A Sublexicon Approach to Morphological Dependencies

Guy Tabachnick(guyt@nyu.edu)

Department of Linguistics, New York University

10 Washington Place, New York, NY 10003, USA

Abstract

Ackerman et al. (2009) and Ackerman and Malouf (2013) argue that the organization of morphological systems allows speakers to efficiently solve the Paradigm Cell Filling Problem: how to predict a word's inflected form given some of its other forms. For example, a given word's possessive suffix in Hungarian can be predicted in part by its phonology and by its exponent in forms like the plural. I propose a model in which this phonological and morphological predictability is encoded in formal, constraint-based grammars. In particular, I extend the sublexicon model (Gouskova et al., 2015), in which lexically specific behavior is handled using diacritics on lexical items. Thus, I treat the Paradigm Cell Filling Problem as a problem of finding correlations between diacritics in a given lexical entry, without relying on the storage of a paradigm's output forms.

Keywords: morphology, allomorphy, Paradigm Cell Filling Problem, analogical modelling, sublexicons, Hungarian

1 Introduction

Comparisons of different morphological theories often focus on how a particular theory's architectural features allow for an adult speaker's grammar to generate a particular set of forms. For example, Kramer (2016) looks at how two theories, Paradigm Function Morphology (Stump, 2019) and Distibuted Morphology (Halle and Marantz, 1993), account for *syncretisms*—that is, instances where two words with different morphosyntactic features have the same form. However, there is another aspect to speakers' knowledge of morphology: Ackerman et al. (2009) and Ackerman and Malouf (2013) argue that learners create inferential networks among members of a morphological paradigm that allow for the accurate prediction of an unseen inflected form of a word given knowledge of other of its inflected forms (what they call the Paradigm Cell Filling Problem). The formal architecture of theories like Distributed Morphology cannot provide an account of these inferences, which do not involve the deterministic generation or selection of output forms from abstract morphosyntactic structures. Nonetheless, the process of associating lexical items with their arbitrary inflectional classes is one crucial component of morphological learning, and the inferences that develop during this process of learning can be used to generate previously unseen forms of familiar words. Accordingly, we want to capture this inferential knowledge in our theories of linguistic competence.

One such case where speakers must learn to associate individual lexical items with the appropriate allomorph of an inflectional affix is the possessive marker in Hungarian. The possessive has two basic allomorphs that differ in the presence of [j]: *-ɒ/ɛ* and *-jɒ/jɛ* (henceforth *-V* and *-jV*, respectively). As Rácz and Rebrus (2012) describe, in some pockets of the lexicon, a noun's phonological features make it easy to associate each noun with its correct possessive allomorph: like the examples in (1a), nouns ending in sibilants always take *-V*, while nouns ending in vowels universally have *-jV*. In other cases, a noun's phonology is not fully determinative: even (near-)minimal pairs can differ in their possessive, as we see in (1b). A noun's morphology can also provide clues to its possessive: nouns in irregular stem classes (like those that have [ɒ] in the plural instead of regular [o]) mostly have possessives with *-V*. This is the case with the forms in (1c).

(1) *Phonological and morphological differences in Hungarian possessive allomorphy*

a. phonological predictability: gaːz-ɒ 'gas', kɒpu-**j**ɒ 'gate'

b. lexical idiosyncrasy: kaːr-ɒ 'damage', paːr-**j**ɒ 'pair'

c. morphological predictability: ɟaːr-ɒ 'factory', paːr-**j**ɒ 'pair' (cf. plurals ɟaːr-ɒ̲k, paːr-o̲k)

Thus, the Hungarian possessive is a case of *partial predictability*. Speakers can use cues in a given noun's phonology and morphology to predict the possessive allomorph it selects. In most cases, however, they cannot be sure until they actually hear the form, and ultimately must explicitly learn the selectional pattern for all, or at least most, nouns.

In this paper, I propose a model of morphological learning that addresses the empirical problem of paradigm cell filling—namely that learners efficiently use knowledge of inflected forms of a word to infer other inflected forms of that word—and provides a concrete implementation using a highly elaborated lexicon. In particular, I extend the *sublexicon* model (Gouskova et al., 2015), which makes use of diacritics attached to individual lexical items to govern lexically specific behaviors like Hungarian possessive allomorphy. The model assumes that these diacritics appear in contextually dependent operations (such as rules of exponence) that comprise part of the morphological or morphosyntactic grammatical module. In the case of the possessive, nouns that select for *-V* would receive one diacritic, while those that select for *-jV* would receive another. In the sublexicon model, the learner divides the lexicon into sublexicons based on these diacritics, each of which is described by sublexical grammars. Learners then decide to place new words into sublexicons based on how they are evaluated by the various sublexical grammars. Thus, each sublexicon is linked to a specific behavior (through the diacritic in the rule of exponence) and allows the speaker to learn generalizations (through the sublexical grammars).

In previous work, these sublexical grammars contained phonotactic constraints, allowing them to capture generalizations over the *phonological* form of the members of each sublexicon. To account for *morphological* inferences, I extend these grammars to include a set of *SUBLEX constraints, which penalize forms for preexisting membership in a particular sublexicon. Thus, in determining an inflected form of a novel word, learners can be sensitive to its other inflected forms in addition to its phonological shape.

The *SUBLEX constraints treat the Paradigm Cell Filling Problem as a problem of correlation between diacritics. In Hungarian, the Paradigm Cell Filling Problem requires establishing an implicational relation between two cells of the paradigm, the plural and the possessive. Rather than directly comparing output forms, the *SUBLEX constraints implement this by establishing a correlation between two properties that can both be associated with a given morpheme, like [ɟaːr] 'factory': namely, "has plural with *-ɒk*" and "doesn't have possessive with *-jɒ*". These properties are stored in the form of diacritics on items in the lexicon. In this sense, the sublexicon theory is a theory of lexicon structure, rather than a theory of how the grammar generates forms. Accordingly, it is largely independent of questions involving morphological architecture: it is compatible with any theory of morphology that allows for reference to morphosyntactically arbitrary features attached to items in the lexicon.

In Section 2, I present the basic sublexicon model and propose an extension to it in the form of morphological *SUBLEX constraints. Next, in Section 3, I present the case study of Hungarian possessives in more detail, then model this case with sublexicons in Section 4. Section 5 discusses my proposal in the context of the Paradigm Cell Filling Problem research program, and Section 6 concludes.

2 Sublexicon models and morphological dependencies

2.1 The basic sublexicon model

The sublexicon model (Gouskova et al., 2015) has been proposed as a means of grammatically encoding phonological generalizations in cases of lexically specific variation—for example, words that select the same allomorph of a given affix. This allows learners to pick up on the partial phonological predictability determining a given lexical item's choice of allomorph. As such, it follows in the path of previous formal models, like the Minimal Generalization Learner (Albright and Hayes, 2003), that aim to use phonological analogy to determine the set of lexical items to which a given morphophonological rule applies. (See Guzmán Naranjo (2019) for an overview of proposals—in both the generative tradition and others, such as usage-based linguistics—using analogical modelling to predict morphological class affiliation.)

For example, Gouskova et al. (2015) discuss the case of Russian diminutives: masculine nouns form the diminutive with one of three suffixes (*-ók*, *-ʲik*, and *-tʃʲik*), as we see in Table 1:

word	*gloss*	*diminutive*
áng^{j}il	'angel'	áng^{j}il-ók
mónstr	'monster'	mónstr-^{j}ik
ajfón	'iPhone'	ajfón-tʃ^{j}ik

Table 1: Some Russian diminutive forms (see Gouskova et al., 2015)

In the sublexicon model, the learner divides the lexicon into *sublexicons* that pattern together. For example, in Russian, as learners encounter words in diminutive forms, they sort them into sublexicons according to those words' diminutive suffixes:

(2) *Russian diminutive sublexicons*

a. [*ok*]: {áng^{j}il, ...}

b. [*ik*]: {mónstr, ...}

c. [*chik*]: {ajfón, ...}

The sublexicon theory of morphological learning makes several assumptions about the architecture of the grammar. First of all, I assume a *realizational* theory of morphology (Stump, 2001), in which the grammar produces exponents for abstract morphosyntactic properties such as diminutive. Next, these morphosyntactic properties must get spelled out into phonological material through *rules of exponence*—for example, one rule of exponence would dictate that the diminutive is realized as *-ók*. Finally, I assume that these rules of exponence can be *contextually specific*, meaning that they only apply to certain lexical items marked accordingly in the lexicon. For example, one such contextually specific rule would have the diminutive be realized as *-ók* in the context of roots like [áng^{j}il].

This set of assumptions is compatible with theories like Distributed Morphology (Halle and Marantz, 1993) and Paradigm Function Morphology (Stump, 2019), but not, for example, Nanosyntax (Baunaz and Lander, 2018), which handles lexically specific behavior through differences in syntactic structure rather than formal features. Other architectural choices I make in this paper, such as the use of Distributed Morphology–style rules of exponence or my particular (at times oversimplified) analysis, are for the purposes of illustration but are not essential to the learning model I present.

In the case of Russian diminutives, we have three basic rules of exponence corresponding to the three diminutive allomorphs:

(3) *Rules of exponence with sublexical diacritics for the Russian diminutive*

a. DIM ↔ ók / [*ok*] ___

b. DIM ↔ ʲik / [*ik*] ___

c. DIM ↔ tʃʲik / [*chik*] ___

When Russian speakers wish to form the diminutive of a new lexical item, they must determine which of the rules of exponence in (3) applies to this particular item. To do so, they must figure out whether this word gets an [*ok*], [*ik*], or [*chik*] diacritic—that is, which sublexicon in (2) it belongs to. For previously unseen words, there is no lexically listed diminutive form, so they must place the word into a sublexicon on the basis of its phonology.[1] They do so by using a *sublexical grammar*, described in the next section.

2.2 Sublexical phonotactic grammars

Hayes and Wilson (2008) present a model of phonotactic learning in which a constraint-based phonotactic grammar is induced over the lexicon to capture generalizations over a language's forms. In this model, the learner keeps track of sequences of sounds that are rare or absent in the lexicon and proposes constraints against them. For example, Russian has voicing assimilation, so voiced obstruents never appear before voiceless consonants. Thus, words with voiced obstruents in this position will be unattested, so the learner will generate a phonotactic constraint, *[−son,+voice][−voice], which penalizes voiced obstruents before voiceless consonants. The generated constraints allow speakers to get a sense of what a typical word in the lexicon looks like. The more out of place a segment or sequence is in the lexicon, the more heavily weighted the constraint against it. Since Russian never has mixed-voice obstruent clusters, even though voiced obstruents are quite common in the language as a whole, the constraints listed above will be heavily weighted.

[1] Of course, previously established forms can exhibit stable variation as well. Gouskova et al. (2015) posit that in such cases, a lexical item will belong to multiple sublexicons and its output will be assigned probabilistically in any given instance. I do not address such cases further.

The sublexicon model extends the notion of phonotactic learning to capture generalizations over subsets of the lexicon that pattern together—that is, sublexicons. The learner induces a phonotactic grammar not just for the entire lexicon, but also for each sublexicon, allowing the learner to find patterns specific to that sublexicon. For example, Gouskova et al. (2015) note that nouns that take *-ʲik* tend to have word-final stress and do not end in velar consonants. The sublexical grammar for the [*ik*] sublexicon, then, should include constraints penalizing these structures alongside those underattested in the language as a whole, such as mixed-voice clusters. Indeed, we see in Table 2 that the Phonotactic Learner did find constraints corresponding to these claims in the literature (in addition to many other, more specific constraints). The stress constraint (*[−stress][−stress]#), which operates on a projected stress tier only including stress-marked vowels, is somewhat more specific, penalizing words whose last two syllables are unstressed. The dorsal effect is captured with a constraint against word-final segments marked [−low]. Among Russian consonants, this set defines the dorsals: velars are marked [−low], while uvulars (which Russian does not have) are [+low]; other consonants are undefined for the [±low] feature. Finally, the learner also finds a strong constraint against mixed-voice clusters, *[−son,+voice][−voice], which applies throughout the Russian lexicon, not just for the [*ik*] sublexicon.

constraint		*description*	*weight*
*[−stress][−stress]#	(Stress tier)	no final stress lapses	5.192
*[−son,+voice][−voice]		voice agreement	4.567
*[−low]#		no word-final dorsals	2.918

Table 2: Selected constraints from the [*ik*] sublexical grammar induced by the UCLA Phonotactic Learner (Hayes and Wilson, 2008), from Gouskova et al. (2015)

As new words are learned and added to sublexicons, the sublexical grammars are updated to account for the updated generalizations over the sublexicons' members—this can include recalibrating the weights of individual constraints or inducing new ones.

