Selectional restrictions as phonotactics over sublexicons[*]

Maria Gouskova and Luiza Newlin-Łukowicz (New York University)

October 16, 2014

Abstract

Affixation and allomorphy are often phonologically predictable: thus, the English indefinite "a" appears before consonants, and "an" before vowels. We propose a theory of phonological selection that separates rules of morpheme realization from phonological knowledge about the bases and the derived words. This phonological knowledge is encoded in miniature phonotactic grammars, which are learned over sublexicons defined by morphological generalizations. Each sublexical phonotactic grammar determines the likelihood that a new word will follow the associated rule. We examine a complex case of suppletive allomorphy in Russian, whose diminutive suffixes define sublexicons differing in constraints on final consonant place and manner, presence and location of consonant clusters, vowel hiatus, and stress. In elicitation, Russians choose allomorphs for words without diminutives based on how these words and the derived diminutives fare in the sublexical phonotactic grammars. This approach handles aspects of the pattern that are missed by alternative approaches such as emergence of the unmarked, insertion rules that refer directly to phonological information, and the Minimal Generalization Learner.

1 Introduction

Affixes sometimes impose phonological selectional restrictions on bases (Siegel 1974, Carstairs 1988), requiring them to have certain phonological properties for affixation. For example, in English, the comparative and superlative suffixes attach to monosyllabic words (*smart-er*) and some disyllabic ones (*silli-er*), but not trisyllabic or longer words (**intelligent-er*). In Russian, masculine nouns can form diminutives with three allomorphs, each of which imposes a host of phonological selectional restrictions on the location of stress, place and manner of articulation of the last consonant, and the presence and location of consonant clusters—to name a few:

(1) Russian diminutive allomorphy, in brief

	Unaffixed	Diminutive	Gloss	Some restrictions imposed by suffix:
a.	áng^{j}il	ang^{j}il-ók	'angel'	*stress-initial /stress-final base, no vowel hiatus, no CC#*
b.	mónstr	mónstr-^{j}ik	'monster'	*stress-final base, no final dorsal consonants*
c.	ajfón	ajfón-ʧ^{j}ik	'iPhone'	*stress-final base, no clusters, no final coronal obstruents*

[*]We would like to thank Michael Becker, Ryan Bennett, Stuart Davis, Kenneth de Jong, Steven Franks, Gillian Gallagher, Stephanie Harves, Junko Ito, Sofya Kasyanenko, Jeremy Kuhn, Sang-Im Lee, Alec Marantz, Elliott Moreton, Neil Myler, Jaye Padgett, Katya Pertsova, Jen Smith, Jonathan North Washington, Matt Wolf, and audiences at NYU, UNC Chapel Hill, UC Santa Cruz, the 22nd Formal Approaches to Slavic Linguistics conference in Hamilton, Ontario, and the Annual Meeting on Phonology 2013 at UMass Amherst. Thanks also to Anna Aristova for helping with participant recruitment, and to our Russian participants. Our work was supported in part by NSF grant BCS-1224652.

There is a debate in the morphological and phonological literature as to how such restrictions should be stated in the grammar and what people actually know about phonological conditions on affixation. We propose that these restrictions are learned from the lexicon, much as in the influential proposals by Albright and Hayes (2003) and Hayes and Wilson (2008), and that the phonological generalizations about affixation arise from phonotactic learning over *sublexicons* defined by morphological operations (Becker and Gouskova 2012).

We argue that selectional restrictions are encoded diacritically in morphological realization rules. Learners figure out which morphemes are marked to combine with which allomorphs; once the sublexicons are assembled, people learn sublexical phonotactic grammars that characterize each sublexicon, but the phonological information is not encoded in the insertion rules themselves. Thus, we assume that morphological realization is formally separated from phonological knowledge—but phonological knowledge about the content of sublexicons can be called upon in certain situations. For example, nonce words and loanwords are not indexed for diminutive formation or any other morphological rules, but people can consult the sublexical phonotactic grammars to decide on the probability that the item will receive a particular diacritic (this can be modeled as grammar inference; see Becker and Gouskova 2012, Albright 2008).

We argue that both the shape of the base and the shape of the affixed word inform sublexical phonotactic generalizations. In the terminology of Bybee (1995), both source and product-oriented generalizations matter. This has been a matter of some controversy, since the morphological literature has largely focused on source-oriented generalizations about selectional restrictions (Bobaljik 2000, Paster 2006, Embick 2010), whereas the phonological literature, especially constraint-based frameworks, focus on product-oriented generalizations (Kager 1996, Mascaró 2007, Bye 2007; see Nevins 2011 for an overview). We demonstrate that both play a role in the Russian case as well as in other cases; the key to our solution to this contradiction is that rules refer to diacritics only, but learners keep track of information about both base and affixed sublexicons in deciding which diacritics to put on new items.

The case of Russian diminutive formation provides some important evidence for our theory. First, the phonotactic properties of bases that predict diminutive allomorphs do not need to be local to the suffixes themselves. For example, the presence of vowel hiatus in the base diminishes its likelihood of combining with [-ok]: thus, words such as [klóun] 'clown' do not diminutivize easily even though the hiatus sequence is not phonologically local to the [-ok] suffix. We identify a similar effect of medial consonant clusters, which highlights another aspect of selectional restrictions that our theory captures: restrictions are often orthogonal to the phonology of the affixes and cannot be reduced to markedness-based explanations that attribute the failure of affixation to the phonological ill-formedness of the resulting word. We argue that this is because restrictions are formed over bases, not affixed words.

Our theory can be distinguished from alternative approaches to selectional restrictions, which include subcategorization frames, Generalized Alignment, and Emergence of the Unmarked. We argue that these theories get some aspects of phonological selectional restrictions right, but they are not sufficiently powerful for cases such as Russian diminutive allomorphy, which involves both restrictions on the site of affixation and less local properties of the word. The work on the typology of phonologically conditioned suppletive allomorphy has established conclusively that selectional restrictions of allomorphs cannot always be reduced to universal phonological markedess (Paster 2006, Bobaljik 2000, Embick 2010, Nevins 2011), but we suggest they can be reduced to phonotactic constraints over word shapes, which often have a language-specific character.

The paper is organized as follows. We start by presenting the proposal in section 2. We next describe Russian diminutive formation (section 3) and an analysis of the Russian lexicon and the sublexicons formed by bases and diminutives (section 4). The elicitation study is described in section

(5); our interpretation of the results and the alternatives to our proposal are discussed in section 6.

2 Proposal: phonotactic grammars for morphological sublexicons

2.1 Assumptions about selection and morpheme insertion

This section outlines the assumptions about morphological theory that underlie several features of our proposal. We assume a realizational theory of morphology (Distributed Morphology, Halle and Marantz 1993, Halle 1994, Embick and Marantz 2008, inter alia). According to this theory, phonological and morphosyntactic information is specified not for lexemes or words but for morphemes. For us, a morpheme is an abstract association between semantic representations, morphosyntactic features, and phonological representations. The morphosyntactic features are manipulated by the syntax, which supplies the inputs to morphological realization rules. These realization rules specify which (allo)morphs are inserted in which contexts. We make three assumptions about the morphology-phonology interface that we explain in more detail here: first, realization rules can refer to diacritics on inner but not outer morphs. Second, realization rules are stated on morphemes, and they refer to diacritic features rather than directly to phonological information. Third, diacritic features are discovered in the process of learning the lexicon, when phonological words containing an affix are associated with related phonological words that lack the affix. Any phonological generalizations over words that combine with an affix are learned from these phonological words.

2.1.1 Morpheme insertion follows scope

We assume that realization rules apply to syntactic trees in the reverse order of scope, from the most deeply embedded node outward (Bobaljik 2000, Wolf 2008, Myler to appear). Inside-out insertion captures a typological observation: phonologically conditioned suppletive allomorphy is conditioned by inner morphemes, not by outer ones. Thus, suppletive allomorphy of affixes, clitics, and function words such as determiners is quite common cross-linguistically (e.g., “a” vs. “an” in English), and the choice of allomorph can be determined by the phonological properties of the base. On the other hand, root suppletion, such as “person”/“people” or “good”/“better”, is only ever conditioned by the outer morphosyntactic context (Bobaljik 2000). Phonologically conditioned suppletive allomorphy of stems is unattested. The explanation for this typological gap in Distributed Morphology is that the stem is realized before the affix: the choice between two stem allomorphs is decided before the exact phonological shape of the affix is known, then the phonological shape of the affix cannot condition the stem allomorphy.

We illustrate these two types of allomorphy in (2) and (3). The input to the morphology is a morphosyntactic hierarchical structure; the lexical realization rules specify how to fill the structures out. Insertion starts with the most deeply embedded node—the root. The Russian roots in (2) have phonologically invariant realization (see (2a–b)), but the diminutive in Russian is realized differently depending on the identity of the root (see (2c–d)). The context is specified in the realization rule as a diacritic feature, so any root that is marked as a *+ik* morpheme will condition the diminutive allomorph [-^{j}ik], and any root that is marked as *+ok* will condition the realization of [-ok]. On the other hand, the English root in (3) has two realizations, sensitive to the external syntactic context stated as features (plural vs. singular; see (3a–b)). The plural in English also has several allomorphs, which include the null marking that appears on nouns such as “sheep” and “goose” and as [-s] on other nouns (among others). Once the phonological shape of the root is known, its diacritic determines which plural allomorph is inserted—in the case of “people”, the plural is null. Thus, in the case of Russian affix allomorphy, the phonological realization of the roots determines that of the diminutive allomorphs. In English, on the other hand, the syntactic context determines the realization of the roots, which then in turn determine the realization of the plural allomorphs.

