite set of indices, and each character $\sigma \in \Sigma_{\bowtie}$ 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) $\mathbf{S} = \langle D; \sigma_1, \sigma_2, \ldots, \sigma_n, p, s\rangle$
$D = \{1, 2, \ldots, \mathrm{n}\}$

To illustrate how input predicates are evaluated, I use the vowel-final stem of *bure* 'barrel' in Table 4. Some example input predicates are given, 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	 +
$\texttt{[neu]}_{\mathrm{i}}(x)$	⊤	⊤	⊤	⊤	⊥
$\texttt{e}_{\mathrm{i}}(x)$	⊥	⊥	⊥	⊤	⊥
$+_{\mathrm{i}}(x)$	⊥	⊥	⊥	⊥	⊤
$\texttt{r}_{\mathrm{i}}(p(x))$	⊥	⊥	⊥	⊤	⊥
$\texttt{r}_{\mathrm{i}}(s(s(x)))$	⊤	⊥	⊥	⊥	⊥

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 $\texttt{[neu]}_{\mathrm{i}}(x)$ 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 $\texttt{e}_{\mathrm{i}}(x)$ is true for the input position occupied by the stem-final vowel of bure (position 4), and no other position. $+_{\mathrm{i}}(x)$ 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. $\texttt{r}_{\mathrm{i}}(p(x))$ is true for e in position 4, as $\texttt{r}_{\mathrm{i}}(p(x))$ can be understood as "r is the predecessor of x". Similarly, $\texttt{r}_{\mathrm{i}}(s(s(x)))$ 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 Section 4.

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

	1	2	3	4	5	6	7
input	[neu]	[neu]	[neu]	[neu]			
	b	u	r	e	+	[dat]	[sg]
output							
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 `[dat/loc]`$_i(x)$ in (13) to have it return ⊤ for any instance of the dative `[dat]` or locative `[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 `[dat/loc]`$_i(x)$ in (13): if, for a given input position x, the predicate `[dat]`$_i(x)$ evaluates to ⊤, `[dat/loc]`$_i(x)$ will also return ⊤; if the predicate `[dat]`$_i(x)$ returns ⊥, the predicate `[loc]`$_i(x)$ is evaluated, and the boolean value it returns will be the evaluation of `[dat/loc]`$_i(x)$.

(13) `[dat/loc]`$_i(x)$ = if `[dat]`$_i(x)$ then ⊤ else `[loc]`$_i(x)$

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

(14) $\texttt{[nom/acc/voc]}_{i}(x)$ = if $\texttt{[nom]}_{i}(x)$ then $\top$ else

if $\texttt{[acc]}_{i}(x)$ then $\top$ else $\texttt{[voc]}_{i}(x)$

The necessary feature bundles are defined in (15). I start with $\texttt{[dat/loc sg]}_{i}(x)$, which is needed for the realization of the `u` suffix invoked in Table 5. The predicate $\texttt{[dat/loc sg]}_{i}(x)$ evaluates to $\top$ *iff* the user-defined input predicate $\texttt{[dat/loc]}_{i}(x)$ returns $\top$ in the same position, and the position that directly succeeds is associated with the input number feature `[sg]` (the "else $\bot$" 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) $\texttt{[dat/loc sg]}_{i}(x)$ = if $\texttt{[dat/loc]}_{i}(x)$ then $\texttt{[sg]}_{i}(s(x))$

$\texttt{[nom/acc/voc sg]}_{i}(x)$ = if $\texttt{[nom/acc/voc]}_{i}(x)$ then $\texttt{[sg]}_{i}(s(x))$

$\texttt{[ins sg]}_{i}(x)$ = if $\texttt{[ins]}_{i}(x)$ then $\texttt{[sg]}_{i}(s(x))$

$\texttt{[nom/acc/voc pl]}_{i}(x)$ = if $\texttt{[nom/acc/voc]}_{i}(x)$ then $\texttt{[pl]}_{i}(s(x))$