These sublexical grammars are then reflected in learners' behavior. When the learner wishes to form the diminutive of a novel word, they evaluate the stem against each sublexicon's grammar, where each sublexical grammar yields a score for that word. The better a word fares on the [*ik*] sublexicon relative to the other two, the more likely it is to be placed into this sublexicon. Once a word is placed into a sublexicon, the appropriate rule of exponence in (3) can be applied to form the word's diminutive.

In Figure 1, we can see an example of nonce words being tested on the [*ik*] sublexical grammar. Since [tʃʲadúx] ends in a velar, it violates *[−low]#, incurring a penalty of 2.918. On the other hand, since one of its last two syllables is stressed, it does not violate *[−stress][−stress]#. The opposite is true for [pákuval], whose antepenultimate stress means it incurs a costly violation of the first constraint, worth 5.192.

constraint *weight*	*[−stress][−stress]# 5.192	*[−son,+voice][−voice] 4.567	*[−low]# 2.918	...	total
tʃʲadúx	0	0	−2.918	...	−2.918
pákuval	−5.192	0	0	...	−5.192

Figure 1: Nonce word evaluation on the [*ik*] sublexical grammar

Since [tʃʲadúx] has a better score on the [*ik*] sublexical grammar than [pákuval], it is a better fit for this sublexicon. Of course, whether or not either of these forms would actually be placed into the [*ik*] sublexicon, and thus have an [*ik*] diacritic on their lexical items and a diminutive form with -ʲ*ik*, depends on how they are evaluated on the [*ok*] and [*chik*] sublexical grammars as well: one of these latter sublexicons may turn out to be a substantially better fit for either or both of these nonce words.

The sublexicon model with phonotactic sublexical grammars is designed to capture generalizations over the phonological shape of each sublexicon's members. This raises the question of whether these grammars have access to any information about a word other than its phonological form. In particular, the phonology seems to be able to identify certain lexical items as undergoing item-specific morphophonological patterns—for example, some Russian nouns shift their stress onto the suffix in some portion of teh paradigm. This information is most easily contained in the lexicon, in the form of markings on lexical items. But these are precisely the diacritics that can appear in morphological contexts like selectional restrictions for the Russian diminutive. Accordingly, it is reasonable to assume that the sublexical grammars also have access to lexical diacritics. I take advantage of this by proposing a family of constraints, *SUBLEX, which reference a given lexical item's *morphological* organization, allowing learners to consider the set of diacritics already placed on a lexical item in determining how to assign the word a new diacritic and place it into a new sublexicon.

2.3 A sublexicon model with morphology

2.3.1 Multiple sublexicons

Looking at it another way, when we add *SUBLEX constraints, each diacritic now does double duty. On the one hand, it serves as the foundation for a sublexicon: for example, in Russian, the [*ok*] sublexicon contains all words with the [*ok*] diacritic. On the other hand, each diacritic is also referenced in a *SUBLEX constraint in the sublexical grammars of *other* sublexicons. To understand this, we need to take a broader look at morphological systems than we have so far: we will now expand our example of Russian nouns from just the diminutives to encompass the plural as well.

The Russian plural for masculine nouns, just like the diminutive, has more than one allomorph, although the distribution of allomorphs is quite different: the majority of nouns, like [naród] 'people', take -*ɨ* in the plural, while a small number, including [górod] 'city', instead mark the plural with -*a* (Zaliznjak, 1977). These examples can be found in Table 3, where I derived the type counts by reanalyzing the data from Gouskova et al. (2015) to include plural allomorphy.

noun class	*count*	*example*	*gloss*	*plural*
regular	1461	naród	'people'	naród-ɨ
-*a* plural	64	górod	'city'	gorod-á

Table 3: Russian masculine plural forms

Under the sublexicon model, irregular nouns like [górod] 'city' would be marked with a lexical diacritic [PL:*a*] and collected into a corresponding sublexicon.[2] Now we have lexical diacritics along two dimensions: a noun's diminutive is determined by the [*ok*], [*ik*], and [*chik*] diacritics, while their plural is governed by the presence or absence of a [PL:*a*] diacritic. Ackerman et al. (2009) argue that we see effects in natural language where knowledge of one cell in a word's inflectional paradigm can help reduce the possibilities for another cell. Indeed, in Table 4 we see that this holds in Russian as well: further reanalysis of the data from Gouskova et al. (2015) suggests that [PL:*a*] nouns can almost always form diminutives with -*ok*, but are attested less commonly with -^{j}*ik* and -*t*$ʃ^{j}$*ik* diminutives than nouns with regular plurals are.

[2] For discussion of whether regular nouns like naród 'people' would similarly get a [PL:*ɨ*] diacritic, or whether the plural exponence would instead be subsumed under some broader declension class marker, see Section 5.2.

As shown in Table 4, roughly 30% of nouns with regular plurals (435 out of 1461) allow for diminutives with *-tʃʲik*, while only 11% of nouns with plural *-a* (7 out of 64) do. Similarly, *-ʲik* is an acceptable diminutive suffix for 32% of regular nouns (467), but only 19% of *-a* plurals. On the other hand, *-ók* diminutives are overrepresented among *-a* plural nouns: 95% (61) accept them, compared to only 56% of other nouns (827). Thus, *-ók* diminutives are substantially *overrepresented* among nouns with plural *-a*, while *-ʲik* and *-tʃʲik* are substantially *underrepresented*.

		attested?		
stem class	*diminutive*	*yes*	*no*	*% yes*
regular (1461 stems)	*-ók*	827	634	56.6%
	-ʲik	467	994	32.0%
	-tʃʲik	435	1026	29.8%
-a plural (64 stems)	*-ók*	61	3	95.3%
	-ʲik	12	52	18.9%
	-tʃʲik	7	57	10.9%

Table 4: Distribution of Russian plural *-a* stems with attested diminutives

The *SUBLEX constraints come in to capture the bias found in Table 4 by allowing the sublexical grammars for the diminutive diacritics [*ok*], [*ik*], and [*chik*] to check whether a lexical entry has the [PL:*a*] diacritic.

2.3.2 Sublexical dependencies

In my proposal, as the sublexical grammars are created and adjusted, they are seeded with *SUBLEX constraints that penalize membership in other sublexicons.[3] These constraints will be active (i.e, weighted heavily enough to affect the overall evaluation of a form) when knowledge of one derived form of a word is informative about another. Each sublexical grammar includes *SUBLEX constraints referencing sublexicons in the language's other partitions of the lexicon. For the Russian example, we have one plural sublexicon, [PL-*a*], and three diminutive sublexicons, [*ok*], [*ik*], and [*chik*]. The sublexical grammars for [*ok*], [*ik*], and

[3]There are two reasons why *SUBLEX constraints penalize the *presence* of a diacritic rather than its *absence*. The first is conceptual: negative *SUBLEX constraints match the phonotactic constraints in the sublexical grammars, which also penalize the presence of certain segments or sequences. The second is more technical: suppose a constraint penalizes a word for *not* having a [PL:*a*] diacritic. Two sets of words would fall into this category: words that the speaker knows take a different plural form (and have thus been assigned a different diacritic governing the plural), and words whose plural is unknown to the speaker (which thus have no diacritic associated with the plural in the lexicon). Since absence of information is the initial, default state, it should not itself be penalized.

[*chik*] would all contain *SUBLEX-[PL:*a*], which penalizes the small number of nouns that take -*a* in the nominative plural. On the other hand, the sublexical grammar for the [PL-*a*] sublexicon would include constraints penalizing membership in each of the diminutive sublexicons: *SUBLEX-[*ok*], *SUBLEX-[*ik*], and *SUBLEX-[*chik*].

How heavily would *SUBLEX-[PL:*a*] be weighted in each of the diminutive sublexicons? In Table 5, the data from Table 4 is reformatted to show the proportion of words with a [PL:*a*] diacritic in each diminutive sublexicon. Of the 888 words with attested -*ók* diminutives—which would thus be given [*ok*] diacritics and placed in the [*ok*] sublexicon—61 (6.9%) form the plural with -*a* and thus have the [PL:*a*] diacritic. In the [*ik*] and [*chik*] sublexicons, the proportion of words with [PL:*a*] is much smaller: 2.5% (12 out of 479) and 1.5% (7 out of 442), respectively.

Words with plural -*a* are thus significantly *more underattested* in the [*ik*] and [*chik*] sublexicons than in the [*ok*] sublexicon. Accordingly, the learner will weight *SUBLEX-[PL:*a*], which penalizes words with plural -*a*, *more heavily* in the [*ik*] and [*chik*] sublexical grammars than in the [*ok*] sublexical grammar.

sublexicon	*plural* -ɨ	*plural* -a	% [PL:*a*]	*SUBLEX-[PL:*a*] *weight*
[*ok*]	827	61	6.9%	low
[*ik*]	467	12	2.5%	high
[*chik*]	435	7	1.5%	high

Table 5: Distribution of Russian plural [PL:*a*] across diminutive sublexicons

The Russian example can give us a sense of what it means for *SUBLEX constraints to be active. When they are weighted irrelevantly low, we have the situation described by previous iterations of the sublexicon theory: when a word's exponent in a given environment is not already known, learners place it into a sublexicon based solely on its phonological characteristics, not its exponent in other environments. As Ackerman et al. (2009) argue, this is an inefficiently organized morphological system: one form provides no information about another. On the other hand, when a *SUBLEX constraint is weighted heavily enough as to be inviolable, this means that a speaker encountering one form of a word can entirely rule out a certain exponent of that word in a different environment. That is, if a word has exponent *X* in, say, the dative, it cannot have exponent *Y* in the genitive. To put it another way, higher ranked *SUBLEX constraints are indicative of a smaller overlap between two sublexicons.

In cases such as the Russian plural, the *SUBLEX constraint is given an intermediate weight: active, but not inviolable. When this happens, knowing a lemma's behavior in one environment can bias the prediction of its behavior in another environment, but does not fully determine it: the speaker will still, occasionally, guess wrong. For example, the Russian learner would see the word [rukáv] 'sleeve', whose plural is [rukav-á], and guess that it could only form a diminutive with *-ók*. However, this is the rare word that takes plural *-a* and does allow for a different diminutive, in this case [rukáv-tʃʲik].

In the next section, I show a case study in greater depth: the choice of allomorphy for a given Hungarian noun's possessive form can be predicted in part by other forms like the plural; moreover, this relationship is a strict correlation not due to other factors like incidental shared phonology. The standard sublexical phonotactic grammars are insufficient to predict a given noun's choice of possessive allomorph, but adding *SUBLEX constraints allows for a much better fit.

3 Morpholexical behavior in Hungarian possessives

3.1 Possessive allomorphy

The Hungarian possessive suffix, exemplified by noun forms with third-person singular possessors,[4] shows lexically specific allomorphy: some nouns take possessive forms beginning with [j] (*-jV*), while others have just a bare vowel (*-V*). Since the possessive suffix is also subject to standard backness harmony, there are a total of four possessive suffixes: back-harmonizing words select for *-ɒ* or *-jɒ*, while front-harmonizing words take *-ɛ* or *-jɛ*. Some words can take either suffix.[5] We can see examples of *-V* and *-jV* possessives in Table 6.

[4]The standard syntactic analysis (Szabolcsi, 1994; É. Kiss, 2002) argues that Hungarian has one morpheme expressing a Possessive head—which is *-(j)ɒ/(j)ɛ* in some cases and null in others—and another agreeing with the possessor, expressed as *-Vm* for 1SG and *-Ø* for 3SG. As such, given the noun [tɒnaːr] 'teacher', 'my teacher' is morphologically [tɒnaːr-Ø-om], while 'her teacher' is [tɒnaːr-ɒ-Ø].