(2) Russian: phonologically conditioned affix allomorphy

	Russian: "ik"/"ok"	
Input:	[N [N $\sqrt{\text{HOUSE}}$ N] DIM]	[N [N $\sqrt{\text{SON}}$ N] DIM]
Root:	[N [N dom N] DIM]	[N [N sin N] DIM]
Aff 1:	[N [N dom ∅] DIM]	[N [N sin ∅] DIM]
Aff 2:	[N [N dom ∅] ʲik]	[N [N sin ∅] ok]
To phonology:	/dom-ʲik/	/sin-ok/

Realization rules for Russian:

a. $\sqrt{\text{HOUSE}}$↔/dom/

b. $\sqrt{\text{SON}}$↔/sin/

c. DIM↔/-ʲik/ [+ik]
[+ik] | {dom$_{+\text{ik}}$, stolb$_{+\text{ik}}$, ...}

d. DIM↔/-ok/ [+ok]
[+ok] | {sin$_{+\text{ok}}$, pirog$_{+\text{ok}}$, ...}

e. N↔∅

(3) English: syntactically conditioned stem allomorphy

	English "person"/"people"	
Input:	[NumP [N $\sqrt{\text{PERSON}}$ N] PL]	[NumP [N $\sqrt{\text{PERSON}}$ N] SG]
Root:	[NumP [N pipl N] PL]	[NumP [N pɝsn N] SG]
Aff 1:	[NumP [N pipl ∅] PL]	[NumP [N pɝsn ∅] SG]
Aff 2:	[NumP [N pipl ∅] ∅]	[NumP [N pɝsn ∅] ∅]
To phonology:	/pipl/	/pɝsn/

Realization rules for English:

a. $\sqrt{\text{PERSON}}$↔/pipl/ /__PL

b. $\sqrt{\text{PERSON}}$↔/pɝsn/

c. PL ↔[-s]/ __[+s]
[+s] | | {paɪp$_{+\text{s}}$, sɑk, ...}

d. PL ↔∅/ __
+∅
+∅ | {ʃip, gus, pipl, ...}

e. N↔∅

f. SG ↔∅

As Bobaljik (2000) argues, it is not possible for allomorphy to be phonologically conditioned by the outer context—for example, it would not be possible for suppletive allomorphy of stems such as *go/went* to be conditioned by whether the following suffix begins in a vowel or a consonant, as in **go-ed* vs. **went-ing* (though see Wolf 2013 for a version of this approach that does allow for some limited outward conditioning). The architecture we assume precludes this possibility, but some theories of phonological conditions on allomorphy allow for it (see section 6.2.1). The well-formedness of the derived word can determine the likelihood of an allomorph being chosen for a novel item, however. We discuss this next.

2.1.2 Morpheme insertion rules refer to diacritic features

In our analysis of (2) and (3), the conditioning of affix allomorphs is never really phonological. Rather, it is lexical: the rules refer to diacritic features of the root, not to its phonological properties. We assume that in all cases of phonologically conditioned allomorphy, the phonological conditioning is indirect: the lexical representations for affixes include information about which morphemes they appear with, but not the full list of all the phonological properties that these morphemes share. Even for cases where the insertion rule could refer to a simple phonological context, actual usage is often variable and complex. For example, the English indefinite is supposed to be realized as “a” before consonant-initial words and as “an” before vowel-final words, but there are exceptions in both directions (which we discuss shortly). We would formulate the rule for “a” insertion as referring to a list of morphemes that share the diacritic [+A], as in (4). The phonological generalizations are extracted separately.

(4) Our theory: insertion rules refer to diacritics

INDEF $\leftrightarrow$ ə /___ [+A] | {kæt_{+A}, dɑg_{+A}, teɪbl_{+A}, ...}

(5) Insertion rule that refers to phonological information:

INDEF$\leftrightarrow$ ə / ___ [+consonantal]

The more usual alternative is (5), where the insertion rule refers to phonological features of the context. But in many cases of allomorphy, the list of conditioning factors can be quite large, is often riddled with exceptions, or defines properties that can only be stated as negative constraints that cannot be elegantly summarized in a simple rule (see Wolf 2008 pp. 106–107). We are particularly concerned with cases that do not lend themselves to the elegant treatment in (5) because no one phonological property fully determines allomorph choice. Our analysis works uniformly for both simple and complex types of phonological conditioning: the formal rules refer to diacritics, and the phonological generalizations are extracted through phonotactic learning.

We exemplify this approach first on the case of English *a/an*. We assume that the learner detects an inconsistency in the realization of the indefinite article.[1] This morphological inconsistency triggers the creation of lists of phonological words that each allomorph occurs with—lists that we call *sublexicons*. Eventually, the learner posits diacritic rules that refer to morphemes adjacent

[1] We assume that there is a mechanism for identifying the lack of a language-wide phonotactic motivation for the suppletive alternation of the indefinite. Cf. the [ði]/[ðə] alternation of ‘the’, which is apparently phonotactically motivated by the presence and height of the following vowel (see Raymond et al. 2002), and [ɪn] ‘in’, which does not alternate in the same context where the indefinite does alternate. In such cases, no inconsistency is going to arise for the learner; the alternations can be ascribed to phonotactics and learned from surface distributional patterns (Peperkamp et al. 2006, Calamaro and Jarosz 2014, and others).

to the allomorphs (as shown in (6) and (7)), but the learning procedure starts with phonological words. The sublexicons for each allomorph are quite large, since the indefinite is fully productive in English and can cliticize onto any NP-initial phonological word; as we show in the next section, these large sublexicons are almost as diverse as English words in general—except that in the "A" base sublexicon, the words tend to not begin in a vowel, and in the "AN" base sublexicon, they tend to not begin in a consonant. The size and diversity will allow the learner to find phonological generalizations about the shapes of words that combine with each allomorph.

(6) INDEF ↔ə /___ [A]

a. Sublexicon of bases for [A]: {kæt$_{\mathrm{A}}$, dɑg$_{\mathrm{A}}$, teɪ.bl$_{\mathrm{A}}$, bɹæt$_{\mathrm{A}}$, ...}

b. Sublexicon of derived [A] forms: {ə.kæt, ə.dɑg, ə.teɪ.bl, ...}

(7) INDEF ↔ən /___ [AN]

a. Sublexicon of bases for [AN]: {æ.pl$_{\mathrm{AN}}$, o.ɹənʤ$_{\mathrm{AN}}$, i.gl$_{\mathrm{AN}}$, ...}

b. Sublexicon of derived [AN] forms: {ʔə.næ.pl, ʔən.o.ɹænʤ, ən.ʔi.gl, ...}

Separating the formal statement of the rule from the phonological generalizations is necessary to capture variability. Even though "a"/"an" is supposed to be a textbook case of phonologically conditioned allomorphy, there are exceptions to the "consonant/vowel" pattern in some dialects, which generalize the use of [ə] to vowel-initial contexts for at least some words (Raymond et al. 2002, Gabrielatos et al. 2010). Conversely, for some speakers, [ən] is exceptionally conditioned when the first syllable is unstressed and begins with [h], as in "[ən] historical". Our realization rules do not directly encode these phonological factors—rather, the phonotactic generalizations are encoded and learned separately. The learning input can be variable ("[ə]∼[ən] historical"), and it can contain exceptions to phonological generalizations ("[ə] horrendous" but "[ən] historical"). In the case of variation, the item will appear in both sublexicons, and the choice of suppletive allmorph will be determined probabilistically whenever the grammar applies (see Coetzee and Pater 2011, Coetzee and Kawahara 2013). The same is true for generalizations over the derived sublexicons: work on this alternation shows that the nasal of [ən] is not always syllabified with the following phonological word; the glottal stop is variably present but is actually more likely to be there after function words in English (Garellek 2012). The primary effect of both exceptions and variation is that they weaken phonotactic generalizations over the sublexicons.

2.1.3 Learning phonological generalizations over sublexicons

Language learning starts with general phonotactics. We assume that the phonotactic generalizations are learned as in Hayes and Wilson's (2008) proposal:[2] the learner tracks the co-occurrence probabilities of certain natural classes in various positions and extracts constraints against un- and under-attested sequences. The constraints are weighted (using a Maximum Entropy procedure, Goldwater and Johnson 2003) rather than ranked, and we will adopt Hayes and Wilson's assumption that the phonotactic grammars contain markedness alone. For example, the learner will notice that while [ŋ] is allowed in English in medial and word-final position, English phonological words do not begin in [ŋ. The learner will posit the constraint *[ŋ and give it a weight proportional to the evidence supporting the constraint. The novel aspect of our proposal is that phonotactic learning is restarted for each sublexicon, in order to identify which structures are un- or under-attested in each

[2]There are several approaches to phonotactic learning, such as Albright (2009a) and Adriaans and Kager (2010), that could be compatible with our proposal. We pick Hayes and Wilson's proposal because their computational implementation is particularly convenient to use on our phonological word-based sublexicons.

morphologically defined sublexicon. The idea that grammars contain constraints stated over subsets of the lexicon is of course not itself novel—it goes back at least to Morpheme Structure Constraints (Chomsky and Halle 1968, Booij 2011). What is new about our proposal is that we reduce both morphophonological knowledge (usually thought to involve rules) and phonotactic knowledge to the same mechanism.