$\texttt{[gen pl]}_{i}(x)$ = if $\texttt{[gen]}_{i}(x)$ then $\texttt{[pl]}_{i}(s(x))$

$\texttt{[dat/loc pl]}_{i}(x)$ = if $\texttt{[dat/loc]}_{i}(x)$ then $\texttt{[pl]}_{i}(s(x))$

$\texttt{[ins pl]}_{i}(x)$ = if $\texttt{[ins]}_{i}(x)$ then $\texttt{[pl]}_{i}(s(x))$

Feature bundles can also be defined so as to capture syncretic patterns at the same time: consider the definition of $\texttt{[dat/loc/ins pl]}_{i}(x)$ 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, $\texttt{[gen/ nom/acc/voc pl]}_{i}(x)$ will evaluate to $\top$ if a

---

[11] Standard logical conjunction could also be used (i.e., $\texttt{[dat/loc sg]}_{i}(x) \wedge \texttt{[sg]}_{i}(s(x))$), 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 [gen] (regardless of number), or the nominative-accusative-vocative plural feature bundle [nom/acc/voc pl].

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

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) is an inflectional realization rule: 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 $\texttt{[dat/loc sg]}_i(x)$, u can only be output faithfully as a counterpart of an input u.

(17) $\texttt{u}_o^1(x)$ = if $\texttt{[dat/loc sg]}_i(x)$ then $\top$ else $\texttt{u}_i(x)$

The definition of $\texttt{a}_o^1(x)$ 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) $\texttt{a}_o^1(x)$ = if $\texttt{[gen/ nom/acc/voc pl]}_i(x)$ then $\top$ else $\texttt{a}_i(x)$
$\texttt{i}_o^1(x)$ = if $\texttt{[ins pl]}_i(x)$ then $\top$ else $\texttt{i}_i(x)$

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 `t`. In order to adequately capture that, in (19) I first define the predicate $\texttt{outseg}(x)$, which returns $\top$ for all instances of segments realized in the first copy of the output string:

(19) $\text{outseg}(x) =$ if $\texttt{a}_\text{o}^1(x)$ then $\top$ else
if $\texttt{b}_\text{o}^1(x)$ then $\top$ else
(...)
if $\texttt{z}_\text{o}^1(x)$ then $\top$ else $\texttt{ʒ}_\text{o}^1(x)$

This user-defined predicate is then referred to in the definition of $\texttt{t}_\text{o}^1(x)$ in (20), which specifies that the stem boundary symbol + will be output as `t` *iff* `e` directly precedes, *and* there is an output segment directly following the stem boundary (as shown in Table 5). Otherwise, `t` will only be output as a faithful counterpart of an input `t`.

(20) $\texttt{t}_\text{o}^1(x) =$ if $+_\text{i}(x)$ then
if $\texttt{e}_\text{i}(p(x))$ then $\text{outseg}(s(x))$
else $\texttt{t}_\text{i}(x)$

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'

input							
	1	2	3	4	5	6	7
	[neu]	[neu]	[neu]	[neu]			
	b	u	r	e	+	[ins]	[pl]
output							
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) $C = \{1, 2, 3\}$

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 of integers used as indices in sets C (21) and D (12). This means that, for any single index i_m, output Copy 1 is ordered first, directly followed by Copies 2 and 3, respectively. These are then followed by the respective output copies of indices i_{m+1}, i_{m+2}, ..., i_n, for all output characters whose functions return $\top$ at each input index.

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 5); otherwise, $\mathtt{m}$ does not appear in Copy 2 of the output. This is captured in the definition of $\mathtt{m_o}^2(x)$ in (22).

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

As per the definition of $\mathtt{a_o}^2(x)$ in (23), $\mathtt{a}$ 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, $\mathtt{a}$ does not surface as a Copy 2 output segment.

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

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 5). This is defined in (24):

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

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 *poʎe* 'field' in Table 8.