[5]As noted by Kiefer (1985) and others, when a word accepts both possessive suffixes, this sometimes leads to a difference in meaning, generally of alienable vs. inalienable possession.

word	*gloss*	*possessive*
paːr	'pair'	paːr-**j**ɒ
ʃor	'line'	ʃor-ɒ
ɒblɒk	'window'	ɒblɒk-(**j**)ɒ

Table 6: Some Hungarian possessive forms (see Rácz and Rebrus, 2012)

The two allomorphs, *-V* and *-jV*, are both quite common: according to Rácz and Rebrus (2012), although *-V* is more common overall (in both type and token frequency), *-jV* is productive, given that it universally appears on novel forms like recent loanwords and nonce words. The authors also describe generalizations and patterns in possessive allomorphy selection (see Table 7). Some of these are categorical—for example, words that end in vowels *always* take *-jV*, while words that end in palatals and sibilants *never* do. Others are gradient tendencies—for example, they found that words ending in complex codas (geminates or clusters) have *-jV* more than words that end in singleton consonants.

variable	*predicted rate of -jV*
harmony class	back $>$ front
final coda	vowel $=$ 100% $>$ geminate, cluster $>$ singleton
final C place	labial $>$ alveolar $>$ velar $\gg$ palatal $=$ 0%
final C manner	sibilant $=$ 0%

Table 7: Predicted phonological effects on possessive allomorphy selection (from Rácz and Rebrus, 2012)

3.2 Lowering stems

Rácz and Rebrus (2012) also note a morphological effect in possessive allomorphy: irregular noun classes generally take *-V*. One such class involves the identity of the *linking vowel*, which typically appears when certain suffixes, such as the plural, accusative, and first-and second-person singular possessive, attach to consonant-final stems. For most nouns, this vowel is mid (e.g., [o]). However, for one lexically specified class known as *lowering stems*, it is instead low (e.g., [ɒ]).

In Table 8, we see the full range of linking vowels available for back-harmonizing words. First, vowel-final words like [kɒpu] 'gate' lack linking vowels entirely. Consonant-final words are then split between lowering stems, with a low linking vowel, and non-lowering stems, with a mid linking vowel. (For front vowels, the distinction between low and mid vowels is more complicated than a simple difference in vowel height—see

Section 3.3.2 and Siptár and Törkenczy (2000) for more details.) Table 8 also includes the possessive forms of these words: [kɒpu], like all vowel-final nouns, has *-jV*, as does the non-lowering stem [paːr] 'pair'; on the other hand, the lowering stem [ɟaːr] 'factory' takes *-V*.

stem class	*example*	*gloss*	*plural*	*possessive*
V-final	kɒpu	'gate'	kɒpu-k	kɒpu-**j**ɒ
regular	paːr	'pair'	paːr-o̲k	paːr-**j**ɒ
lowering	ɟaːr	'factory'	ɟaːr-ɒ̲k	ɟaːr-ɒ

Table 8: Hungarian linking vowels and lowering stems

As we see in (4) below, in Hungarian, possessive allomorphy and lowering stem status is not fully predictable from phonology alone: the four words shown are identical except for the initial consonant, yet each behaves differently with respect to linking vowel height and possessive allomorphy. The plural and possessive dimensions are cross-classifying: the forms in (4a) and (4b) have the regular plural *-ok*, while the lowering stems in (4c) and (4d) have the plural *-ɒk*. On the other hand, the nouns in (4a) and (4c) have *-V* in the possessive, while those in (4b) and (4d) take *-jV*.

(4) *Phonologically similar Hungarian nouns can differ in plural and possessive allomorphy*

a. kaːr-o̲k, kaːr-ɒ 'damage' b. paːr-o̲k, paːr-**j**ɒ 'pair'

c. ɟaːr-ɒ̲k, ɟaːr-ɒ 'factory' d. ɲaːr-ɒ̲k, ɲaːr-**j**ɒ 'poplar'

Nonetheless, the overwhelming majority of lowering stems take *-V*, as seen in Table 9, while non-lowering stems trend in the other direction. Out of 119 consonant-final monomorphemic nouns in Papp (1969) that are not marked as being variable in either the plural or the possessive, only 15 have *-jV*; on the other hand, nearly 60% of non-lowering stems have *-jV*. Looking along the columns of Table 9, the difference is also quite stark: 13.6% of all nouns taking *-V* are lowering stems, compared to only 1.6% of those that take *-jV*.

		possessive		
		-V	*-jV*	*% -jV*
stem class	*regular*	656	943	59%
	lowering	104	15	13%
	% lowering	13.6%	1.6%	

Table 9: Distribution of Hungarian lowering stems and possessive allomorphy

There is a clear asymmetry in the intersection between Hungarian possessive allomorphy and stem class seen in Table 9: lowering stems take *-jV* much less frequently than non-lowering stems. This, however, does not necessarily represent a learned correlation between two morphological patterns. An alternative possibility is that lowering stems tend to have a phonological form that is associated with *-V*. For example, nouns ending in palatals and sibilants always take *-V*, so if lowering stems overwhelmingly end in these consonants, we would expect numbers like we see in Table 9. Of the 119 lowering stems in the table, 48 (approximately 40%) end in a palatal or sibilant, while of the 1599 non-lowering stems, 478 (about 30%) do. While it is true that lowering stems are somewhat more likely to take *-V* based on their phonology alone, the difference is not large enough to fully explain the patterns in Table 9. Indeed, as I show in Section 3.3, the morphological effect of lowering stem status on possessive allomorphy cannot be reduced to coincidences of phonology.

3.3 Phonological and morphological factors influencing possessive allomorphy

In this section, I take a quantitative look at the features of a word that can predict its choice of possessive allomorph. I do this by fitting two linear regression models showing how various factors can be used to predict a given lexical item's possessive allomorph: the first includes exclusively phonological factors such as the manner and place of the stem's final consonant, while the second adds the morphological factor of whether a noun is a lowering stem.[6] If the morphological effect were itself reducible to phonological factors, we would expect the second model to look quite different from the first, as the explanatory power of the phonological factors would be partially attributed to the morphological factor. On the other hand, if stem class is a strong predictor independent of any phonological effect, the second model should include stem class as a strong, significant factor, without substantial interactions between stem class and the phonological factors.

3.3.1 Data

For my corpus, I manually transcribed data from Papp (1969), a morphological dictionary of Hungarian. Each entry, representing a lemma, includes information about its part of speech, morphological breakdown,

[6] The models were fitted using the `lrm` function in R's `rms` package (R Core Team, 2020; Harrell Jr., 2020).

stem changes, and—for nominals—the suffixes used in the accusative, plural, and possessive. These latter entries can also be used to infer a stem's vowel harmony and lowering status. My database includes 58,326 entries, of which 35,169 are listed as declining like nominals. This includes not just nouns, but also words that behave like both adjectives and nouns (as Rebrus and Szigetvári (2018) note, the syntactic distinction between the two is not always so clear in Hungarian), some numerals and pronouns, etc.

I then filtered the data set in a number of ways. First, I only looked at words ending in consonants. I also removed the small number of words ending in orthographic *h*, as they present a phonological complication: this letter usually represents /h/, which is often not pronounced word-finally, and at times it also represents /x/, at best a marginal phoneme in Hungarian. I also limited my search to monomorphemic nouns, under the non-lexicalist assumption shared by models like Distributed Morphology (Halle and Marantz, 1993) that lexically specific information about a noun's behavior is generally associated with particular morphemes rather than with complex words; the latter are not listed in the lexicon, but instead built up by the syntax. Finally, I only looked at words listed exclusively as nouns. By default, nouns are non-lowering stems, while all but a handful of adjectives are lowering stems, and some forms behave differently depending on the syntactic environment (Rebrus and Szigetvári, 2018). Accordingly, looking at words that can be used as adjectives can introduce additional complications. Finally, I excluded any stems whose possessive form is either variable or unknown. This left a total of 2,433 nouns. However, I also fitted models with the same factors on two larger data sets: one including all nouns regardless of morpheme structure, and one including all nominals regardless of morpheme structure. Although the model coefficients were different, the crucial finding in Section 3.3.3—the fact that adding stem class as a factor improves the model—was true across all three data sets. Thus, the results in Section 3.3.3 are not dependent on the particular way in which I filtered my data.

My data set, which is based on a structured dictionary, differs significantly in its design from that of Rácz and Rebrus (2012), who use a web corpus. Their search includes all consonant-final nominals that appear in the possessive, including morphologically complex stems and stems listed by Papp (1969) as including adjectival uses, for a total of 22.8 thousand consonant-final types and about 10 million tokens. Indeed, they explain some of their results with reference to the selectional restrictions of individual suffixes, including the comparative. To the extent that my results in Section 3.3.3 differ from theirs, then, much of this difference can be reduced to the effect of excluding morphologically complex words, which reflects different assumptions about the lexicon.

3.3.2 A note on the coding of stem class

Any quantitative treatment of lowering stems in Hungarian runs into the problem that the distinction between lowering and non-lowering stems breaks down for certain words: as shown in Table 10, back-harmonizing words have mid ([o]) and low ([ɒ]) linking vowel variants. Front-harmonizing words with front rounded vowels [ø øː y yː] in the last syllable also have mid and low variants, [ø] and [ɛ]. However, front-harmonizing words with unrounded vowels [ɛ eː i iː] in the last syllable only have one linking vowel, [ɛ].

harmony	*non-lowering*		*lowering*	
back	paːr-ok	‘pairs’	ɟaːr-ɒk	‘factories’
front rounded	ʃyl-øk	‘porcupines’	fyl-ɛk	‘ears’
front unrounded	bɛj-ɛk	‘beys’	tɛj-ɛk	‘milks’

Table 10: Linking vowels in Hungarian plurals according to stem harmony class

Siptár and Törkenczy (2000) include some words with front unrounded harmony, such as [tɛj] ‘milk’ from Table 10, in their list of lowering stems, even though these words’ linking vowel is not actually phonetically distinct from that of stems that the authors do not consider to be lowering stems, like [bɛj] ‘bey’. This determination is based on a number of factors:

1. Some classes of nouns with stem changes always take low linking vowels when the distinction is visible, so this implicational relationship is extended to nouns with these stem changes that show front unrounded harmony as well.
2. Low linking vowels generally appear in the accusative, while mid linking vowels may not. Accordingly, the accusative of [bɛj] is [bɛj-t], while the accusative of [tɛj] is [tɛj-ɛt], suggesting that the linking vowel in the latter may be “phonologically low”.
3. Some Hungarian dialects do distinguish between the mid short vowel [e] and the low short vowel [ɛ] where standard Hungarian only has [ɛ]. In these dialects, [tɛj] has the low linking vowel like [fyl] ‘ear’ (plurals [tɛj-ɛk] and [fyl-ɛk]), while [bɛj] has the mid linking vowel [e], as we see in the plural

[bɛj-ɛk] (Bárczi and Országh, 1962).

Theoretically speaking, I do not wish to claim that speakers necessarily have an entailment between irregular behavior in one morphophonological environment and irregular behavior in another *in the generative grammar* (regarding the first two points above). I certainly do not wish to assume that the organization of one dialect's morphophonology necessarily tells us anything about the linguistic system of speakers of another, crucially different, dialect (regarding the third point). In addition, practically speaking, my data source does not fully distinguish between front unrounded stems that linguists mark as lowering stems and those that they mark as non-lowering stems. Accordingly, I mark *all* stems with front unrounded harmony—including both [bɛj] 'bey' and [tɛj] 'milk'—as undetermined for lowering stem class. This means that there is a large class of nouns in my data whose categorization in the lowering stem factor is determined phonologically, rather than marking its morphological behavior as is the case for other nouns. This will be important to keep in mind in interpreting some of the results below.