In the unmarked case,[3] the constraints within each sublexicon will be more restrictive than the constraints of the language in general. Thus, the learner has to notice which structures are banned in each sublexicon even if they are allowed in the language as a whole. For example, in English, both vowel- and consonant-initial words are allowed, and in Russian, words can end in dorsal obstruents. But in the sublexicon of bases for the indefinite allomorph [ə], vowel-initial words are under-attested; likewise, in the sublexicon of bases for the Russian diminutive allomorph [-ik], there are virtually no dorsal-final bases. The phonotactic learning procedure will identify these gaps.

The sublexical phonotactic grammars characterize the static knowledge that speakers have about types of words (really, morphemes) that combine with phonologically selective morphemes. The grammars can also be used to assign the likelihood to nonce words: when the grammar is given a new input for which the diacritic is not known, each sublexical phonotactic grammar can assign the probability to the item based on its phonotactic well-formedness in the grammar.

For example, the sublexicons defined by English [ə] and [ən] are so large that they are mostly consistent with English phonotactics in general, but there are subset generalizations concerning the first syllable of the word: in the [ə] sublexicon, almost none of the items are vowel-initial, so a strong constraint against initial vowels will be posited, *[V. The [ən] sublexicon motivates a constraint against initial consonants, *[C. For a learner whose [ən] sublexicon exceptionlessly obeys *[C, the constraint will be weighted heavily and be influential in assigning new items to the [ən] sublexicon, but for learners whose sublexicons are variable, generalizing will be harder, and more so when the number of exceptions is considerable (Albright 2009b, Albright and Hayes 2003).

Assigning diacritics to novel items. When a novel morpheme (a wug, Berko 1958) is submitted to the grammar, it does not have a diacritic required for the application of insertion rules. The phonotactic sublexical grammars can be used to assign diacritics to wugs using probabilistic inference (Albright 2008, Becker and Gouskova 2012, inter alia). In general, the phonological restrictions of an affix are more reliable as a predictor of its productivity when the affix itself combines with a large number of morphemes—the larger the number, the more evidence the learner has for the restrictions (see Plag, 1999, Hay and Baayen, 2002, Albright and Hayes, 2003, 2006, Embick and Marantz, 2008). Take the classic example of an unproductive affix in English, "-th$_{\mathrm{N}}$", which attaches to just eight adjectives {*broad*, *dead*, *deep*, *long*, *strong*, *true*, *warm*, *wide*} and six verbs {*bear*, *gird*, *grow*, *heal*, *steal*, *till*} (Baayen et al. 1993). Some generalizations can be made about these morphemes: for example, they are all monosyllabic, and none of them end in stridents or [k]. But we would not expect the satisfaction of these generalizations to enable monosyllables to combine with the affix, since the likelihood of any root combining with this affix is so miniscule to begin with.

Applied to the English "a/an" example, this makes a prediction: overall, both allomorphs of the indefinite are quite productive, but more words in English begin in consonants than in vowels, so [ə] is more productive than [ən] (in the Carnegie Mellon dictionary, 81% of all words are consonant-initial). To the extent that learners might generalize one of the allmorphs beyond its baseline distribution in the learner input, we would expect that they should extend the use of [ə]. This appears to be correct for the English dialects examined by Gabrielatos et al. (2010): [ə] is becoming more general, and it is used before both consonant-initial words and some vowel-initial words.

[3]The nested structure of sublexicons has been noted for some time (see Ito and Mester 1995, Fukazawa et al. 1998), although there are some disjunctive relationships between sublexicons (Kawahara et al. 2002, Jurgec 2010).

3 Case study: Russian diminutives

3.1 Overview

We now turn to a more complicated case: Russian diminutive allomorphy. We start by presenting the background on the morphological and phonological properties of Russian diminutives. We then show how we constructed sublexicons of Russian diminutive bases and derived forms, and summarize the phonological properties of each sublexicon in a statistical study. Finally, we present an elicitation study in which we asked Russian speakers to diminutivize words that do not have commonly used diminutives. We show that people's choices of allomorphs are largely predictable from whether the bases and resulting diminutive forms did well in our sublexical phonotactic grammars. We also show that the choice of allomorph was sensitive not just to features that were local to the suffix, such as the final consonant of the base, but also to some non-local features. This latter result distinguishes our theory from several alternatives.

3.2 Background: morphological properties of diminutives

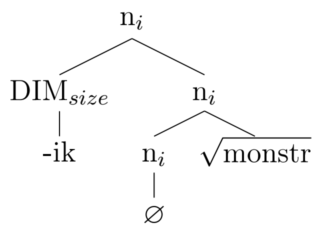
Diminutives are common in Russian speech, and they are productive in the sense that they attach to new loanwords such as [ajfón] 'iPhone' (see (8c)). Russian nouns belong to one of three genders (masculine, neuter, feminine) and can be further subdivided into declension classes; diminutives are formed differently for each gender and declension class (Stankiewicz 1968, Hippisley 1996, Kempe et al. 2003, Steriopolo 2008). We discuss only the masculine nouns here. Masculines form diminutives with four allomorphs, as shown in (8). The allomorphs differ in their stress properties and in whether they undergo vowel deletion when a vowel-initial suffix follows.

(8) Russian masculine nouns form diminutives with four allomorphs: [-(o)k], [-ik], [-ʧ^{j}ik] and [-(ə)k]/[-(i)k].

	Nom sg	Dim nom sg	Dim gen sg	Gloss
a.	áng^{j}il	ang^{j}il-ók	ang^{j}il-k-á	'angel'
b.	mónstr	mónstr-^{j}ik	mónstr-^{j}ik-ə	'monster'
c.	ajfón	ajfón-ʧ^{j}ik	ajfón-ʧ^{j}ik-ə	'iPhone'
d.	r^{j}ib^{j}ónək	rib^{j}ónəʧ^{j}-ik	rib^{j}ónəʧ^{j}-k-ə	'child, baby'

Diminutives do not change the gender of the noun they attach to, and they have only a limited effect on the declension class. Steriopolo (2008) analyzes Russian diminutives as morphosyntactic adjuncts to already categorized roots; for example, the Russian masculine noun [monstr] 'monster' consists of a root and a null nominalizing head (see (9); we assume that the n_i head determines the declension class). In the vast majority of cases, diminutive words have corresponding bases that are also free-standing phonological words. When we talk about morphological bases and derived words below, we mean phonological words such as [mónstr] 'monster (nom sg)' and [mónstr^{j}ik] 'monster-dim (nom sg)'.

(9) The morphosyntactic structure of diminutives, after Steriopolo (2008)



The [-(ə)k]/[-(i)k] allomorph has a morphologically restricted distribution: it occurs after other diminutives or affixes that end in [k]. Thus, the double diminutive of [ang^{j}il] is [ang^{j}il-óʧ^j-ik], as shown in (10). There are a few sporadic exceptions ([karmán] 'pocket' ∼[karmáʂ-ək], [m^{j}iʂók] 'sack' ∼[m^{j}iʂóʧ^{j}ik]; see Polivanova 1967). This suffix had the lowest productivity in our elicitation study (as well as in the sublexicons we collected), and we do not analyze it in detail in the remainder of the paper.

(10) Restricted distribution for [-(ə)k]/[-(i)k]: after k-final morphemes

			with affix		Diminutive nom. sg.	Diminutive gen. sg.
a.	rabót-ə	'work'	rabót-n^{j}ik	'worker'	rabótn^{j}iʧ^j-ik	rabótn^{j}iʧ^j-k-ə
b.	áng^{j}il	'angel'	ang^{j}il-ók	'angel-dim'	ang^{j}ilóʧ^j-ik	ang^{j}ilóʧ^j-k-ə

3.3 Phonological properties of diminutives

As can be seen from the examples in (8), the diminutive allomorphs differ in their stress properties and the effect on the last consonant of the stem. Russian stress is lexical and contrastive (Halle 1973a, Melvold 1989, Alderete 1999, Revithiadou 1999, Gouskova and Linzen to appear, inter alia): morphemes differ in the presence of stress, in stress location, and stress mobility. This is illustrated in (11) and (12). The [-ok] suffix is stress-dominant: even if the stem is lexically stressed, as in [áng^{j}il], the diminutive pulls stress onto itself (as in [ang^{j}ilók]) or, when its own vowel deletes, onto the following vowel (as in /áng^{j}il-ok-a/ →[ang^{j}ilká]). These examples also show regular features of Russian phonology such as vowel reduction in unstressed syllables and devoicing of stem-final consonants (both seen in /gorod/ [górət]∼[gəradók]; see Crosswhite 1999, Padgett 2002, and others).