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

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

Table 8. Input and output strings of the instrumental plural form of SC *poʎe* 'field'

input						
	1	2	3	4	5	6
	[neu]	[neu]	[neu]			
	p	o	ʎ	+	[ins]	[sg]
output						
Copy 1:	p	o	ʎ		e	
Copy 2:					m	

In order to capture the pattern illustrated in Tables 7 and 8, I utilize the user-defined input predicate $\mathsf{FC}_i(x)$ (25), as well as output Copy 1 predicates $\mathtt{e}_o{}^1(x)$ (26) and $\mathtt{o}_o{}^1(x)$ (27). Let us first look at $\mathsf{FC}_i(x)$: it returns $\top$ for all members of the set of fronting consonants, which was defined in Section 2.[12]

(25) $\mathsf{FC}_i(x)$ = if $\text{ʦ}_i(x)$ then $\top$ else
if $\text{ʧ}_i(x)$ then $\top$ else
if $\text{ʨ}_i(x)$ then $\top$ else
if $\text{ʤ}_i(x)$ then $\top$ else
if $\text{ʥ}_i(x)$ then $\top$ else
if $\mathtt{j}_i(x)$ then $\top$ else
if $\text{ʎ}_i(x)$ then $\top$ else

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

if $\text{ɲ}_i(x)$ then $\top$ else

if $\text{ʃ}_i(x)$ then $\top$ else $\text{ʒ}_i(x)$

Then, in (26), the definition of $\mathtt{e}_o^1(x)$ states that $\mathtt{e}$ 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 `[neu]` (line 2) can have $\mathtt{e}$ 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 $\mathtt{[neu]}_i(p(p(x)))$ in line 3 returns $\bot$, 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 $\mathtt{e}$'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, $\mathtt{e}$ can only surface as a faithful output of an input $\mathtt{e}$ (line 6).

(26) $\mathtt{e}_o^1(x)$ = if $\mathtt{[ins\ sg]}_i(x)$ then $\mathrm{FC}_i(p(p(x)))$ else

if $\mathtt{[neu]}_i(p(p(x)))$ then

if $\mathtt{[nom/acc/voc\ sg]}_i(x)$ then $\mathrm{FC}_i(p(p(x)))$

else

if $\mathtt{[voc\ sg]}_i(x)$ then $\top$ else

if $\mathtt{[acc\ pl]}_i(x)$ then $\top$ else $\mathtt{e}_i(x)$

In (27), the definition of $\mathtt{o}_o(x)$ states that $\mathtt{o}$ is the exponent of the instrumental singular (line 1), provided that the conditions that would produce an $\mathtt{e}$ in the same position in output Copy 1 do not hold (line 2). This makes $\mathtt{o}$ the default exponent of the instrumental singular in Class I –

it gets output in the absence of more specific conditions. o 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 $\mathtt{o}_o{}^1(x)$ (the hierarchical order of predicates) is parallel to that of $\mathtt{e}_o{}^1(x)$: this captures the generalization that, in this inflectional class, suffix-initial default o is overwritten by e in a specific context (after FC-final stems).[13]

(27) $\mathtt{o}_o{}^1(x)$ = if $\mathtt{[ins\ sg]}_i(x)$ then
if $\mathtt{e}_o{}^1(x)$ then $\perp$ else $\top$
else
if $\mathtt{[neu]}_i(p(p(x)))$ then
if $\mathtt{[nom/acc/voc\ sg]}_i(x)$ then
if $\mathtt{e}_i(p(p(x)))$ then $\perp$ else
if $\mathtt{e}_o{}^1(x)$ then $\perp$ else $\top$
else $\mathtt{o}_i(x)$

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.