3.3.3 Results

The resulting full models can be found in Appendix A and include a number of phonological factors: the place and manner of the final consonant, the height and length of the final syllable's vowel, the stem's vowel harmony class, the complexity of the final coda, and whether the stem was monosyllabic. In Table 11, I use one factor, final consonant place, to demonstrate how adding stem class as a factor affects the other factors in the model. In both models, the effect size for individual values of the final consonant place variable are negative, meaning that a noun ending in a labial, palatal, or velar consonant is *less* likely to take *-jV* in the possessive than the baseline for this factor, alveolar-final nouns. This result differs from that of Rácz and Rebrus (2012), who found that words ending in labials use *-jV* more frequently than alveolar-final ones (see Table 7). This difference is likely due to a different input, as discussed in Section 3.3.1: their corpus includes adjectives, as well as morphologically complex words, while mine does not. Indeed, the authors attribute the labial effect to the comparative suffix *-bb*, which usually takes *-jV*. Thus, the discrepancy between my results and those of Rácz and Rebrus is not problematic.

In Table 11, the effect size represents the strength of the effect; the strongest effect can be seen in palatals, which categorically trigger selection of *-V*.

predictor	*effect size in model without stem class*	*effect size in model with stem class*	*predicted rates*
final C: alveolar	(baseline)		
final C: labial	−2.02	−2.22	tsiːm-jɛ < siːn-jɛ
final C: palatal	−8.88	−9.25	ryɟ-jɛ < tʃyd-jɛ
final C: velar	−3.26	−3.54	gøːg-jɛ < tʃøːd-jɛ
stem class: regular	(baseline)		
stem class: lowering	—	−3.71	ɲaːr-jɒ < paːr-jɒ

Table 11: The effect of final consonant place on Hungarian possessive allomorphy, with and without stem class as a predictor

When we add stem class to the model, we are looking at several factors. First of all, adding stem class to the model substantially and significantly improves its fit, even when we factor in a penalty for adding another factor to the model (see Appendix A for details). That is, stem class is *doing something* in our model, allowing it to make a more accurate prediction of possessive allomorphy. Next, we can see in Table 11 that being a lowering stem has a fairly substantial negative effect, confirming the statistical tendency for lowering stems to take *-V*. Finally, we must determine whether the factor of stem class is acting independently from the phonological factors—or whether, on the contrary, it is duplicating the work of the phonological factors. As we can see in Table 11, adding stem class to the model does not make a substantial difference for the sizes of the other effects. (Recall that, here, final consonant place is representative of all of the phonological factors in the full original model found in Appendix A, including consonant manner, monosyllabicity, etc. All of these phonological factors behave similarly to final consonant place.) This may suggest that stem class is largely independent of the other factors.

We can confirm this intuition with further tests. First, I looked at collinearity among the model's variables. When factors are collinear, they are describing the same effect. If the *lowering* variable showed high collinearity, that would indicate that its effect is reducible to some combination of the phonological effects, rather than being a genuine morphological effect. I measured collinearity using the variance inflation factor (VIF).[7] A VIF of 1 indicates a lack of collinearity, while a VIF higher than either 5 or 10 for a given variable

[7]For each coefficient $\hat{\beta}_j$ in the model, VIF compares the amount of variance exhibited by $\hat{\beta}_j$ in the model to the amount of variance $\hat{\beta}_j$ would exhibit in a model with no other factors.

indicates that the amount of collinearity is problematic (James et al., 2013). In both the basic phonological model and the model with stem class, no variable has a VIF exceeding 4 (the *approximant* value for word-final consonant manner has a VIF of 3.99 and 3.66 in the two models, respectively). Crucially, the VIFs for different values of the stem class variable are low: 1.16 for lowering stems and 2.53 for undetermined stems—which, as discussed in Section 3.3.2, mainly consists of stems with front unrounded harmony.

I also looked to see whether stem class has any significant interactions with phonological factors. I created a series of new models that added an interaction term between stem class and each phonological factor, respectively. Although some of these additional terms did significantly improve the model, the only significant interactions involved the *undetermined* stem class—which, again, consists of stems with front unrounded harmony, making this a phonological effect rather than a morphological one. Information about the models with interaction terms can be found in Appendix A. To summarize, then, the class of lowering stems cannot be decomposed into a series of phonological factors, and it does not significantly interact with any of the phonological factors, either.[8]

This suggests, then, that being a lowering stem really is predictive of possessive allomorph selection *independent* of the effect of the various phonological factors: taking stem class into consideration allows speakers to more accurately and efficiently predict a given noun's choice of possessive allomorph. Or, to put it another way, if learners sees a noun's phonological form (for example, the bare nominative singular), this form will be helpful to them in formulating a guess for the possessive. However, if learners encounter a form like the plural to learn that the noun is a lowering stem, this inflected form is even more helpful in generating the correct possessive form.

4 Sublexicons and *SUBLEX constraints in Hungarian

In Section 3 above, I showed that the choice of possessive allomorph in Hungarian is sensitive to both phonological and morphological effects. Let us now see how the sublexicon model presented in Section 2

[8] Of course, as we make our terms more complicated, the small number of lowering stems makes it difficult for such interaction terms to reach statistical significance. However, this plausibly means that the learner would be unable to make such inferences as well given the state of the lexicon.

applies to this case.

4.1 Diacritics and rules of exponence

We have encountered two dimensions of lexically specific allomorphy in Hungarian: possessive allomorphy, which involves the presence or absence of [j] in *-jV* and *-V*, and lowering stem status, which involves the height of the "linking vowel" in forms like the plural. As described in Section 2.1, this allomorphy is implemented through lexically specific rules of exponence including diacritics, here [±j] and [±lower]. One such implementation of possessive allomorphy is seen in (5)—though this is by no means the only option, and is at any rate an oversimplification, ignoring vowel harmony:

(5) *Rules of exponence for the Hungarian possessive*

a. POSS ↔ jɒ / [+j] ___

b. POSS ↔ ɒ / [−j] ___

Capturing lowering stems is slightly more complicated. Recall that the basic form of the plural is *-k*; this surfaces in vowel-final stems like [kɒpu] 'gate' (plural [kɒpu-k]). The linking vowel appears after consonant-final stems in certain suffixes like the plural. For this analysis—again, by no means the only option—I assume that suffixes that take linking vowels, like plural *-k*, are lexically marked with a diacritic [*LV*], and that the linking vowel is inserted through a readjustment rule when placed before a morpheme marked with [*LV*]. The choice of linking vowel depends on the value of the [±lower] diacritic on the stem. This yields the set of rules in (6):

(6) *Rules of exponence and readjustment involving linking vowels*

a. PL ↔ $\mathrm{k}_{[LV]}$

b. Ø → o / [−lower] ___ [*LV*]

c. Ø → ɒ / [+lower] ___ [*LV*]

Entries in the lexicon include diacritics that determine which rules in (5) and (6) apply to a particular word, thus determining their possessive and plural forms. We can see some examples of this in Table 12. The lexical entry for [paːr] 'pair' contains the diacritics [+j] and [−lower]. The presence of [+j] ensures that rule (5a) will apply in the context of this word, meaning that the possessive will be spelled out as *-jɒ*, yielding the inflected form [paːr-jɒ]. In the formation of the plural, two rules will apply: first, (6a) spells out the plural as *-k* (for all words—it is not context-specific). Next, since [paːr] has a [−lower] diacritic, the context-specific readjustment rule (6b) inserts the mid linking vowel [o] in this context, which yields the plural form [paːr-ok].

lexical entry		*possessive*	*plural*
kaːr$_{[-\text{j},\ -\text{lower}]}$	'damage'	kaːr-ɒ	kaːr-o̲k
paːr$_{[+\text{j},\ -\text{lower}]}$	'pair'	paːr-**j**ɒ	paːr-o̲k
ɟaːr$_{[-\text{j},\ +\text{lower}]}$	'factory'	ɟaːr-ɒ	ɟaːr-ɒ̲k
ɲaːr$_{[+\text{j},\ +\text{lower}]}$	'poplar'	ɲaːr-**j**ɒ	ɲaːr-ɒ̲k

Table 12: Lexical entries and surface forms for some Hungarian nouns

4.2 Sublexicons and sublexical grammars

As before, each diacritic is associated with a sublexicon containing all words that have that diacritic, regardless of any other diacritics appearing on a given word's lexical entry:

(7) *Possessive and lowering sublexicons*

a. [−j]: {kaːr$_{[-\text{j},\ -\text{lower}]}$, ɟaːr$_{[-\text{j},\ +\text{lower}]}$, ...}

b. [+j]: {paːr$_{[+\text{j},\ -\text{lower}]}$, ɲaːr$_{[+\text{j},\ +\text{lower}]}$, ...}

c. [−lower]: {kaːr$_{[-\text{j},\ -\text{lower}]}$, paːr$_{[+\text{j},\ -\text{lower}]}$, ...}

d. [+lower]: {ɟaːr$_{[-\text{j},\ +\text{lower}]}$, ɲaːr$_{[+\text{j},\ +\text{lower}]}$, ...}

In my version of the sublexicon model, each sublexicon is described by a sublexical grammar including both phonotactic and *SUBLEX constraints, the latter of which penalize lexical items that have a particular diacritic. To demonstrate this, I trained the UCLA Phonotactic Learner (Hayes and Wilson, 2008) on a data set of 4,193 words, identical to the one described in Section 3.3.1 but with the addition of vowel-final words. The data set was split into two sublexicons comprising words that take *-jV* (2,956 words) and *-V* (1,237 words) in the possessive, respectively. This process thus produced sublexical grammars for the two possessive sublexicons, [+j] and [−j].

The Phonotactic Learner does not directly allow for the manipulation of constraints that do not reference segments and features, such as my *SUBLEX constraints. One solution is to adopt the approach to diacritics assumed by Chomsky and Halle (1968), in which lexical diacritics like [±j] filter down to every segment of lexical items that bear them. This, however, leads to a combinatorial explosion in the number of distinct segments in the phonological inventory: each phonologically distinct segment, such as [t], must now come in flavors representing every possible combination of lexical features. Thus, Hungarian would have a [t] marked [+j,−lower] that appears in [+j,−lower] words, a [t] marked [+j,+lower] that appears in [+j,+lower] words, and so on. Although this raises the potentially interesting possibility of constraints defined by an interaction of lexical and phonological features (for example, $*t_{[+\text{lower}]}$), these are peripheral to my research question and taxing in their implementation.

Instead, I implemented the *SUBLEX constraints by putting a dummy segment at the beginning of each word that was given a [+sublex] feature, while all "real" segments were given [−sublex]. This segment also had a [±lower] feature: lowering stems started with a [+sublex, +lower] dummy segment, while non-lowering stems had a [+sublex, −lower] dummy segment. Words that were undetermined or variable for lowering (see Section 3.3.2) were left unvalued for [±lower], leaving only the [+sublex] feature. Segments with [+sublex] were then projected onto a separate Sublex tier which allowed for unigram constraints, meaning that the only possible constraints that could be induced on this tier were *[+lower] and *[−lower]—equivalent to *SUBLEX-[+lower] and *SUBLEX-[−lower], respectively. Of course, this implementation detail did leave its trace on the induced constraints: the most heavily weighted constraint induced in both sublexical grammars was *#[−sublex], penalizing every word that began with a true segment that was not one of the dummy [+sublex] pseudophonemes.

Table 13 shows some constraints found by the Phonotactic Learner for the [+j] and [−j] sublexical grammars. Here I only include constraints that do not express general phonotactic principles of Hungarian more broadly—that is, constraints that the Phonotactic Learner does not induce for the full data set comprising the union of the [+j] and [−j] sublexicons. Some of the most heavily weighted constraints in the [+j] sublexical grammars are the preferences discussed in Section 3: words ending in sibilants are heavily penalized, as are those marked [−anterior], which includes the palatals [c ɟ j ɲ] (and their geminate forms) and the post-alveolar sibilants [tʃ dʒ ʃ ʒ] (and their geminates). Together, these constraints cover the categorical ban on *-jV* with words ending in palatals and sibilants. The induced [+j] sublexical grammar also includes a strongly weighted constraint against lowering stems.