(11) Dominant diminutive suffix [-ok]

	Base: fixed	Dim: final	Base: mobile	Dim: final
Nom sg	**á**ng^{j}il	ang^{j}il-**ók**	g**ó**rət	gərad-**ók**
Gen sg	**á**ng^{j}il-ə	ang^{j}il-**k-á**	g**ó**rəd-ə	gərat-**k-á**
Nom pl	**á**ng^{j}il-i	ang^{j}il-**k-^{j}í**	gərad-**á**	gərat-**k-^{j}í**
Gen pl	**á**ng^{j}il-əf	ang^{j}il-**k-óf**	gərad-**ó**f	gərat-**k-óf**
	'angel'	'angel (dim)'	'city'	'city (dim), town'

The [-ik] and [ʧ^{j}ik] allomorphs pattern differently from [-ok]. For the most part, they attach to words that already have stress on the last syllable, but there are a few examples like [sókəl], where stress moves to the last syllable in the diminutive. All such stems either already follow the mobile stress pattern or have a fixed and a mobile variant:

(12) Stress patterns in diminutives derived with [-ik]

	Base: final	Dim: fixed	Base: fixed/mobile	Dim: fixed
Nom sg	kr**ó**t	kr**ó**t-^{j}ik	s**ó**kəl	sak**ó**l-^{j}ik
Gen sg	krat-**á**	kr**ó**t-^{j}ik-ə	s**ó**kəl-ə	sak**ó**l-^{j}ik-ə
Nom pl	krat-**í**	kr**ó**t-^{j}ik-^{j}i	s**ó**kəl-**í,** səkal-**á**	sak**ó**l-^{j}ik-^{j}i
Gen pl	krat-**ó**f	kr**ó**t-^{j}ik-əf	s**ó**kəl-əf, səkal-**ó**f	sak**ó**l-^{j}ik-əf
	'mole'	'mole (dim)'	'falcon'	'falcon (dim)'

The stress properties of the diminutives are best seen in doublets that can diminutivize with more than one suffix:[4]

(13) Stress alternations in diminutive doublets

	Unaffixed	With case suffix	[-ok]	[-ik]	[-ʧ^{j}ik]	Gloss
a.	gólup^j	galub^{j}éj	galubók	—	galúpʧ^{j}ik	'pigeon'
b.	kav^{j}ór	kavrá	—	kóvr^{j}ik	kav^{j}órʧ^{j}ik	'carpet'
c.	kl^{j}ón	kl^{j}ónə	kl^{j}inók	kl^{j}ón^{j}ik	—	'maple'
d.	kóləkəl	kələkəlá	kələkalók	—	kəlakól^{j}ʧ^{j}ik	'bell'
e.	sókəl	sókəli, səkalá	səkalók	sakól^{j}ik	—	'falcon'

Diminutive affixation brings various changes to the last consonant of the stem. The [-ik] allomorph palatalizes stem-final consonants that have palatalized counterparts, including /ʦ/, which becomes /ʧ^j/ (e.g., [pál^{j}iʦ] 'finger' ∼[pál^{j}ʧ^{j}ik] 'finger-dim'). The [-ok] allomorph does not generally alter the last consonant unless it is a dorsal, in which case the dorsal changes to a corresponding strident: /k, g, x/ become [ʧ^j], [ʐ], [ʂ]. As shown in (14), this rule is triggered by only some morphemes and does not apply in the context of other, phonologically similar suffixes (see Padgett 2003, Iosad and Morén-Duolljá 2010, Kapatsinski 2010). Note also that while the diminutive triggers this consonant mutation on the stems, it does not undergo any mutation itself (unless followed by the diminutive [-ək], which behaves just like [-ok] with respect to mutation). Finally, the [-ʧ^{j}ik] allomorph palatalizes /l/ but not other consonants.

(14) Stem-final consonant mutation in [-ok] diminutives

	Unaffixed	Inst sg	Diminutive	Dim Inst sg	Gloss
a.	ut^{j}úk	ut^{j}u$\underline{\text{g}}$óm	ut^{j}u$\underline{\text{ʐ}}$ók	ut^{j}uʂkóm	'iron'
b.	kazák	kaza$\underline{\text{k}}$óm	kəza$\underline{\text{ʧ}^j}$ók	kəzaʧ^{j}kóm	'cossack'
c.	lapúx	lapu$\underline{\text{x}}$óm	ləpu$\underline{\text{ʂ}}$ók	ləpuʂkóm	'burdock'

We already mentioned that [-ok] loses its vowel when a vowel-initial suffix follows, and so do some other morphemes (e.g., /kov^{j}or/ keeps its last vowel in [kav^j$\underline{\text{ó}}$r] 'carpet', [kav^j$\underline{\text{ó}}$rʧ^{j}ik] 'carpet-dim (1)' but loses it in [kav_rá] 'carpet (gen sg)' and [kóv_r^{j}ik] 'carpet-dim (2)'). Vowel deletion in Russian is complicated and has been the subject of much study (Lightner 1965, Halle 1973a,b,

[4] There is some disagreement in the Slavicist literature as to whether these doublets are all diminutives (Polivanova 1967: fn. 4 on p. 10, Zaliznjak 1985: p. 53); we take the position that people parse them as such, regardless of etymology. There are definitely some cases that started out as diminutives etymologically but have diverged, e.g., the diminutive of [lóp] 'forehead' is [lób^{j}ik], not [labók] 'pubis'. In cases such as this, either the string is derived from a homophonous root, or it has been either reanalyzed as morphologically simple.

Melvold 1989, Yearley 1995, Hermans 2001, 2002, Becker and Gouskova 2012, Gouskova and Becker 2013, Gouskova 2012). The behavior of vowel deletion stems in our experiment is not discussed in this paper for lack of space, but see Gouskova (in preparation).

3.4 Phonological predictability and lexical factors

When diminutive affixes attach, bases alternate dramatically. At the same time, the diminutive itself is not realized uniformly. Figuring out the conditions for these alternations presents a big challenge to the learner. Fortunately, there is a way to predict how a noun will diminutivize based on its phonological shape. Polivanova (1967) identifies a number of phonological generalizations about diminutive formation. She notes, for example, that the [-ok] allomorph tends not to attach to nouns that end in a consonant cluster, or have a stress that is neither initial or final. The [-ik] allomorph happily attaches to CC-final nouns, but it avoids dorsal-final nouns, and it does not easily attach to nouns that have non-final stress. The [-ʧʲik] allomorph disprefers final clusters, final dorsals, coronal obstruents [t, d, s, z, ʂ, ʐ, ʃʃʲ, ʧʲ], and nouns with non-final stress:

(15) Some of Polivanova's generalizations about Russian diminutive allomorphy

		[-ok]	[-ik]	[ʧʲik]
a.	Final syllable	no final CC	—	no final CC
b.	Final C place	—	no final dorsals	no final dorsals, no COR obstr.
c.	Stress location	not medial	must be final	must be final
d.	Final C manner	—	—	prefers sonorants

Recall from (14) that stem-final consonant mutation is affix-specific: the dorsal in /utʲug̲/ spirantizes in [utʲuz̲ók] but not [utʲug̲-óm]. The restrictions listed above are similarly affix-specific—the clearest evidence for this comes from affixes that are homophonous but do not impose such restrictions. Both the [-ik] allomorph and the [-ʧʲik] allomorph have homophones: the borrowed Latin suffix [-ik] '-ic' (see (16)) and the agentive suffix [ʧʲik] '-er' (see (17)). The latter suffix is an allomorph of [-ʃʃʲik]. These examples show that the diminutive [-ik] does not avoid dorsal-final stems because dorsal-[i] sequences are prohibited in Russian—they are allowed. The diminutive [-ik] is in fact often followed by the plural [-i] (recall [krótʲikʲi] 'moles-dim' and [sakólʲikʲi] 'falcons-dim' in (12)). Likewise, sequences of coronal obstruents followed by [-ʧʲik] are attested, and in fact quite common, in agentive nouns.

(16) Diminutive suffixes vs. homophones: [-ik] '-ic'

	Unaffixed		With '-ic' (gen pl)	Gloss
a.	psʲíx	'psycho'	psʲíx̲ʲik	'psyche'
b.	pʲidagók	'pedagogue'	pʲidagóg̲ʲik	'pedagogy'
c.	lʲingvʲist	'linguist'	lʲingvʲistʲik	'linguistics'

(17) Diminutive suffixes vs. homophones: [-ʃʃʲik]/[ʧʲik] '-er'

	Unaffixed		With '-ic' (gen pl)	Gloss
a.	pulʲimʲot	'machine gun'	pulʲimʲótʧʲik	'machine gun operator'
b.	danós	'report to the authorities'	danóʃʃʲik	'informant'
c.	kavʲór	'carpet'	kavʲórʃʃʲik	'carpeter'

Thus, an important feature of Russian diminutive allomorphy is that some of the generalizations are *source-oriented* (Bybee 1995, Kapatsinski 2010): they concern the shape of the base, but not necessarily the shape of the resulting diminutive. Take velar palatalization: dorsal-final stems such as [lapúx] become strident-final once [-ok] is attached to them, [ləpuʂ̺-ók]. The suffix [-ik] attaches to [ʂ̺]-final nouns, [ʂ̺alaʂ̺] 'shack' [ʂ̺aláʂ̺-ik] 'shack-dim', but not if the [ʂ̺] is derived from /x/: thus, [ləpuʂ̺ók] and [ʂ̺aláʂ̺ik] are fine, but /lapux-ik/→*[lapúʂ̺ik] is impossible. The stress restrictions are similarly source-oriented. All [-ok] diminutives are stress-final by definition, so why should [-ok] shirk bases with medial but not initial or final stress? An entirely output-oriented theory of allomorphy would have difficulty with these facts (see section 6.2). As we will show, the phonological shape of the diminutive does matter in selecting allomorphs, but so does the base.