(28) $\mathtt{r}_o{}^1(x) = \mathtt{r}_i(x)$
$\mathtt{m}_o{}^1(x) = \mathtt{m}_i(x)$
$\mathtt{b}_o{}^1(x) = \mathtt{b}_i(x)$
$\mathtt{ʦ}_o{}^1(x) = \mathtt{ʦ}_i(x)$

[13] In an alternative approach, the initial vowel of the [ins sg] and the [nom/acc/voc sg] suffixes is always o, which then gets fronted to e after FCs. I develop that account in Section 4, where I introduce additional levels to the architecture of the morphological module.

$ʧ_o{}^1(x) = ʧ_i(x)$
$ʨ_o{}^1(x) = ʨ_i(x)$
(etc., for {d, ʤ, ʥ, f, g, h, j, k, l, ʎ, n, ɲ, p, s, ʃ, v, z, ʒ})

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 t-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 ⋉ is similarly output as the desired consonant); in fact, it specifically captures the understanding that this consonant 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.

4 Architecture of the grammar: basic assumptions and implications

The work outlined in Section 3 is part of a larger research project, which conceptualizes morphology as an independent module of transductions on strings that lies between syntax and (pure) phonology. Following Ermolaeva & Edmiston (2018), I assume that the flattening of syntactic tree structure happens above the morphological module, not post-morphology as assumed in frameworks like Distributed Morphology (DM). As already discussed in Section 3.1, the strong generative capacity of a system operating over binary trees is above regular (i.e. at least context-free on the Chomsky hierarchy), and predicts patterns which are unattested in morphology (e.g. iterable nested dependencies).

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. All examples in Section 3 assume the input string to have a "stem + case + number" structure, which is what these components get rearranged to pre-flattening in Serbo-Croatian. 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 9, which is an alternative version of what was shown in Table 5:

Table 9. Input (alternative) and output strings of the dative singular form of SC *bure* 'barrel'

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

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 paper, 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: 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 Section 3, by mirroring Stump's PFM model more closely. Incorporating PFM's assumption that the realization of a word form occurs in a step-wise fashion (stem selection – exponent realization – morphophonological processes; Stump 2001: 33-38) makes it possible to straightforwardly model phenomena like contextual stem allomorphy, affix allomorphy that is based on the form of the stem, as well as phonological processes that are dependent on morphological information.

Assuming a step in the architecture in which elements like stems or morphosyntactic features have not been spelled out yet allows us to easily model variable morphotactics. Let us take (a slightly simplified slice of) Georgian verb inflection as an example, where verbal stems can take a single affix in the present tense: first person singular is marked with the prefix *v-*, whereas third person singular is marked with the suffix *-s* (Harris & Samuel 2021). If we assume that the input string to the Georgian morphological module has a structure such as “number + person + stem” (with the stem boundary symbol present on both sides of the stem), the stem selection step for *ban* ‘bathe’ will look like the illustrations in Tables 10 and 11. `[1]` and `[3]` represent morphosyntactic person features (first and third, respectively), while $\texttt{Stem}_{\texttt{BATHE}}$ represents the present indicative stem for the Georgian verb ‘bathe’, which gets spelled out as `ban` (Harris & Samuel 2021: 164).

Table 10. Stem selection step for the first person singular form of Georgian *ban* ‘bathe’

	1	2	3	4	5
<u>input</u>	[1]	[sg]	+	$\texttt{Stem}_{\texttt{BATHE}}$	+
<u>output</u>					
Copy 1:		[1sg]	+	b	+
Copy 2:				a	
Copy 3:				n	

Table 11. Stem selection step for the third person singular form of Georgian *ban* ‘bathe’

	1	2	3	4	5
<u>input</u>	[3]	[sg]	+	$\texttt{Stem}_{\texttt{BATHE}}$	+
<u>output</u>					
Copy 1:			+	b	+
Copy 2:				a	[3sg]
Copy 3:				n	

Given that the stem node is exactly one symbol long ($\texttt{Stem}_{\texttt{BATHE}}$) in the input string (i.e. before spellout) for any lexical entry, defining morphosyntactic feature bundles that are to come right before or right after it is relatively straightforward, as it requires constant, finite counting. The predicates for first and third person singular feature bundles in Georgian are defined in (29) and (30), respectively. As multiple steps have now been introduced to the process of word-form realization, from this point on I mark stem selection output predicates with a subscript S, for clarity.