The [−j] sublexical grammar, on the other hand, captures the fact that words ending in vowels never take *-V* in the possessive with a constraint penalizing word-final vowels. This is the only phonologically natural constraint found in the [−j] grammar. For the purposes of illustration, I include one "unnatural" constraint found by the Learner: a prohibition against voiced geminate sibilants [zː ʒː dzː dʒː]. To a linguist, there is no phonologically sensible reason why words that contain a voiced geminate sibilant at any point in the word should favor one possessive over another—perhaps these segments are simply underattested in Hungarian on the whole, even though this constraint did not make it into the [+j] sublexical grammar. In general, the grammars induced by the Phonotactic Learner include many similar constraints, unexpected for a linguist's analysis, which penalize specific sounds or sound sequences. More examples can be seen in the full induced grammars found in Appendix B.

sublexicon	*constraint*		*description*	*weight*
[+j]	*[+lower]	(Sublex tier)	no lowering stems	4.330
	*[−anterior]#		no word-final palatals or post-alveolars	4.285
	*[+strident]#		no word-final sibilants	3.914
[−j]	*[+syllabic]#		no word-final vowels	3.718
	*[+voice,+strident,+long]		no voiced geminate sibilants	1.333

Table 13: Selected constraints from the [+j] and [−j] sublexical grammars induced by the UCLA Phonotactic Learner (Hayes and Wilson, 2008)

4.3 Using the sublexical grammars

In the sublexicon theory, the learner applies the sublexical grammars to nonce words to determine which sublexicon to put them into—that is, which diacritic to place onto these words' lexical entries—and thus what rule of exponence to apply to them. Let us consider an example. In Table 14 we have a speaker's information about two very phonologically similar nonce words, [luʒːud] and [luzːoːd]. However, there is an important *morphological* distinction between the two words: for [luʒːud], the speaker has only seen the singular form, while for the other word, the speaker has previously encountered the plural form [luzːoːd-ɒk]. From this, the speaker can infer that [luzːoːd] is a lowering stem, and has placed a [+lower] diacritic on its lexical entry.

singular	*plural*	*lexical entry*
luʒːud	*(unknown)*	luʒːud
luzːoːd	luzːoːd-ɒ̲k	luzːoːd$_{[+\text{lower}]}$

Table 14: Known forms and lexical entries of the nonce words [luʒːud] and [luzːoːd]

Next, Figure 2 and Figure 3 show how each of the nonce words is evaluated on the sublexical grammars for the [+j] and [−j] sublexicons, respectively, presented in Table 13. The two behave identically on the [−j] grammar: since both contain a geminate voiced sibilant, they incur a violation of *[+voice,+strident,+long], which gives both a score of −1.333. However, they diverge on the [+j] grammar: [luʒːud] does not end in a palatal or sibilant, nor does it have a [+lower] diacritic, so it incurs no violations. On the other hand, since [luzːoːd] does have a [+lower] diacritic on its lexical entry, it is penalized for violating *SUBLEX-[+lower], yielding a score of −4.330.

constraint / *weight*	*SUBLEX-[+lower] 4.330	*[−anterior]# 4.285	*[+strident]# 3.920	...	total
luʒːud	0	0	0	...	0
luzːoːd$_{[+\text{lower}]}$	−4.330	0	0	...	−4.330

Figure 2: Evaluation of nonce words [luʒːud] and [luzːoːd] on the [+j] sublexical grammar

constraint / *weight*	*[+syllabic]# 3.718	*[+voice,+strident,+long] 1.333	...	total
luʒːud	0	−1.333	...	−1.333
luzːoːd$_{[+\text{lower}]}$	0	−1.333	...	−1.333

Figure 3: Evaluation of nonce words [luʒːud] and [luzːoːd] on the [−j] sublexical grammar

The two words thus vary in their relative evaluation on the sublexical grammars: [luʒːud] does better on the [+j] grammar ($0 > -1.333$), while [luzːoːd] is rated better by the [−j] grammar ($-1.333 > -4.330$). The sublexicon theory predicts that speakers' behavior should differ accordingly: they should give [luʒːud] a [+j] diacritic and [luzːoːd] a [−j] diacritic, meaning that they will output the possessive forms [luʒːud-jɒ] and [luzːoːd-ɒ] after the application of rules of exponence (5a) and (5b), respectively.

In Section 3, I showed that the distribution of Hungarian possessives is sensitive to both phonological and morphological factors. The phonotactic learning process applied to the Hungarian possessive sublexicons shows that the sublexicon theory can capture these sensitivities in a formal model that allows speakers to use the sublexical grammars to generate novel forms of nonce words, and incorporate both their phonological form and their observed paradigm in doing so.

5 Discussion

5.1 The Paradigm Cell Filling Problem

As discussed in Section 2.1, the sublexicon theory proposes a model for how speakers learn and keep track of phonological and morphological factors influencing morphological behavior. Importantly, it lies outside any generative grammar that derives phonological outputs from abstract morphosyntactic features and/or structures. This makes it compatible with any theory of morphology that marks lexically specific morphological patterning through diacritical information attached to individual items. This includes many disparate morphological theories—such as Distributed Morphology (Halle and Marantz, 1993) and Paradigm Function Morphology (Stump, 2019)—but by no means all. For example, in Nanosyntax (Baunaz and Lander, 2018), allomorphy is attributed to differences in fully elaborated syntactic structures; lexical items are associated with syntactic trees of greater or lesser size rather than diacritics.

In the taxonomy of Stump (2001), Distributed Morphology is also a *lexical* theory, in which words and sentences are built up from affixes that are lexically associated with particular morphosyntactic features.[9] In the formal derivation, a word is built up from a syntactic structure and thus stands in a part–whole relation to the root and affixes of which it is composed, but does not have any formal relation to other words sharing a root. As Embick and Marantz (2008) point out, this means that Distributed Morphology cannot formulate global blocking rules like "The word **gloriosity* is invalid because its 'slot' is already filled by *glory*": there

[9]Not to be confused with *lexicalist* theories, in which the morphological word holds a privileged position over smaller or larger phonological, syntactic, or semantic units—a position which Distributed Morphology rejects.

is no paradigm cell filled by *glory* that is blocking the creation of **gloriosity*. In this sense, paradigms in Distributed Morphology are an epiphenomenon: any paradigmatic effects present in the formal grammar do not come from formal objects representing paradigms.

Ackerman et al. (2009) and Ackerman and Malouf (2013) criticize views of morphology in which paradigmatic relations play no role, arguing that crosslinguistic patterns in morphological organization only make sense when viewed through relations between members of an inflectional paradigm. In particular, they suggest that morphological systems are organized in a way that facilitates solving the Paradigm Cell Filling Problem—that is, how to correctly identify one member of an inflectional paradigm given one or more others. In other words, they find that morphological dependencies within a language such as the implicational relationship between lowering stems and *-V* possessives in Hungarian are much stronger than we would expect from a random distribution of morphological exponents.

The sublexicon model with *SUBLEX constraints is a proposal for how speakers can incorporate and take advantage of paradigmatic information to aid morphological learning. Sublexicons do not allow direct access to related forms in the course of a derivation, so rules like the global blocking constraint described above would still be impossible in Distributed Morphology with sublexicons. Instead, a speaker's experience of having seen multiple forms of a related word is represented through diacritics on that word's lexical entry; the sublexical grammars can then make generalizations over these diacritics. Thus, the lexical entries and sublexical grammars act as stores of paradigmatic information related to previously encountered forms, but they do not directly store these forms. Instead, all output forms are derived online by the formal grammar, which associates the lexical diacritics with their corresponding exponents through contextually specified rules of exponence. In a Distributed Morphology model with sublexicons, then, paradigmatic relations are captured indirectly in the form of lexical diacritics, and the distribution of exponents is captured by *SUBLEX constraints that can derive the implicational relationships within a language's morphological organization.

Finally, it is worth discussing an example of what *SUBLEX constraints do not do. One important debate in morphology is how to handle *syncretism*: when a word has the same form in two different morphological environments. For example, in many Slavic languages, the accusative and genitive forms of masculine animate nouns are often identical. Moreover, this is not a feature of one particular accusative/genitive exponent, but rather holds across multiple such exponents (hence, it is a *metasyncretism*). In Paradigm Function Morphology, metasyncretisms can be handled by means of a cross-reference in the mapping between the content paradigm (which contains morphosyntactic information like case) and the form paradigm, so that accusative content will be mapped to genitive form, or vice versa (Stump, 2016). On the other hand, theories like Distributed Morphology and Nanosyntax will appeal to shared structure between the accusative and the genitive: perhaps the two cases decompose into a number of subatomic features, one of which is shared between the two; or, alternatively, perhaps the accusative structurally contains the genitive (or vice versa). All of these approaches to syncretism are located in the architecture of the formal grammar; thus, the *SUBLEX constraints have nothing to say about them. (See Kramer (2016) for an illustration of how syncretisms can be used to compare morphological theories.) If identity between accusative and genitive is built into the formal grammar, then one diacritic will govern both of them; that is, a given noun's membership in one sublexicon will determine its form in both cases. If, on the other hand, this identity is viewed as coincidental in the formal grammar, then it may include two separate groups of diacritics for the accusative and genitive, respectively—this would allow a noun to fall in different sublexicons for the two cases. Of course, in this latter instance, there would be very strong *SUBLEX constraints discouraging learners from placing a noun in a genitive sublexicon corresponding to a form not identical to its accusative (and vice versa). The point, though, is that sublexicons are not themselves a grammatical approach to issues like syncretism; rather, the use of sublexicons and diacritics is determined by the architecture of the grammar.

5.2 Sublexicons and declension classes

The Hungarian case study discussed in Section 4 involves two diacritics, [±j] and [±lower], which are each associated with a single dimension of variation: the presence of a glide in the possessive, and the height of the "linking vowel" in cases like the plural, respectively. In Hungarian, the choice of *-V* or *-jV* in the possessive is not part of a larger pattern of behavior that extends to other morphosyntactic environments.[10]

[10]The glide—or lack thereof—appears in possessive forms other than the third person singular: for example, singular nouns with third person plural possessors ([paːr-j-uk] 'their pair' vs. [ɟaːr-Ø-uk] 'their factory') and across the paradigm of plural possessed nouns ([paːr-j-ai] 'his pairs' and [ɟaːr-Ø-ai] 'his factories'). However, under standard morphosyntactic analyses (see É. Kiss, 2002), all of these cases involve spellout of the same Poss head, so this is different from Russian-style declension classes as I discuss them here.

This is very different from languages like Russian, where words are grouped into four major classes that determine their forms across the different cases and numbers:[11]

	'law' *zakon*	*'room'* *komnata*	*'bone'* *kost^j*	*'wine'* *v^{j}ino*
class	I	II	III	IV
nominative	zakon	komnat-a	kost^j	v^{j}in-o
accusative	zakon	komnat-u	kost^j	v^{j}in-o
genitive	zakon-a	komnat-i	kost^j-i	v^{j}in-a
dative	zakon-u	komnat-e	kost^j-i	v^{j}in-u
instrumental	zakon-om	komnat-oj	kost^j-ju	v^{j}in-om
locative	zakon-e	komnat-e	kost^j-i	v^{j}in-e

Table 15: Major noun declension classes of Russian in phonemic transcription (from Corbett and Fraser, 1993)

Declension classes are, essentially, diacritics that appear in more than one rule of exponence. For example, the class II diacritic would appear in the context of rules spelling out the nominative as *-a*, the accusative as *-u*, etc.:

(8) *Rules of exponence for Russian class II nouns*

a. NOM ↔ a / [*II*] ___

b. ACC ↔ u / [*II*] ___

Of course, each of these diacritics is associated with a sublexicon that contains the words in a given declension class. Accordingly, the formation of declension classes during the learning process can be accomplished through the manipulation of sublexicons. We can imagine a couple of ways in which this process might happen. One possibility is that declension classes are formed through *merger*. In this case, at first the learner collects all stems that exhibit the same behavior in a given case into a sublexicon, so [komnat-] 'room', which has nominative [komnata] and accusative [komnatu], would be placed into sublexicons [NOM:*a*], [ACC:*u*], etc. However, these two sublexicons will contain an almost identical set of nouns, and thus nearly identical sublexical grammars as well. This makes them candidates for merger: when the two sublexicons have crossed some established threshold in size and overlap, they are merged into a single sublexicon, with

[11] As Parker and Sims (2020) demonstrate, these four classes break down into a number of exceptional subclasses—-for example, the class I nouns like [górod] 'city' discussed in Section 2.3.1 that take *-a* in the nominative plural—and cross-classify with other dimensions of morphological variation, such as stem changes and stress placement.

any exceptions spun off into separate, much smaller sublexicons. One such "exceptional" sublexicon would be comprised of class I nouns like [górod] 'city' that take *-a* in the nominative plural instead of the regular *-i*, as discussed in Section 2.3.1. In this way, the class sublexicons that appear in (8) and similar rules of exponence can be built up one case at a time.