An insight Polivanova brings is that diminutive formation is lexically idiosyncratic: some nouns have more than one diminutive (recall (13)), and others do not have diminutives at all. The last category is particularly interesting: in some cases, diminutives are just absent, even though no phonological constraints preclude their formation (e.g., [son] 'sleep, dream' has no diminutive). In other cases, diminutives are impossible because each allomorph's phonological preferences are violated. Polivanova classifies as "difficult" words with medial stress, such as [aftóbus] 'bus', which is at best ?[aftóbus^{j}ik] in the diminutive. She groups under "impossible" words that end in consonant clusters and have a dorsal as their last consonant, such as [l^{j}ém^{j}ink] 'lemming' and [mal^{j}úsk] 'mollusk'. This distinction between "difficult" and "impossible" is traditional in Russian linguistics—paradigm gaps are often classified into those categories (Zaliznjak 1977 and others).

It should be kept in mind, however, that Polivanova's generalizations are based on traditional linguistic methodology. She is reporting her own intuitions and generalizations she identified by examining a dictionary—and it is not clear that Russian usage obeys these generalizations. Thus, we found examples of diminutives that she judges to be "difficult" in the Russian National Corpus, e.g., her *[glóbus^{j}ik] has two hits in the corpus, and ?[aftóbus^{j}ik] has 12 hits. For most of the generalizations she identifies, she lists counterexamples—and this is true for phonological as well as semantic and stylistic factors. Our interpretation is that the list of diminutives in current use informs the phonological generalizations, but satisfying the phonological trends does not mean a lexical diminutive will automatically sound good to a Russian speaker. Indeed, it appears that certain stems are specified as not combining with diminutives at all. But one linguist's intuitions are not enough to conclude that diminutive formation is impossible for any given item. This motivates the elicitation study that we describe in section 5, as well as the lexicon study we turn to next.

4 A study of the lexicon

4.1 Assembling the diminutive sublexicons

To examine the phonological characteristics of Russian nouns that combine with each diminutive allomorph, we need proxies for the sublexicons that Russian speakers would be exposed to. We constructed sublexicons by doing an exhaustive search for all possible diminutive forms for each masculine noun in a large electronic dictionary. We started with Usachev's (2004) electronic paradigm list, which is based on Zaliznjak's (1977) dictionary. The dictionary lists approximately 20,000 masculine nouns; taking each of them as a base, we constructed three orthographic diminutive forms. Table 1 shows some example constructed diminutives, which incorporate some stem-final consonant changes (e.g., k→ʧ^j in diminutives with [-ok]). Stress is marked in Zaliznjak's dictionary but is not normally shown in Russian orthography and in the corpus we used. For some stem type+affix combinations, Russian has gaps, so it is not clear what the diminutive forms are supposed to look like—in such cases, we searched for linear concatenations of stems and suffixes. One such example

	[-ok]	[-ik]	[-ʧʲik]	
kazak	**kazaʧʲok**	*kazakʲik	kazakʧʲik	‘Cossack’
vʲetʲer	**vʲetʲerok**	veterʲik	veterʧʲik	‘wind’
golubʲ	**golubok**	golubʲik	**golubʧʲik**	‘pigeon’
pʲenal	pʲenalok	**penalʲik**	**pʲenalʲʧʲik**	‘pencil box’

Table 1: Examples of constructed diminutives we searched for

is marked with an asterisk in the table. The boldfaced examples got hits in the search for these particular bases.

To identify which of these possible diminutive forms are attested in Russian usage, we searched for singulars and plurals of each of the constructed diminutives in every one of six possible case forms in the Russian 1-gram portion of the Google Ngrams corpus (Michel et al. 2011). This corpus contains orthographic words obtained by optical character recognition from scanned books, and it is quite large (11 million words) and downloadable, allowing for quick offline searches. The corpus is noisy, however, so the results required some cleaning.[5] The total number of searches we performed was almost a million (20,000 nouns $\times$ 4 diminutive suffixes $\times$ 12 case/number forms). We eliminated all hits older than 1950, and before cleaning, we got 771 [-ʧʲik] words that had occurred in at least one case form, 1319 [-ik] words, and 3626 [-ok] words. After cleaning (described next), we got 547 diminutive forms with [-ʧʲik], 640 [-ik], and 1197 [-ok]; 17850 nouns did not occur in what we considered to be diminutive forms in the Ngrams corpus.

The biggest challenge in cleaning the results was eliminating homographs that were not diminutives. As mentioned in section 3.2, each of the allomorphs is homographous with a non-diminutive suffix: [-ʧʲik] is also an occupational suffix (e.g., [pulʲemʲot] ‘machine gun’, [pulʲemʲotʧʲik] ‘machine gun operator’), [-ik] is also the cognate of the Latinate *-ic,* which forms abstract feminine nouns in Russian (e.g., [lʲingvʲist]‘linguist’, [lʲingvʲistʲik] ‘linguistics (gen. pl.)’), and [-ok] is homographous with the suffix [-(ə)k], used to make feminine nouns (e.g., [udmurt] ‘an Udmurt man’, [udmurtək] ‘an Udmurt woman (gen pl)’. The latter two types of cases were eliminated if they were attested in case forms that disambiguate masculines from feminines (e.g., [udmurtk-oj] ‘an Udmurt woman (inst sg)’ or *[lingvist-k-ov] ‘linguist (masc gen pl)’). As for [-ʧʲik], we eliminated all bases that ended with coronal obstruents, since manual inspection confirmed that all of them were instances of the occupational suffix.

4.2 Analysis

We converted the Russian orthographic representations from the corpus into a broad phonemic transcription. We did not transcribe vowel reduction or voicing assimilation/devoicing, but we eliminated orthographic artifacts such as sequences of identical consonants in words such as <килограмм> [kʲilogram] ‘kilogram’, which are generally pronounced as singletons (Dmitrieva 2012). We annotated the diminutives for various features such as stress location, presence of clusters in initial,

[5] We searched the Russian National Corpus for diminutives, as well. The corpus allows to search for diminutives by tag, which appears to be assigned by NLP methods rather than by human checkers; the resulting list of 1800 or so unique case forms identified as masculine diminutives included many forms that were not diminutive or not masculine.

medial and final position, place and manner of the final consonant, and the presence of vowel hiatus. For base stress, we followed Zaliznjak (1977); for diminutive stress, we assumed it was on the last syllable for [-ok] diminutives, and in the same location as in the base in the diminutives with [-ik] and [-ʧʲik]. Monosyllables were coded as having final stress. The distribution of these phonological features across the sublexicons is plotted in Fig. 1. The figure shows proportions of each feature in the sublexicons. We start with descriptive observations; a statistical analysis of these trends appears at the end of this section.

Place and manner of articulation of the last consonant. Of all masculine nouns, about 28% end in dorsals, but the proportion of dorsals in the [-ʧʲik] and [-ik] sublexicons is close to zero. There are considerably more dorsal-final words among words that combine with [-ok], however. The most frequent manner of articulation of the last consonant in lexical masculines is stop (43%), followed by liquid (19%), fricative (17%), and nasal (14%). These proportions do not differ dramatically in the [-ik] and [-ok] sublexicons, but sonorants are noticeably more frequent than obstruents in the [-ʧʲik] sublexicon. Partly this is because final coronal obstruents—including stridents—are completely absent from the [ʧʲik] set, in an OCP-like pattern of dissimilation. Words that end in [j] most often diminutivize with [-ʧʲik], and almost never with [-ik] (again, in a dissimilatory pattern). Final consonant clusters occur in 15% of masculines and in 22% of the nouns that combine with [-ik], but they are practically unattested in the [-ʧʲik] sublexicon (<1%), and they are rare in the [-ok] sublexicon (4%).

Stress and non-local properties. Masculine nouns as a whole usually have final stress (55.7%), followed by medial (27.6%) and initial stress (16.6%). Stress is overwhelmingly final on nouns that combine with [-ʧʲik] and [-ik]; words with medial or initial stress are underattested in all of the diminutive sublexicons, though they are comparatively more common in the [-ok] subset than in the other two. Hiatus is fairly rare—it occurs in 8% of the masculines, and it is even more rare in the [-ʧʲik] sublexicon (4%) and in the [-ik] and [-ok] sublexicons (2% and 1% respectively). Medial consonant clusters occur in 18% of all masculines, but they are far less common in the diminutive sublexicons. We also considered the distribution of vowels of different quality in final syllables and identified some trends towards height and rounding harmony (cf. Becker 2009 on a similar trend in Hebrew noun plurals). Last syllable high vowels are considerably more common among words that take [-ik], and round vowels are considerably more common among nouns that take [-ok] (this is not shown on the graph).

There were a few features whose distribution was balanced across the sublexicons and not statistically different from the set of masculines as a whole: word-initial consonant clusters and first syllable vowels did not differ significantly (monosyllables' vowels were counted as last vowels for the purposes of this analysis).