(29) $\mathtt{[1sg]}_{\mathrm{So}}{}^{1}(x) = \text{if } \mathtt{[sg]}_{\mathrm{i}}(x) \text{ then } \mathtt{[1]}_{\mathrm{i}}(p(x))$

(30) $\mathtt{[3sg]}_{\mathrm{So}}{}^{2}(x) = \text{if } +_{\mathrm{i}}(x) \text{ then}$

$\text{if } \mathtt{[sg]}_{\mathrm{i}}(p(p(x))) \text{ then } \mathtt{[3]}_{\mathrm{i}}(p(p(p(x))))$

The segments that comprise the realization of $\mathtt{Stem}_{\mathtt{BATHE}}$ are defined in (31), (32), and (33). The stem boundary symbol + also needs to be retained for the following step (exponent realization); this is reflected in the definition of $+_{\mathrm{So}}{}^{1}(x)$ in (34).

(31) $\mathtt{b}_{\mathrm{So}}{}^{1}(x) = \mathtt{Stem}_{\mathtt{BATHE}\ \mathrm{i}}(x)$

(32) $\mathtt{a}_{\mathrm{So}}{}^{2}(x) = \mathtt{Stem}_{\mathtt{BATHE}\ \mathrm{i}}(x)$

(33) $\mathtt{n}_{\mathrm{So}}{}^{3}(x) = \mathtt{Stem}_{\mathtt{BATHE}\ \mathrm{i}}(x)$

(34) $+_{\mathrm{So}}{}^{1}(x) = +_{\mathrm{i}}(x)$

The outputs defined in (29–34) and illustrated in Tables 10 and 11, `[1sg]+ban+` and `+ban+[3sg]`, can therefore be realized as `v+ban+` and `+ban+s`, respectively, in the exponent realization step that follows.

Going back to Serbo-Croatian, consider the process of suffix-initial vowel fronting in that language, illustrated in Tables 7 and 8, and modeled by the predicates $\mathtt{e}_{\mathrm{o}}{}^{1}(x)$ and $\mathtt{o}_{\mathrm{o}}{}^{1}(x)$ in (26) and (27). 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 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 12 and 13:

Table 12. Exponent realization step for the instrumental singular form of SC *poʎe* 'field'

input					
	1	2	3	4	5
	[neu] p	[neu] o	[neu] ʎ	 +	 [ins sg]
output					
Copy 1:	[neu] p	[neu] o	[neu] ʎ	 +	[ins sg] o
Copy 2:					[ins sg] m

Table 13. Morphophonological processes step for the instrumental singular form of SC *poʎe* 'field'

input						
	1	2	3	4	5	6
	[neu] p	[neu] o	[neu] ʎ	 +	[ins sg] o	[ins sg] m
output						
Copy 1:	p	o	ʎ		e	m

The process is thus altered in several ways compared to what was outlined in Section 3. As shown in Table 10, in the exponent realization step, o surfaces consistently as the Copy 1 output segment if the corresponding input position is occupied by [ins sg], or [nom/acc/voc sg] following a neuter stem. The predicate $\mathtt{o}_{\mathrm{Eo}}{}^{1}(x)$ in (35) is therefore simplified in comparison to $\mathtt{o}_{\mathrm{o}}{}^{1}(x)$ in (27); similarly, $\mathtt{e}_{\mathrm{Eo}}{}^{1}(x)$ in (36) is simplified in comparison to $\mathtt{e}_{\mathrm{o}}{}^{1}(x)$ in (26), as e can now only be realized as the exponent of the [voc sg] or the [acc pl] feature bundle. For clarity, the exponent realization output predicates are henceforth marked with a subscript E.