Another possibility is that declension classes are never formed in the grammar at all, but are simply emergent, resulting from significantly overlapping sublexicons. In the Russian example, this would mean that the adult grammar doesn't contain a class II diacritic at all; instead, class II nouns would be defined by sharing a collection of lexical diacritics: [NOM:*a*], [ACC:*u*], etc.. In this case, morphological inferences seemingly attributable to declension class would instead be the result of one (or more) of a number of *SUBLEX constraints that are nearly collinear, given that the sublexicons they describe are almost identical: *SUBLEX-[NOM:*a*], *SUBLEX-[ACC:*u*], and so on. Which of these largely identical constraints ends up bearing the weight in a given grammar could be largely random, varying from speaker to speaker, an artifact of the learning process that leaves little or no real observable difference in behavior.

Thus, using sublexicons to address issues of morphological organization can shed light on a range of topics, including providing a framework for a readily implementable algorithm of learning declension classes—or, alternatively, imagining declension classes as an emergent property.

6 Conclusion and future directions

In this paper, I proposed an extension to the phonotactic sublexicon grammars of Gouskova et al. (2015), adding morphological constraints, called *SUBLEX that coreference sublexicon membership along different dimensions. I presented allomorphy selection in the Hungarian possessive as a case study showing that these *SUBLEX constraints allow the model to more fully capture the factors influencing it: words that choose their linking vowel from the series of low vowels [ɒ ɛ] instead of the usual mid vowels [o ɛ ø] are also more likely to take *-V* in the possessive. I showed that this previously observed effect (Rácz and Rebrus, 2012; Rebrus, 2013; Rebrus et al., 2017) is statistically robust, though not categorical, and stands orthogonal to the many effects of phonological form. I then discussed how the *SUBLEX constraints solve the Paradigm Cell

Filling Problem (Ackerman et al., 2009; Ackerman and Malouf, 2013) by finding correlations among lexical diacritics, and considered implications of my model for languages like Russian, where words are assigned to declension classes that govern their behavior in a number of morphosyntactic environments.

The model described in this work makes concrete claims about the behavior of adults and child learners when confronted with new words: a set of sublexical grammars induced from patterns in the lexicon will evaluate novel forms and assign them probabilities of being placed into each sublexicon based on how good a fit a novel form is for that sublexicon's grammar. Although Rebrus et al. (2017) and others have claimed that *-jV* is productive and that the set of nominal stems with low linking vowels is a closed class that does not accept nonce words (contra MacWhinney, 1975), the model presented in this work provides a way to test the morphological relationship between linking vowel height and possessive allomorphy all the same: if speakers are shown evidence that a novel word takes low linking vowels—if they see a plural *-ɒk* ending, for example—will they generate a possessive form for this word without [j]? The sublexicon model with morphophonological sublexical grammars provides a concrete framework for approaching this question and similar issues in Hungarian, as well as other languages that show several kinds of lexically specific allomorphy.

A Appendix: Full models

This appendix contains the full models calculated for predicting possessive allomorphs among monomorphemic Hungarian nouns without variable possessive forms, with data taken from Papp (1969). The models were calculated using the `lrm` function in R's `rms` package (R Core Team, 2020; Harrell Jr., 2020). The first model (Table 16) contains only phonological factors, while the second (Table 17) also includes the morphological factor of lowering stem class.

I assembled the models using stepwise comparison, where each additional factor significantly improved the model ($p < .0001$) according to the same package's `lrtest` function. For both cases, an ANOVA confirmed that each factor in the final models was significant ($p < .0001$). Furthermore, each addition to the models improved their Akaike Information Criterion (AIC), a metric that rewards model likelihood but penalizes model complexity—that is, the improvement in model fit brought by adding each factor outweighed the cost of having a more complex model.

In addition, this appendix includes information about the models with interaction terms, which are built upon the model in Table 17. Each such model contains one interaction term between stem class and a phonological variable; I include the coefficients for the intercept, stem class, that variable, and significant interaction terms. Coefficients of other phonological variables are omitted. In addition, Table 24 contains a summary of comparisons between the models with interaction terms and the base model in Table 17, including whether the model is an improvement in terms of basic fit (χ^2 with associated p-value) as well as the difference in AIC between the base model and the interaction model—when this is positive, the AIC for the base model is *higher* than that of the interaction model, suggesting that the interaction model is *better* at modelling the underlying distribution, even taking into account the model's additional complexity.

	β coef	*SE*	*Wald z*	*p*	*predicted rates*
Intercept	**3.02**	**.32**	**9.55**	**<.0001**	**bot-jɒ > bot-ɒ**
C Manner (default: plosive)					
fricative	**–1.44**	**.39**	**–3.73**	**.0002**	**raːf-jɒ < laːp-jɒ**
sibilant	**–10.69**	**.80**	**–13.36**	**<.0001**	**hɒs-jɒ < lɒt-jɒ**
nasal	**–1.95**	**.27**	**–7.16**	**<.0001**	**oːn-jɒ < toːt-jɒ**
approximant	**–4.08**	**.30**	**–13.47**	**<.0001**	**tʃɛr-jɛ < tsɛt-jɛ**
C Place (default: alveolar)					
labial	**–2.02**	**.26**	**–7.94**	**<.0001**	**tsiːm-jɛ < siːn-jɛ**
palatal	**–8.88**	**1.10**	**–8.06**	**<.0001**	**ryɟ-jɛ < tʃyd-jɛ**
velar	**–3.26**	**.29**	**–10.96**	**<.0001**	**gøːg-jɛ < tʃøːd-jɛ**
Harmony (default: back)					
front	**–2.03**	**.18**	**–10.96**	**<.0001**	**køb-jɛ < dob-jɑ**
variable	**2.26**	**.97**	**2.33**	**.0197**	**ɒnkeːt-jɒ/jɛ > klɒrineːt-jɒ**
V Height (default: mid)					
high	**1.73**	**.22**	**7.89**	**<.0001**	**cuːk-jɒ > tʃoːk-jɒ**
low	0.28	.19	1.50	.1342	kɒr-jɒ > kor-jɒ
V Length (default: short)					
long	**1.40**	**.17**	**7.98**	**<.0001**	**boːr-jɒ > bor-jɒ**
Coda (default: singleton)					
geminate	**2.47**	**.40**	**6.25**	**<.0001**	**ʃikː-jɛ > sik-jɛ**
cluster	0.04	.21	0.18	.8602	domb-jɒ > dob-jɒ
Syllables (default: monosyllabic)					
polysyllabic	**1.15**	**.17**	**6.67**	**<.0001**	**ɛlɛm-jɛ > sɛm-jɛ**

Table 16: Regression model with phonological predictors of possessive [j], with significant effects bolded

	β coef	*SE*	*Wald z*	*p*	*predicted rates*
Intercept	**3.53**	**.33**	**10.60**	**<.0001**	**bot-jɒ > bot-ɒ**
Stem class (default: non-lowering)					
lowering	**–3.71**	**.44**	**–8.44**	**<.0001**	**ɲaːr$_{[+\text{lower}]}$-jɒ < paːr$_{[-\text{lower}]}$-jɒ**
undetermined	–0.25	.25	–0.98	.3278	(no direct comparison)
variable	**–2.76**	**.69**	**–4.00**	**<.0001**	**taːr$_{[\pm\text{lower}]}$-jɒ < tsaːr$_{[-\text{lower}]}$-jɒ**
C Manner (default: plosive)					
fricative	**–1.03**	**.44**	**–2.37**	**.0179**	**raːf-jɒ < laːp-jɒ**
sibilant	**–11.07**	**.80**	**–13.86**	**<.0001**	**hɒs-jɒ < lɒt-jɒ**
nasal	**–2.07**	**.28**	**–7.39**	**<.0001**	**oːn-jɒ < toːt-jɒ**
approximant	**–4.06**	**.31**	**–13.10**	**<.0001**	**tʃɛr-jɛ < tsɛt-jɛ**
C Place (default: alveolar)					
labial	**–2.22**	**.27**	**–8.35**	**<.0001**	**tsiːm-jɛ < siːn-jɛ**
palatal	**–9.25**	**1.13**	**–8.22**	**<.0001**	**ryɟ-jɛ < tʃyd-jɛ**
velar	**–3.54**	**.31**	**–11.55**	**<.0001**	**gøːg-jɛ < tʃøːd-jɛ**
V Height (default: mid)					
high	**1.85**	**.23**	**8.09**	**<.0001**	**cuːk-jɒ > tʃoːk-jɒ**
low	**0.77**	**.21**	**3.66**	**.0003**	**kɒr-jɒ > kor-jɒ**
Harmony (default: back)					
front	**–1.98**	**.27**	**–7.41**	**<.0001**	**køb-jɛ < dob-jɑ**
variable	**2.25**	**1.04**	**2.17**	**.0297**	**ɒnkeːt-jɒ/jɛ > klɒrineːt-jɒ**
Coda (default: singleton)					
geminate	**2.43**	**.41**	**5.97**	**<.0001**	**ʃikː-jɛ > sik-jɛ**
cluster	–0.08	.22	–0.36	.7161	domb-jɒ < dob-jɒ
V Length (default: short)					
long	**1.30**	**.19**	**6.97**	**<.0001**	**boːr-jɒ > bor-jɒ**
Syllables (default: monosyllabic)					
polysyllabic	**0.79**	**.18**	**4.31**	**<.0001**	**ɛlɛm-jɛ > sɛm-jɛ**

Table 17: Regression model with phonological and morphological predictors of possessive [j], with significant effects bolded

	β coef	*SE*	*Wald z*	*p*	*predicted rates*
Intercept	**3.72**	**.37**	**10.04**	**<.0001**	**bot-jɒ > bot-ɒ**
Stem class (default: non-lowering)					
lowering	**–3.66**	**.72**	**–5.12**	**<.0001**	**ɲaːr$_{[+lower]}$-jɒ < paːr$_{[-lower]}$-jɒ**
undetermined	**–0.67**	**.34**	**–1.97**	**.0490**	**(no direct comparison)**
variable	**–2.99**	**1.28**	**–2.34**	**.0194**	**taːr$_{[\pm lower]}$-jɒ < tsaːr$_{[-lower]}$-jɒ**
C Manner (default: plosive)					
fricative	0.30	.87	0.34	.7344	raːf-jɒ < laːp-jɒ
sibilant	**–11.22**	**.82**	**–13.75**	**<.0001**	**hɒs-jɒ < lɒt-jɒ**
nasal	**–2.30**	**.35**	**–6.50**	**<.0001**	**oːn-jɒ < toːt-jɒ**
approximant	**–4.55**	**.38**	**–11.99**	**<.0001**	**tʃɛr-jɛ < tsɛt-jɛ**
. . .					
Stem class * C Manner (default: non-lowering * plosive)					
. . .					
undetermined * approximant	**1.08**	**.39**	**2.74**	**.0062**	**(no direct comparison)**
. . .					