To assess which trends were statistically significant, we fitted a polytomous regression (Arppe 2013, *polytomous* package) to the lexicon data in R (R Development Core Team 2013). Polytomous regressions are logistic regressions with dependent variables that have more than two categorical outcomes—in our case, [ʧʲik], [-ik], [-ok], and "none". We used the *one-vs-rest* heuristic, which fits a series of logistic regressions comparing each level of the dependent variable with the rest of the group (e.g., [-ʧʲik] takers vs. [-ik]+[-ok]+"none"). The set of nouns that do not occur with diminutives is an order of magnitude larger than the diminutive sublexicons, which can skew the results when the one-vs-rest heuristic is used. Thus, we took a random sample of 700 nouns from the "none" category for this regression.[6] Table 2 shows the log odds for each predictor, by suffix. We include all the

[6] We also fit a polytomous regression to the entire dataset, without sampling, and the results were qualitatively similar to the regression we report. Two predictors were significant in this full regression but not in the randomly

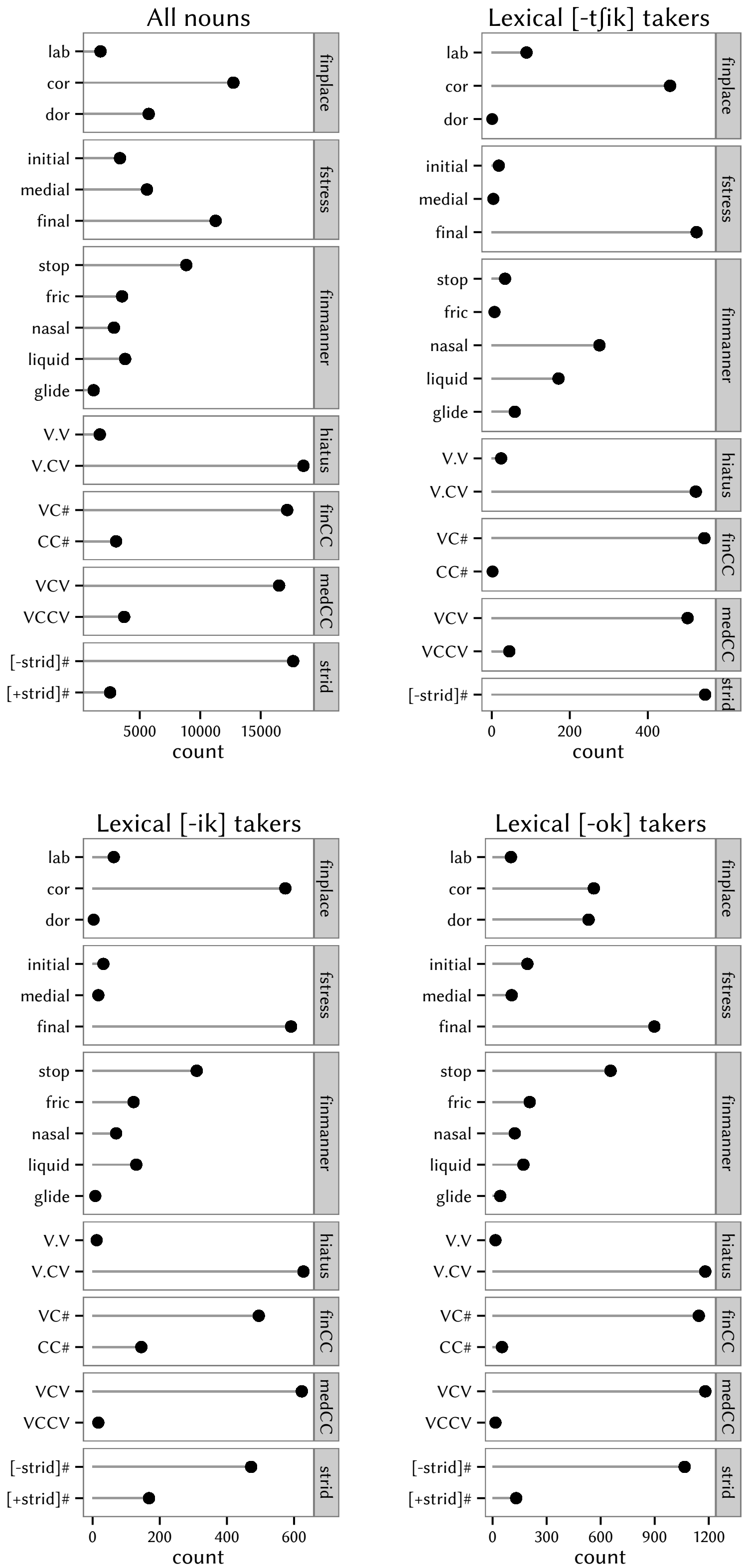


Figure 1: Diminutive sublexicons: properties of bases

	"none"	[-ʧʲik]	[-ik]	[-ok]
(Intercept)	-0.8317	-4.629	-1.359	(-0.0397)
lastC=dor	-0.4398	-3.156	-4.827	2.181
lastC=nasal	-0.538	3.095	-1.441	-0.5667
lastC=approx	(-0.2305)	2.356	-0.792	-0.2521
lastC=glide	0.7263	(0.3457)	-1.651	(-0.2204)
final.stress	-0.8049	1.78	1.153	-0.6731
medial.stress	0.9728	(-0.523)	(0.6542)	-2.008
final.CC	0.8348	-3.538	0.742	-1.11
initial.CC	0.2936	-0.3431	(-0.1416)	(-0.03847)
medial.CC	1.881	0.4724	-1.336	-2.344
lastV=high	0.3928	(-0.2038)	-0.3794	(-0.06273)
lastV=round	-0.2901	(0.0202)	(-0.06133)	0.3315
medial.stress:hiatus	(-0.6041)	(-12.63)	(-0.8467)	1.781
hiatus	0.9674	(0.3329)	(-0.4464)	-1.41

R^2 likelihood: 0.26, AIC: 6025, BIC: 6362. Null deviance: 8038 on 12000 df
Residual (model) deviance: 5913 on 11944 df

Table 2: Polytomous regression table for the lexicon

predictors tested in the model, even those that were not significant for any of the suffixes, to give a complete picture of the lexicon. Log odds appear in parentheses for predictors whose p values exceed 0.05. When a predictor is significant in this model, it means that the likelihood of ending up in the relevant subcategory increases (positive coefficients) or decreases (negative coefficients) for bases that have the feature encoded by the predictor compared to the likelihood of ending up in the category without the feature.

As can be seen from the model, the likelihood of [-ok] increases when the base noun ends in a dorsal, and the likelihood of the other suffixes (including 'none') decreases significantly. The effect of labials is not significant. As for manner, compared to stops, noun-final fricatives do not affect the affixation rates, but nasals increase the chances of [-ʧʲik]. Approximants (l, r) likewise increase the chances [-ʧʲik] and decrease the chances of [-ik]. Word-final glides (j, v) decrease the chances of [-ik]. The quality of the vowel in the last syllable has an effect on lexical affixation: high vowels decrease the likelihood of [-ik], and round vowels increase the likelihood of [-ok], in a kind of vowel harmony pattern. The presence of a consonant cluster also affects affixation rates, and this effect is not confined to the last syllable: medial clusters are significantly more likely in the "none" set than in the [-ik] and [-ok] sets. Initial clusters were more likely to occur in the randomly selected subset of words that do not combine with diminutives than in any of the diminutive sublexicons. Finally, vowel hiatus was significantly less likely to occur in bases that combine with [-ok] than in the other two suffix sublexicons; hiatus was significantly more likely in the "none" subsets. This effect went the opposite way, however, in words that had medial stress and hiatus.

sampled subset regression: final labial consonants were significantly more likely to occur in the [-ok] sublexicon but not in the others, and fricatives were significantly more likely to occur in the 'none' sublexicon than in the others. There were also minor differences in the significance of effects for hiatus and stress. TheR^2 likelihood of the full model was 0.20, AIC 15770, BIC 16280, residual model deviance 15640 on 80872 degrees of freedom.

4.3 Training the UCLA Phonotactic Learner on the sublexicons

We next trained the UCLA Phonotactic Learner (Hayes and Wilson 2008) on each sublexicon of bases and affixed diminutives. The learner is given a lexicon (a list of phonological words) of a language, and a set of features to describe all the segmental and suprasegmental distinctions transcribed in the lexicon. The learner then posits a grammar of weighted markedness constraints that characterize the structures that are unattested or underattested in the lexicon. The constraints generally refer to n-grams of natural classes and word edges; and for example, given a lexicon of English, the learner will notice that velar nasals are unattested word-initially and posit a constraint *[word boundary][+nasal, Dorsal]. One limitation in the computational implementation of the learner is that it needs at least 3,000 phonological words to learn from; we thus multiplied the sublexicons that we got from the N-grams corpus until that threshold was reached. To describe the Russian sublexicons, we used a feature set similar to that in Hayes (2009). Stress in bases was transcribed as in Zaliznjak (1977). We obtained affixed diminutives from an orthographic corpus, and Russian orthography does not normally indicate stress. Thus, we added stress to the diminutives following the general rules for stress in words with these affixes (see Zaliznjak 1985): we placed it on final syllables in all the [-ok] words; in the [-ik] and [-ʧ^{j}ik] words, it was left on the same syllable as in the base and corrected by hand for the few yer words in which stress moves to the left when the vowel deletes (e.g., [kov^{j}ór] 'carpet' vs. [kóvr^{j}ik] 'carpet-dim').

The output of the learner is a sublexical phonotactic grammar of constraints, and it can also be used to assign probability scores to new items it has not been trained on. These scores reflect the likelihood that the new item belongs to the lexicon. Since we obtained six different sublexicons with 30 constraints each, we will not describe the constraints in detail (the full grammars can be viewed on the project site at https://files.nyu.edu/mg152/public/russian/diminutives/). As we will show, however, the cumulative well-formedness of each base and derived word in the respective sublexicon largely determined how people diminutivized the words in the elicitation study.