(35) $\mathtt{o}_{\mathrm{Eo}}{}^{1}(x)$ = if $\mathtt{[ins\ sg]}_{\mathrm{i}}(x)$ then $\top$ else
if $\mathtt{[neu]}_{\mathrm{i}}(p(p(x)))$ then
if $\mathtt{[nom/acc/voc\ sg]}_{\mathrm{i}}(x)$ then
if $\mathtt{e}_{\mathrm{i}}(p(p(x)))$ then $\perp$ else $\top$
else $\mathtt{o}_{\mathrm{i}}(x)$

(36) $\mathtt{e}_{\mathrm{Eo}}{}^{1}(x)$ = if $\mathtt{[voc\ sg]}_{\mathrm{i}}(x)$ then $\top$ else
if $\mathtt{[acc\ pl]}_{\mathrm{i}}(x)$ then $\top$ else $\mathtt{e}_{\mathrm{i}}(x)$

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 (37) will suffice:

(37) $+_{\mathrm{Eo}}{}^{1}(x) = +_{\mathrm{i}}(x)$
$\mathtt{[neu]}_{\mathrm{Eo}}{}^{1}(x) = \mathtt{[neu]}_{\mathrm{i}}(x)$
$\mathtt{[ins\ sg]}_{\mathrm{Eo}}{}^{1}(x) = \mathtt{[ins\ sg]}_{\mathrm{i}}(x)$
$\mathtt{[ins\ sg]}_{\mathrm{Eo}}{}^{2}(x) = \mathtt{[ins\ sg]}_{\mathrm{i}}(x)$
(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 $\mathtt{T}$ for a position x in the same output copy (e.g., both $\mathtt{o}_{\mathsf{Eo}}{}^1(x)$ and $\mathtt{[ins\ sg]}_{\mathsf{Eo}}{}^1(x)$ evaluate to $\mathtt{T}$ for position 5 in Table 12). Such transductions by themselves do not fall under the definition of well-defined transductions (Bhaskar et al. 2020: 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 $\mathtt{T}$ 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 11, I define $\mathtt{e}_{\mathsf{Mo}}(x)$ in (38): the predicate targets input $\mathtt{o}$'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, $\mathtt{e}$'s are only output faithfully (line 3). No special conditions are imposed on $\mathtt{o}_{\mathsf{Mo}}{}^1(x)$ other than the presence of an $\mathtt{o}$ in the input position (39), and the definition of $\mathsf{FC}_{\mathsf{i}}(x)$ is the one given in (25).

(38) $\mathtt{e}_{\mathsf{Mo}}{}^1(x)$ = if $\mathtt{o}_{\mathsf{i}}(x)$ then
if $+_{\mathsf{i}}(p(x))$ then $\mathsf{FC}_{\mathsf{i}}(p(p(x)))$
else $\mathtt{e}_{\mathsf{i}}(x)$

(39) $\mathtt{o}_{\mathsf{Mo}}{}^1(x) = \mathtt{o}_{\mathsf{i}}(x)$

Predicates like the one in (39) 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, -∅ case endings). Any other morphophonological process would be defined via predicates similar to the ones in (38) and (39).

As can be seen from the analysis offered in this article, 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 2019), and algorithms are crucial to linguistic theory; further work on the inflectional system of Serbo-Croatian focuses precisely on the algorithms that play a major role in this system. Specifically, Author (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. Consider the (slightly simplified) paradigms of inflectional suffixes for nouns in Table 14:

Table 14. Inflectional classes of SC nouns

	CLASS I						CLASS II		CLASS III	
	masculine *zavod-* ‘institute’		neuter (+ masculine) C-final *sel-* ‘village’		neuter (+ masculine) *e*-final *bure-* ‘calf’		feminine (+ masculine) *ʒen-* ‘woman’		feminine *reʧ-* ‘word’	
	SG.	PL.	SG.	PL.	SG.	PL.	SG.	PL.	SG.	PL.
NOM.	∅	i	o	a	∅	**t**a	a	e	∅	i
GEN.	a	aa	a	aa	**t**a	**t**aa	e	aa	i	ii
DAT.-LOC.	u	ima	u	ima	**t**u	**t**ima	i	ama	i	ima
ACC.	∅	e	o	a	∅	**t**a	u	e	∅	i
VOC.	e	i	o	a	∅	**t**a	o	e	i	i
INS.	om	ima	om	ima	**t**om	**t**ima	om	ama	i	ima

It is apparent that, without lexically listed class features, the system cannot rely solely on the form of the stem and the morphosyntactic property of gender of a given lexeme to correctly produce the desired inflected output forms. However, the approach adopted in Author (under review) takes into account a *leading form* or *base* (Wurzel 1990; McCarthy 2005; Albright 2008) for every lexical entry. The article argues that Serbo-Croatian is sensitive to the nominative singular as the leading form; in combination with the lexically listed gender and stem properties, it provides sufficient information for correct declension placement. With the 'if... then... else...' syntax of BMRS, such a class assignment algorithm is easy to embed in a system of nominal inflection; such a system, then, makes predictions as to what it considers the default inflectional pattern, and how new words (loanwords, nonce words) get assigned to an inflectional class. This is therefore a clear path for a further implementation of BMRSs in an analysis of the complete inflectional system of Serbo-Croatian. The formalism allows for a direct representation of morphological substance and generalizations, while 'if... then... else...' statements necessarily impose a hierarchical structure of predicates.

5 Conclusion

This article provides an analysis of morphologically conditioned consonant insertion in Serbo-Croatian; as shown in Section 2, 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.

Furthermore, defining morphological processes as string transductions allows for an approach that makes use of concepts from different frameworks like Distributed Morphology and Paradigm Function Morphology. 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. There is little research on BMRSs, which are new, and open questions remain; however, this formalism is promising, as it is not only rather intuitive, but also easily implementable, and extendable to a wider range of phenomena.

References

Albright, A. (2008). Inflectional Paradigms Have Bases Too. Arguments from Yiddish. In A. Bachrach & A. Nevins (Eds.), *The Bases of Inflectional Identity* (pp. 271–312). Oxford University Press.

Anderson, S. R. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.

Aronoff, M. (1976). *Word Formation in Generative Grammar.* Cambridge, MA: MIT Press.

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.

Author, O. (under review). TBD. In *TBD*.

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. (1995). *Hrvatska gramatika* [*Croatian grammar*]. Zagreb: Školska knjiga.

Bhaskar, S., Chandlee, J., Jardine, A., & Oakden, C. (2020). Boolean monadic recursive schemes as a logical characterization of the subsequential functions. In *International Conference on Language and Automata Theory and Applications* (pp. 157-169). Springer, Cham.

Blake, B. J. (2001). *Case*. Cambridge: 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.

Broselow, E. (1984). Default consonants in Amharic morphology. *MITWPL, 7*, 15-31.

Brozović, D. (2006). *Neka bitna pitanja hrvatskoga jezičnog standarda* [*Some important aspects of the Croatian language standard*]. Zagreb: Školska knjiga.

Chandlee, J. (2017). Computational locality in morphological maps. *Morphology*, *27*, 599-641.

Chandlee, J., & Jardine, A. (2021). Computational universals in linguistic theory: Using recursive programs for phonological analysis. *Language*, *97*(3), 485-519.

Dolatian, H., & Heinz, J. (2020). Computing and classifying reduplication with 2-way finite-state transducers. *Journal of Language Modelling*, *8*(1), 179-250.

Filiot, E. (2015). Logic-automata connections for transformations. *Indian Conference on Logic and Its Applications*, 30–57.