Table 18: Details from regression model with phonological and morphological predictors of possessive [j] and an interaction term betewen Stem class and C Manner, with significant effects bolded

	β coef	*SE*	*Wald z*	*p*	*predicted rates*
Intercept	**3.45**	**.35**	**9.83**	**<.0001**	**bot-jɒ > bot-ɒ**
Stem class (default: non-lowering)					
lowering	**–3.29**	**.49**	**–6.74**	**<.0001**	**ɲaːr$_{[+lower]}$-jɒ < paːr$_{[-lower]}$-jɒ**
undetermined	–0.07	.29	–0.25	.8025	(no direct comparison)
variable	**–2.55**	**.80**	**–3.19**	**.0014**	**taːr$_{[\pm lower]}$-jɒ < tsaːr$_{[-lower]}$-jɒ**
. . .					
C Place (default: alveolar)					
labial	**–1.80**	**.32**	**–5.62**	**<.0001**	**tsiːm-jɛ < siːn-jɛ**
palatal	**–9.12**	**1.14**	**–8.02**	**<.0001**	**ryɟ-jɛ < tʃyd-jɛ**
velar	**–3.54**	**.35**	**–10.04**	**<.0001**	**gøːg-jɛ < tʃøːd-jɛ**
. . .					
Stem class * C Place (default: non-lowering * alveolar)					
. . .					
undetermined * labial	**–0.97**	**.44**	**–2.21**	**.0268**	**(no direct comparison)**
. . .					

Table 19: Details from regression model with phonological and morphological predictors of possessive [j] and an interaction term betewen Stem class and C Place, with significant effects bolded

	β coef	*SE*	*Wald z*	*p*	*predicted rates*
Intercept	**3.73**	**.35**	**10.69**	**<.0001**	**bot-jɒ > bot-ɒ**
Stem class (default: non-lowering)					
lowering	**–3.04**	**1.01**	**–3.00**	**.0027**	**ɲaːr$_{[+\text{lower}]}$-jɒ < paːr$_{[-\text{lower}]}$-jɒ**
undetermined	**–0.76**	**.29**	**–2.64**	**.0083**	**(no direct comparison)**
variable	–6.68	66.32	–0.10	.9198	taːr$_{[\pm\text{lower}]}$-jɒ < tsaːr$_{[-\text{lower}]}$-jɒ
. . .					
V Height (default: mid)					
high	**1.12**	**.29**	**3.81**	**<.0001**	**cuːk-jɒ > tʃoːk-jɒ**
low	**0.56**	**.23**	**2.48**	**.0001**	**kɒr-jɒ > kor-jɒ**
. . .					
Stem class * V Height (default: non-lowering * mid)					
. . .					
undetermined * high	**1.66**	**.46**	**3.63**	**.0003**	**(no direct comparison)**
. . .					

Table 20: Details from regression model with phonological and morphological predictors of possessive [j] and an interaction term betewen Stem class and V Height, with significant effects bolded

	β coef	*SE*	*Wald z*	*p*	*predicted rates*
Intercept	**3.44**	**.33**	**10.29**	**<.0001**	**bot-jɒ > bot-ɒ**
Stem class (default: non-lowering)					
lowering	**–3.31**	**.49**	**–6.77**	**<.0001**	**ɲaːr$_{[+\text{lower}]}$-jɒ < paːr$_{[-\text{lower}]}$-jɒ**
undetermined	–0.40	.27	–1.52	.1288	(no direct comparison)
variable	**–2.06**	**.89**	**–2.61**	**.0090**	**taːr$_{[\pm\text{lower}]}$-jɒ < tsaːr$_{[-\text{lower}]}$-jɒ**
. . .					
Coda (default: singleton)					
geminate	**2.55**	**.72**	**3.53**	**.0004**	**ʃikː-jɛ > sik-jɛ**
cluster	–0.20	.26	–0.77	.4389	domb-jɒ < dob-jɒ
. . .					
Stem class * Coda (default: non-lowering * singleton)					
. . .					

Table 21: Details from regression model with phonological and morphological predictors of possessive [j] and an interaction term betewen Stem class and Coda, with significant effects bolded

	β coef	*SE*	*Wald z*	*p*	*predicted rates*
Intercept	**3.52**	**.34**	**10.44**	**<.0001**	**bot-jɒ > bot-ɒ**
Stem class (default: non-lowering)					
lowering	**–4.07**	**.63**	**–6.45**	**<.0001**	**ɲaːr$_{[+\text{lower}]}$-jɒ < paːr$_{[-\text{lower}]}$-jɒ**
undetermined	–0.08	.27	–0.28	.7824	(no direct comparison)
variable	**–2.99**	**1.08**	**–2.78**	**.0055**	**taːr$_{[\pm\text{lower}]}$-jɒ < tsaːr$_{[-\text{lower}]}$-jɒ**
...					
V Length (default: short)					
long	**1.49**	**.25**	**6.05**	**<.0001**	**boːr-jɒ > bor-jɒ**
...					
Stem class * V Length (default: non-lowering * short)					
...					

Table 22: Details from regression model with phonological and morphological predictors of possessive [j] and an interaction term betewen Stem class and V Length, with significant effects bolded

	β coef	*SE*	*Wald z*	*p*	*predicted rates*
Intercept	**3.40**	**.35**	**9.85**	**<.0001**	**bot-jɒ > bot-ɒ**
Stem class (default: non-lowering)					
lowering	**–3.38**	**.47**	**–7.24**	**<.0001**	**ɲaːr$_{[+\text{lower}]}$-jɒ < paːr$_{[-\text{lower}]}$-jɒ**
undetermined	–0.16	.31	–0.52	.6035	(no direct comparison)
variable	**–2.82**	**.85**	**–3.33**	**.0009**	**taːr$_{[\pm\text{lower}]}$-jɒ < tsaːr$_{[-\text{lower}]}$-jɒ**
...					
Syllables (default: monosyllabic)					
polysyllabic	**0.91**	**.22**	**4.04**	**<.0001**	**ɛlɛm-jɛ > sɛm-jɛ**
Stem class * V Length (default: non-lowering * short)					
...					

Table 23: Details from regression model with phonological and morphological predictors of possessive [j] and an interaction term betewen Stem class and Syllables, with significant effects bolded

phonological factor in interaction term	χ^2	*p*	$\Delta(AIC)$
C Manner	19.69	.0731	–4.31
C Place	11.72	.2295	–6.28
V Height	16.53	.0112	4.53
Harmony	(model doesn't converge)		
Coda	14.54	.0242	2.54
V Length	3.85	.2785	–2.15
Syllables	4.96	.1750	–1.04

Table 24: Comparison between base model with stem class and respective regression models including a given interaction term with stem class and a phonological factor

B Appendix: Phonotactic sublexical grammars

This appendix contains the grammars induced by the UCLA Phonotactic Learner Hayes and Wilson (2008) as described in Section 4.2 (including dummy [+sublex] segments at the beginning of each word). First I list the constraints for the baseline phonotactic grammar induced on all words in my data set, then those for the [+j] and [−j] sublexical grammars induced on words that take possessive *-jV* and *-V*, respectively. For the latter two, I include for each constraint both the weight and the residual weight, which I define as the difference between a constraint's weight in a sublexicon and its weight in the baseline phonotactic grammar. In each case, I set the basic parameters as: maximum 40 constraints, maximum 3-gram, maximum OE for constraints 3. In any constraints against trigrams, (at least) one of the segments had to be a word boundary. Segments marked [+syllabic] were projected onto a Vowel tier allowing trigram constraints and including the features [±high], [±low], [±back], and word boundaries. Segments marked [+sublex] were projected onto a Sublex tier allowing only unigram constraints and including the features [±lower] annd word boundaries.

In the Phonotactic Learner, words are padded with [+word_boundary] segments, while all phonological segments are given the feature [−word_boundary]. Here I indicate [+word_boundary] segments as # and [−word_boundary] segments as [+segment]. However, in my implementation, the first segment of the word follows the [+sublex] dummy segment, so I refer to the environment following [+sublex] as word-initial as well.

constraint		*description*	*weight*
*#[−sublex]		no word-initial non-dummy segments	8.414
*[+segment][+sublex]		no non-word-initial dummy segments	6.229
*[−syllabic][+long,−syllabic]		no geminates after consonants	5.361
*[−back,+round,+long]		no long front rounded vowels	5.273
*[+sublex][+long,−syllabic]		no word-initial geminates	4.693
*[+long,+syllabic][+syllabic]		no long vowels before vowels	4.552
*[+long,−syllabic][−syllabic]		no geminates before consonants	4.457
*##	(Vowel tier)	no vowelless words	3.989
*[+voice][−voice]		no mixed-voice (voiced–voiceless) clusters	3.563
*[−cont,+voice,+strident]		no voiced affricates	3.548
*[+voice,+long]		no voiced geminates	3.426
*[+long][+long,−syllabic]		no geminates after long segments	3.373
*[+glottal,+long]		*hː	3.346
*[+nasal][+nasal]		no adjacent nasals	3.173
*[−lateral,+long]		*rː	3.158
*[−son,−cont,+coronal][−son,+coronal]		specific sequence constraint	3.066
*[−cont,−anterior][+coronal]		specific sequence constraint	3.057
*[−son,+cont][+strident]		specific sequence constraint	3.055
*[−son,−anterior][+cont,−voice]		specific sequence constraint	3.025
*[−anterior,+long][−long]		specific sequence constraint	3.000
*[−high][−back]		no mid or low vowels before front vowels	2.837
*[−cont,+coronal][+cont,−anterior]		specific sequence constraint	2.822
*[−son,+labial][+labial]		no labial obstruents before labials	2.792
*[+cont,−anterior,+long]		*sː, *ʒː, *jː	2.732
*[−cont,−voice][−cont,+labial]		specific sequence constraint	2.671
*[−cont,+long][−back]		specific sequence constraint	2.563
*[−low][−low]		no adjacent non-low vowels	2.532
*[+labial][+nasal]		no labials before nasals	2.529
*[−cont,−anterior][−cont,+voice]		specific sequence constraint	2.467
*[+labial][−anterior]		no post-alveolars or palatals after labials	2.459
*[−son][+voice]		no obstruents before voiced consonants	2.300
*[−son,−cont][−cont,−anterior]		specific sequence constraint	2.298
*[+high,−back][+high,−back]#	(Vowel tier)	vowel harmony constraint	2.231
*[+syllabic][−back,+round]		no front rounded vowels after vowels	2.040
*[−anterior][−anterior]		no adjacent palatals or post-alveolars	2.005
*[+long][−son,+long]		no obstruent geminates after long segments	2.002
*[−son,+coronal][−anterior]		specific sequence constraint	1.711
*[−voice][+voice,+strident]		no voiced sibilants after unvoiced consonants	1.367
*[+voice,+strident][−voice]		no voiced sibilants before unvoiced consonants	1.095
*[+strident][+strident]		no adjacent sibilants	0.781