5 An elicitation study

5.1 Hypotheses to test

We expect that in an elicitation study, *people's choice of affixes will depend on the phonotactic well-formedness of the base and the derived word in each affix-defined sublexicon.* To put it differently, if a noun sounds like an [-ok] base, and its diminutive sounds like an [-ok] diminutive, then the noun will combine with [-ok]—and here, "sounds like" means "receives a high probability in the sublexical phonotactic grammar". We expect that people will have difficulty diminutivizing words that receive a low probability from all of the diminutive subgrammars. On the other hand, if a word receives a high probability from more than one subgrammar, the choice of diminutive affix should vary.

The elicitation task encourages people to produce words that they might not necessarily use in normal speech, so we ask people to rate the words they give us. We expect the rating to reflect the phonotactic well-formedness of the diminutive in its sublexical grammar.

Our theory makes a prediction that distinguishes it from other theories of selectional restrictions. Some approaches to phonologically predictable morphology are *edge-oriented*; in such approaches, the choice of affix is determined by phonological properties of the word edge that the affix attaches to. For suffixes, the ends of words should matter; beginnings should not. Examples of such approaches include Alignment (McCarthy and Prince 1993) and the affixation rules of the Minimal Generalization Learner (Albright and Hayes 2003). Such approaches would have difficulty with a pattern in which properties of the middles or beginnings of words determine suffixation patterns.

There are two generalizations about the Russian diminutive sublexicons that are relevant: in bases that combine with [-ok], hiatus is under-attested, and so is medial stress. If diminutive affix choice depends on overall phonotactic well-formedness, we expect words with hiatus or medial stress to not combine with [-ok] as often as with other affixes. In an edge-oriented approach, such a pattern would require a different explanation. We will show that the Minimal Generalization Learner fails to learn the pattern altogether, because it is unable to characterize the stress alternations or even notice the rather strong generalizations about stress position in bases and derived diminutives.

5.2 Participants, design, and materials

5.2.1 Participants

There were 14 participants in the study (11 women, 3 men); aged 20-61 years (mean age of 32). All lived in Moscow at the time of the study and were native speakers of Russian, though most had experience with at least one foreign language. They were compensated the equivalent of $15 for their time. The experiment was conducted in Moscow by an experimenter who spoke Russian natively.

5.2.2 Procedure

The experimenter explained to the participant that various Russian words would appear on the computer screen. The task was to say the word out loud first, and then provide one diminutive of it—the first one that occurs to the participant. We also asked people to rate the diminutives they produced, on a scale of 1 (worst) to 5 (best). People were given examples of common diminutives ([kablúk]∼[kəblutʃʲ-ók] 'heel', [kóləkəl]∼[kəlakólʲ-tʃʲik] 'bell', [mʲátʃʲ]∼[mʲátʃʲ-ik] 'ball'), as well as examples of diminutives that were questionable and might get a rating of "3" (e.g., [prafʲésər] ∼ [prafʲésərʲik] 'professor'). The word list included 167 real nouns; most do not have commonly used diminutives. Since the task was difficult, we included a few control words that do have common diminutives (e.g., [kazák]∼[kəzatʃʲ-ók] 'Cossack'). Each person saw half of the test words (85); two of the participants worked through the entire list.

5.2.3 Materials

The words varied along the dimensions that Polivanova's descriptions predict to be relevant to affix choice: the place and manner of the last consonant, the presence of a cluster at the end of the word, and the location and mobility of stress in the base. The number of words in each category is shown below. No labial-final clusters were included—those tend to be somewhat uncommon, and we did not expect them to pattern differently from labial-final or cluster-final words (both combine with [-ik], as in [górp]/[górbʲik] 'hump'). We also included 11 words that had vowel sequences in various positions ([ópʲium], [áist], [alʲiándr], etc.).

(18) Place/manner of the last consonant; clusters

	Stop	Fricative	Nasal	Lateral	Rhotic
cor	s̯pagát (19)	kamís̯ (25)	bʲidón (14)	tʃixól (31)	abaz̯úr (30)
dor	kazák (21)	kóz̯ux (4)	NA	NA	NA
lab	pʲirʲiskóp (12)	rukáf (1)	astranóm (10)	NA	NA

No cluster	Cluster
bʲérkut (83)	tsimʲént (37)
bátʲik (18)	tʃʲitvʲérk (7)
tʃʲérʲip (22)	0

(19) Stress location in base and stress type

	Fixed type	Final type	Mobile type
final	aváns/aváns-i (45)	barsúk/bərsukʲ-í (37)	NA
initial	átəm/átəm-i (19)	NA	párus/parus-á (31)
medial	aftóbus/aftóbus-i (34)	NA	uʧítʲilʲ/uʧitʲilʲ-á (1)

5.2.4 On using real words

We use real words because we expected diminutive formation to depend on the lexical stress type of the base. That is, people might form diminutives differently depending on whether stress is fixed on some syllable of the stem, is predictably final, or alternates between stem and suffix (Polivanova 1967:11, Stankiewicz 1968). Lexical stress types of real words may vary for some words and for some speakers, but we could expect all of our speakers to exhibit some consistency in how they treated real words—on the other hand, establishing that some nonce words have mobile stress and others have fixed stress would require a complex multi-stage experiment.

But, real words do present a potential problem because they can fail to diminutivize for non-phonological reasons. This might be the case for words belonging to a formal register or ones that people find odd or incompatible with a diminutive/expressive suffix. We found, however, that people were happy to diminutivize [ajfón]→[ajfón-ʧʲik] 'iPhone', [manʲják]→[mənʲjiʧʲ-ók] 'maniac, serial killer' and [kənʲibál] 'cannibal', but had trouble with [klóun] 'clown' and [ótpusk] 'vacation'. Among established lexical diminutives, we find [pənʲidʲélʲnʲik] "Monday", but there is no diminutive for [ʧʲitvʲérk] "Thursday". Thus, we assume that semantic and stylistic factors play at best a minor role in this experiment. (The obvious next step is to conduct a nonce word study; such a study is reported in Anon in preparation).

5.3 Experiment results: overall patterns of production

We collected 1205 base-diminutive pairs with ratings that we could analyze. People formed 395 diminutives with [-ok], 361 with [-ik], 295 with [-ʧʲik], and 34 with [-ək]. They also produced 97 forms with other expressive suffixes (such as [-uʂkə] or [-iʂkə], or [-óniʂ] '-ling'). We coded these as "other". There were 15 gaps, where people could not produce any diminutives at all. The overall productivity of each pattern is summarized in the bar plot in Figure 2. Since the first three suffixes are the main focus of our study and they dwarf the other patterns in productivity, we focus on them in the remainder of the paper.

Figure 3 shows the distribution of phonological properties for the bases that appeared with each suffix; some illustrative examples of words with and without suffixes are given in (20) and (21). The entire dataset we collected is available on the supplemental materials page for the project at https://files.nyu.edu/mg152/public/russian/diminutives/. The first figure in 3 plots the breakdown of properties for all the noun bases in the experiment. The remaining figures show the distribution of these properties among the bases that people affixed with each suffix. These plots allow one to compare the relative proportion of, say, cluster-final nouns that were suffixed with [-ok] vs. [-ik] or [ʧʲik], even though the number of words in each category is different. Thus, out of the 395 bases that people suffixed with [-ok], 41 were cluster-final, and 354 ended in VC; for [-ik], there were about as many CC-final bases as VC-final bases (181 vs. 180). Fewer bases with hiatus were suffixed with

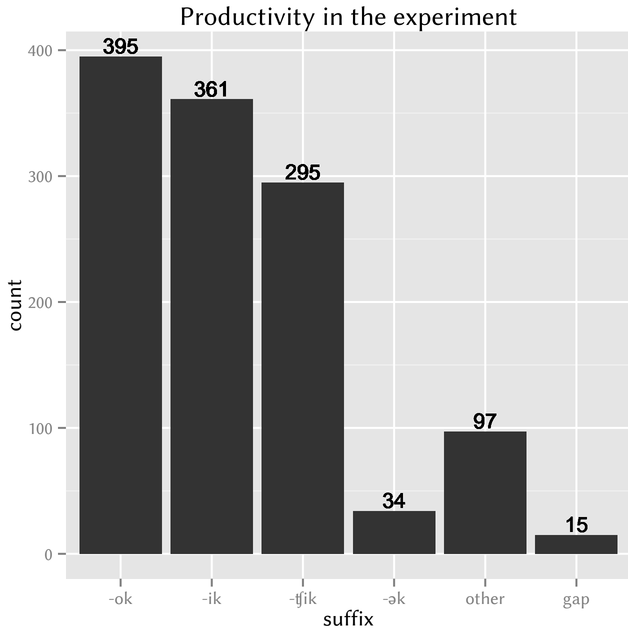


Figure 2: Overall productivity of each suffix in the experiment

[-ok] than with the other two suffixes. The location of stress in the base (initial, medial, final) also affected the choice of suffix: initial bases combined with [-ok] far more often than with [-ik] or [-ʧʲik]. Medial stress bases are less likely to combine with [-ok] than with [-ik] or [-ʧʲik], however. With respect to manner, sonorants preferentially combine with [-ʧʲik], whereas fricatives—which were overwhelmingly strident—tended to avoid that suffix; people suffixed most of the fricative-final bases with [-ik] or [-ok]. For place of articulation, dorsals were most often combined with [-ok], somewhat less often with [-ʧʲik], and almost never with [-ik].