Harris, A. C., & Samuel, A. G. (2021). The suffixing preference: A preliminary report on processing affixes in Georgian. In S. Moradi, M. Haag, J. Rees-Miller, & A. Petrovic (Eds.), *All Things Morphology: Its independence and its interfaces* (pp. 147–168). John Benjamins.

Heinz, J. (2018). The computational nature of phonological generalizations. In *Phonological Typology, Phonetics and Phonology* (pp. 126–195). De Gruyter Mouton.

Johnson, C. D. (1972). *Formal aspects of phonological description*. Mouton.

Kaplan, R. & Kay, M. (1994). Regular models of phonological rule systems. *Computational Linguistics*, 20:331–78.

Karttunen, L. (2003). Computing with realizational morphology. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 203-214). Springer, Berlin, Heidelberg.

Karttunen, L., Kaplan, R. M., & Zaenen, A. (1992). Two-level morphology with composition. In *COLING 1992 Volume 1: The 15th International Conference on Computational Linguistics.*

Kiparsky, P. (1973). "Elsewhere" in phonology. In Stephen R. Anderson and Paul Kiparsky (eds.), *A Festschrift for Morris Halle,* 93-106. New York: Holt, Rinehart, and Winston.

Klajn, I. (2005). *Gramatika srpskog jezika* [*Grammar of the Serbian language*]. Belgrade: Zavod za udžbenike i nastavna sredstva.

Leslau, W. (1997). *Ethiopic Documents: Argobba*. Wiesbaden, Harrassowitz Verslag.

Lombardi, L. (2002). Coronal epenthesis and markedness. *Phonology* 19(2), 219–251.

Marković, I. (2013). *Hrvatska morfonologija* [*Croatian Morphophonology*]. Zagreb: Disput.

Matasović, R. (2008). *Poredbenopovijesna gramatika hrvatskoga jezika* [*Comparative Historical Grammar of the Croatian Language*]. Zagreb: Matica hrvatska.

McCarthy, J. (2005). Optimal Paradigms. In: L. Downing, T. Hall & R. Raffelsiefen (Eds.), *Paradigms in Phonological Theory*. Oxford: Oxford University Press, pp. 170–210.

Moradi, S. (2017). *Non-canonical Epenthesis: Epenthetic quality and the role of morphonology*. Unpublished manuscript, Department of Linguistics, Stony Brook University. https://www.academia.edu/100511567/Non_canonical_Epenthesis_Epenthetic_quality_and_the_role_of_morphonology

Moschovakis, Y. N. (2019). *Abstract recursion and intrinsic complexity*, volume 48 of *Lecture Notes in Logic*. Cambridge University Press.

Pigott, G. (1980). *Aspects of Odawa Morphophonemics*. New York: Garland.

Šljivić-Šimšić, B. (1984). Neuter nouns in -∅ or neuter nouns in -*e* with extended stems in Standard Serbo-Croatian. *Folia Slavica, 6*(3), 372-388.

Staroverov, P. (2014). *Splitting theory and consonant epenthesis* [Rutgers, The State University of New Jersey]. https://rucore.libraries.rutgers.edu/rutgers-lib/45498/

Stevanović, M. (1970). *Savremeni srpskohrvatski jezik I* [*Contemporary Serbo-Croatian language I*]. Belgrade: Naučna knjiga.

Strother-Garcia, K. (2019). *Using model theory in phonology: a novel characterization of syllable structure and syllabification*. PhD thesis, University of Delaware.

Stump, G. T. (2001). *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press.

Vennemann, T. (1972). Rule inversion. *Lingua, 29*(3-4), 209-42.

Wurzel, W. U. (1990). The Mechanism of Inflection: Lexical Representation, Rules, and Irregularities. In: W. U. Dressler (Ed.), *Contemporary Morphology*. Berlin: Mouton de Gruyter.

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.

Żygis, M. (2010). Typology of Consonantal Insertions. *ZAS Papers in Linguistics* 52, 111-140.