Table 25: Baseline grammar induced by the UCLA Phonotactic Learner (Hayes and Wilson, 2008) on the full data set

constraint		*description*	*weight*	*residual weight*
*#[−sublex]		no word-initial non-dummy segments	8.498	0.084
*[+segment][+sublex]		no non-word-initial dummy segments	5.601	−0.628
*[−syllabic][+long,−syllabic]		no geminates after consonants	5.317	−0.044
*[−back,+round,+long]		no long front rounded vowels	5.042	−0.231
*[+sublex][+long,−syllabic]		no word-initial geminates	4.510	−0.183
*[+lower]	(Sublex tier)	no lowering stems	4.330	4.330
*[−anterior]#		no word-final palatals or post-alveolars	4.285	4.285
*[+long,−syllabic][−syllabic]		no geminates before consonants	4.268	−0.189
*[+long][+long,−syllabic]		no geminates after long segments	4.238	0.865
*[+long,+syllabic][+syllabic]		no long vowels before vowels	4.195	−0.357
*[+strident]#		no word-final sibilants	3.914	3.914
*##	(Vowel tier)	no vowelless words	3.719	−0.270
*[+voice,+long]		no voiced geminates	3.613	0.187
*[+voice][−voice]		no mixed-voice (voiced–voiceless) clusters	3.371	−0.192
*[−son,+labial][+labial]		no labial obstruents before labials	3.308	0.516
*[−lateral,+long]		*rː	3.166	0.008
*[+glottal,+long]		*hː	3.160	−0.186
*[−cont,−anterior][+coronal]		specific sequence constraint	3.040	−0.017
*[−cont,+voice,+strident]		no voiced affricates	2.908	−0.640
*[+labial][−cont,−anterior]		specific sequence constraint	2.896	2.896
*[+syllabic][−low]		no mid or high vowels after vowels	2.750	2.750
*[−son,+cont][+strident]		specific sequence constraint	2.746	−0.309
*[−anterior,+long][−long]		specific sequence constraint	2.601	0.399
*[−son,−cont,+coronal][−son,+coronal]		specific sequence constraint	2.600	−0.466
*[−son,−anterior][+cont,−voice]		specific sequence constraint	2.592	−0.433
*[−cont,+strident,+long]		no geminate affricates	2.583	2.583
*[−cont,−anterior][+labial]		specific sequence constraint	2.570	2.570
*[−son][+voice]		no obstruents before voiced consonants	2.451	0.151
*[−son,−cont,+coronal][−ant]		specific sequence constraint	2.376	2.376
*[+high,−back][+high,−back]#	(Vowel tier)	vowel harmony constraint	2.286	0.055
*[−anterior][−anterior]		no adjacent palatals or post-alveolars	2.081	0.076
*[+low,−back][−low,−back]	(Vowel tier)	vowel harmony constrant	2.009	2.009
*[−cont,−anterior][−cont,+voice]		specific sequence constraint	1.970	−0.497
*[+son,−anterior,+long]		*jː, *ɲː	1.952	1.952
*[−son,−cont][−cont,−anterior]		specific sequence constraint	1.897	−0.401
*[+cont,−anterior,+long]		*ʃː, *ʒː, *jː	1.523	−1.209
*[−anterior,+strident,+long]		no geminate post-alveolars	1.324	1.324
*[+cont,+voice][−voice]		no voiced fricatives before unvoiced consonants	1.287	1.287
*[−cont,+voice,+strident][−syll]		no voiced affricates before consonants	1.130	1.130
*[−son,+coronal][+strident]		no sibilants after coronal obstruents	0.779	0.779

Table 26: Sublexical grammar induced by the UCLA Phonotactic Learner (Hayes and Wilson, 2008) on the [+j] sublexicon

constraint		*description*	*weight*	*residual weight*
*#[−sublex]		no word-initial non-dummy segments	7.153	−1.261
*[+segment][+sublex]		no non-word-initial dummy segments	5.301	−0.928
*[−back,+round,+long]		no long front rounded vowels	4.150	−1.123
*[−syllabic][+long,−syllabic]		no geminates after consonants	3.980	−1.381
*[+syllabic]#		no word-final vowels	3.718	3.718
*##	(Vowel tier)	no vowelless words	3.583	−0.406
*[+long,−syllabic][−syllabic]		no geminates before consonants	3.557	−0.900
*[−cont,+voice,+strident]		no voiced affricates	3.489	−0.059
*[+labial,+long]		no geminate labials	3.325	3.325
*[+syllabic,+syllabic]		no adjacent vowels	2.999	2.999
*[+long][+long]		no adjacent long segments	2.868	2.868
*[−son,+coronal][+cont,−voice]		specific sequence constraint	2.768	2.768
*[−cont,−anterior][+coronal]		specific sequence constraint	2.730	−0.327
*[−son,+cont][−anterior]		specific sequence constraint	2.652	2.652
*[+glottal,+long]		*hː	2.597	−0.750
*[−ant,+long][+segment]		no non-final geminate palatals or post-alveolars	2.570	2.570
*[−voice][+voice]		no mixed-voice (voiceless–voiced) clusters	2.554	2.554
*[−nasal,−ant,−strident][+high,−back]		specific sequence constraint	2.478	2.478
*[+voice][−son]		no voiced consonants before obstruents	2.473	2.473
*[−son,−cont][−son,−cont,+labial]		specific sequence constraint	2.448	2.448
*[−lateral,+long]		*rː	2.404	−0.754
*[−cont,+strident][+son]		no affricates before sonorants	2.403	2.403
*[+voice,+strident][+cont]		specific sequence constraint	2.378	2.378
*[+nasal][+nasal]		no adjacent nasals	2.361	−0.812
*[−son,+cont][+strident]		specific sequence constraint	2.336	−0.719
*[−son,+long][+high]		no geminate obstruents before high vowels	2.330	2.330
*[−cont,+anterior,−strident,+long]		*tsː, *dzː	2.295	2.295
*[+voice,+long]		no voiced geminates	2.266	−1.160
*[−back,+round][−nasal,+anterior,+long]		specific sequence constraint	2.258	2.258
*[−son,+labial][+labial]		no labial obstruents before labials	2.215	−0.577
*[+labial][+nasal]		no labials before nasals	2.212	−0.317
*[+voice,−anterior][−low,+long]		specific sequence constraint	2.183	2.183
*[−son,+coronal][−anterior]		specific sequence constraint	2.124	0.413
*[−anterior][−cont,−anterior]		specific sequence constraint	2.082	2.082
*[−son,−cont][−cont,+coronal]		specific sequence constraint	2.081	2.081
*[−low][+high,+back]	(Vowel tier)	specific vowel constrant	1.978	1.978
*[−high]#		no word-final low or mid vowels	1.727	1.727
*[−cont,+coronal,+long][+segment]		specific segment constrant	1.581	1.581
*[+voice,+strident,+long]		no voiced geminate sibilants	1.333	1.333
*[−son,+coronal][+strident]		no sibilants after coronal obstruents	1.193	1.193

Table 27: Sublexical grammar induced by the UCLA Phonotactic Learner (Hayes and Wilson, 2008) on the [−j] sublexicon

References

Ackerman, F., Blevins, J. P., and Malouf, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. In Blevins, J. P. and Blevins, J., editors, *Analogy in Grammar: Form and Acquisition*, chapter 3, pages 54–82. Oxford University Press, Oxford.

Ackerman, F. and Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. *Language*, 89(3):429–464.

Albright, A. and Hayes, B. (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition*, 90(2):119–161.

Bárczi, G. and Országh, L., editors (1959–1962). *A magyar nyelv értelmező szótára*, volume 1–7. Akadémiai Kiadó, Budapest.

Baunaz, L. and Lander, E. (2018). Nanosyntax: The basics. In Baunaz, L., De Clercq, K., Haegeman, L., and Lander, E., editors, *Exploring Nanosyntax*, Oxford Studies in Comparative Syntax, chapter 1, pages 3–56. Oxford University Press, New York.

Bobaljik, J. D. (2000). The ins and outs of contextual allomorphy. In Grohmann, K. and Struijke, C., editors, *University of Maryland Working Papers in Linguistics*, volume 10, pages 35–71, College Park. University of Maryland.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. Studies in Language. Harper & Row, New York, Evanston, and London.

Corbett, G. G. and Fraser, N. M. (1993). Network Morphology: a DATR account of Russian nominal inflection. *Journal of Linguistics*, 29(1):113–142.

É. Kiss, K. (2002). *The Syntax of Hungarian*. Cambridge Syntax Guides. Cambridge University Press, Cambridge.

Embick, D. (2010). *Localism versus Globalism in Morphology and Phonology*, volume 60 of *Linguistic Inquiry Monographs*. MIT Press, Cambridge.

Embick, D. and Marantz, A. (2008). Architecture and blocking. *Linguistic Inquiry*, 39(1):1–53.

Gouskova, M., Newlin-Łukowicz, L., and Kasyanenko, S. (2015). Selectional restrictions as phonotactics over sublexicons. *Lingua*, 167:41–81.

Guzmán Naranjo, M. (2019). *Analogical classification in formal grammar*, volume 5 of *Empirically Oriented Theoretical Morphology and Syntax*. Language Science Press, Berlin.

Halle, M. and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In Hale, K. and Keyser, S. J., editors, *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, volume 24 of *Current Studies in Linguistics*, chapter 3, pages 111–176. MIT Press, Cambridge.

Harrell Jr., F. E. (2020). *rms: Regression Modeling Strategies*. R package version 6.1-0.

Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39(3):379–440.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*, volume 103 of *Springer Texts in Statistics*. Springer, New York.

Kiefer, F. (1985). The possessive in Hungarian: A problem for natural morphology. *Acta Linguistica Academiae Scientiarum Hungaricae*, 35(1–2):85–116.

Kramer, R. (2016). Syncretism in paradigm function morphology and distributed morphology. In Siddiqi, D. and Harley, H., editors, *Morphological Metatheory*, volume 229 of *Linguistic Aktuell/Linguistics Today*, pages 95–120. John Benjamins, Amsterdam / Philadelphia.

MacWhinney, B. (1975). Rules, rote, and analogy in morphological formations by Hungarian children. *Journal of Child Language*, 2:65–77.

Myler, N. (2017). Exceptions to the Mirror Principle and morphophonological 'action at a distance': The role of 'word'-internal phrasal movement and spell-out. In Newell, H., Noonan, M., Piggott, G., and Travis, L. d., editors, *The Structure of Words at the Interfaces*, volume 68 of *Oxford Studies in Theoretical Linguistics*, chapter 5, pages 100–125. Oxford University Press, Oxford.

Papp, F. (1969). *A magyar nyelv szóvégmutató szótára*. Akadémiai Kiadó, Budapest.

Parker, J. and Sims, A. D. (2020). Irregularity, paradigmatic layers, and the complexity of inflection class systems: A study of Russian nouns. In Arkadiev, P. and Gardani, F., editors, *The Complexities of Morphology*, chapter 2, pages 23–51. Oxford University Press, Oxford.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Rácz, P. and Rebrus, P. (2012). Variation in the possessive allomorphy of Hungarian. In Kiefer, F., Ladányi, M., and Siptár, P., editors, *Current Issues in Morphological Theory: (Ir)regularity, analogy and frequency: Selected papers from the 14th International Morphology Meeting, Budapest, 13–16 May 2010*, volume 322 of *Current Issues in Linguistic Theory*, pages 51–64, Amsterdam/Philadelphia. John Benjamins.

Rebrus, P. (2013). Miért nincs *j*? In Benő, A., Fazakas, E., and Kádár, E., editors, *„. . .hogy legyen a víznek lefolyása. . . ": Köszöntő kötet Szilágyi N. Sándor tiszteletére*, pages 383–401. Erdélyi Múzeum-Egyesület, Kolozsvár [Cluj-Napoca].

Rebrus, P. and Szigetvári, P. (2018). Morphosyntactically conditioned morphophonological variation. Presentation at PhonVar 1, Barcelona.

Rebrus, P., Szigetvári, P., and Törkenczy, M. (2017). Asymmetric variation. In Lindsey, G. and Nevins, A., editors, *Sonic Signatures: Studies dedicated to John Harris*, volume 14 of *Language Faculty and Beyond*, pages 164–187. John Benjamins, Amsterdam / Philadelphia.

Siptár, P. and Törkenczy, M. (2000). *The Phonology of Hungarian*. The Phonology of the World's Languages. Oxford University Press, Oxford, New York.

Stump, G. (2001). *Inflectional Morphology: A Theory of Paradigm Structure*, volume 93 of *Cambridge Studies in Linguistics*. Cambridge University Press, Cambridge.

Stump, G. (2016). Paradigms at the interface of a lexeme's syntax and semantics with its inflectional morphology. In Siddiqi, D. and Harley, H., editors, *Morphological Metatheory*, volume 229 of *Linguistic Aktuell/Linguistics Today*, pages 27–58. John Benjamins, Amsterdam / Philadelphia.

Stump, G. (2019). Paradigm Function Morphology. In Audring, J. and Masini, F., editors, *The Oxford Handbook of Morphological Theory*, Oxford Handbooks in Linguistics, chapter 14, pages 285–304. Oxford University Press, Oxford.

Szabolcsi, A. (1994). The noun phrase. In Kiefer, F. and É. Kiss, K., editors, *The Syntactic Structure of Hungarian*, volume 27 of *Syntax and Semantics*, pages 179–274. Academic Press, San Diego / New York.

Zaliznjak, A. A. (1977). *Grammatičeskij slovar' russkogo jazyka*. Russkij Jazyk, Moscow.