(20) Segmental content/syllable structure: some examples of people's productions

	-ʧʲik	-ok	-ik
Cluster	kambájn/kambájn-ʧʲik	pójist/pəjizd-ók	mətatsíkl/mətatsíklʲ-ik
Dorsal	gʲértsək/gʲértsək-ʧʲik	manʲják/mənʲjáʧʲ-ók	patsólnux/patsólnuxʲ-ik
Coronal	astiróit/astiróidʧʲik	xóbət/xəbat-ók	ṣpagát/ṣpagátʲ-ik
Labial	pʲirʲiskóp/pʲirʲiskópʧʲik	ʧʲérʲip/ʧʲirip-ók	jipʲískəp/jipʲískəp-ʲik
Hiatus	klóun/klóun-ʧʲik	ópʲium/apʲium-ók	anʧʲóus/anʧʲóusʲ-ik

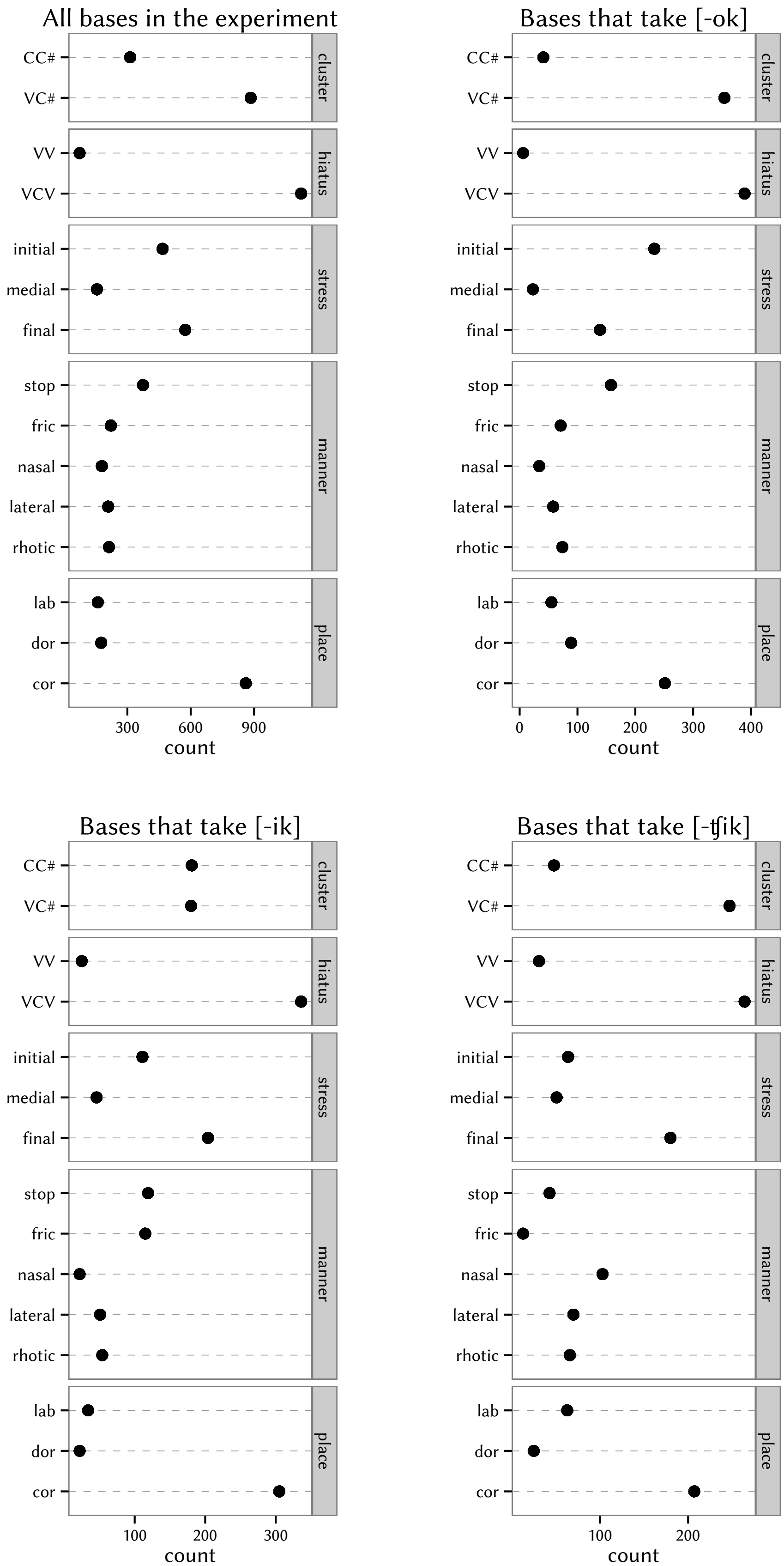


Figure 3: Number of bases people produced with each suffix, organized by phonological property of the base

(21) Stress lexical type and mobility, a few examples of people's productions

	moved	stayed
initial	[órdʲin] vs. [ardʲin-ók]	[kónus] vs. [kónus-ʲik]
medial	[aftóbus] vs. [aftəbus-ók]	[tarántul] vs. [tarántulʲ-ʧʲik]
final	[ʧʲirtʲóz̦] vs. [ʧʲirtʲiz̦-ók]	[jurʲíst] vs. [jurʲíst-ʲik]

5.4 Statistical analysis of suffix choice

To understand what conditioned the choice of suffix, we conducted two analyses. First, we used the probability scores from the sublexical phonotactic grammars trained on base and diminutive phonological words with each suffix. Since these scores are wholistic and do not allow for an easy examination of the factors affecting suffix choice, we also examined the effects of various hand-picked phonological predictors. We start with the more successful analysis, the sublexical phonotactic model.

5.4.1 The sublexical phonotactic model

The predictors we used in the first analysis are shown graphically in Fig. 1. The leftmost graph shows the violation scores assigned to each base and diminutive in the production study by the sublexical grammars trained on bases and diminutives from the Ngrams corpus (recall sections 4.1 and 4.3). For example, take the base [slovák] 'Slovak' and the diminutives people produced for this form, which included [slovaʧʲók], [slovákʧʲik], and [slováʧʲișka]. We trained sublexical phonotactic grammars on Ngram bases for [-ik], [-ok], and [ʧʲik], and diminutives formed with each of these suffixes. Each base subgrammar was then tested on all the experimental bases, and each diminutive subgrammar was tested on all the diminutives people produced. The summed scores for each base and diminutive are shown in the graphs, broken down by production patterns.[7] For example, the leftmost column in the [ʧʲik] graph shows the scores given to [slovák] and [slovákʧʲik], and the rightmost column in the same graph shows the scores given to [slovák] and [slovaʧʲók]. As expected, since the base ends in a dorsal, it is not phonotactically suited to be a base for [-ʧʲik], and its violations would contribute to the height of the rightmost column. The violations of diminutives such as [klóunʧʲik], on the other hand, would be plotted in the leftmost columns in the [ʧʲik] graph and the [-ok] graph. The base [klóun] was affixed with the [-ʧʲik] suffix more than [-ok]. As can be seen from these graphs, the corpus-trained sublexical grammars for a given suffix assign the best (lowest) scores to the forms that people produced with that suffix. This pattern is consistent with the hypothesis that people are guided by the sublexical phonotactic well-formedness of the base and the diminutive when they choose a suffix to put on a base.

A polytomous logistic regression model was fitted for the choice of suffix: [-ik] vs. [-ok] vs. [-ʧʲik] vs. "none/other". There were six predictors, derived from the scores assigned to the experimental bases and the diminutives by sublexical grammars. The model with raw scores produced by the phonotactic learner (plotted in Fig. 1) failed to converge, so we simplified the predictors to "pass" and "fail" scores: if a grammar assigned a score of 0 to a form, it was a pass, otherwise it was a fail.

[7] These graphs do not include gaps—bases for which people failed to produce an affixed form—because we cannot train a diminutive sublexical grammar on them. The forms affixed with [-ək] are excluded to save space.

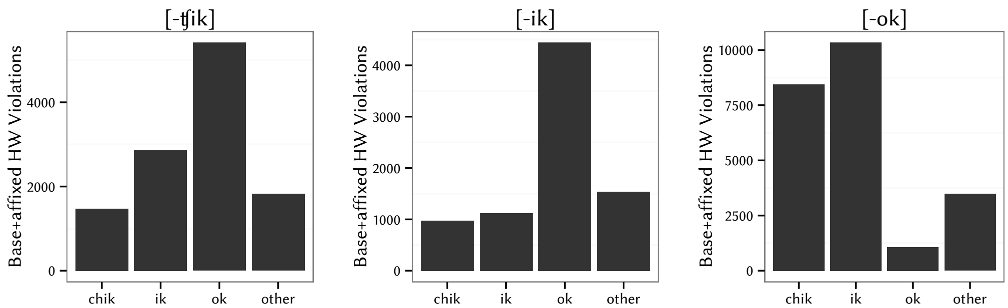


Figure 4: Summed violation scores assigned by grammars trained on base and diminutive sublexicons, grouped by people's production patterns.

The model included random intercepts for participant and word (base); models with more complex random effect structures failed to converge in R.

The log-odds from the model are summarized in Table 3. Predictors that are significant at $p < 0.05$ are given without parentheses. The first two rows show the overall likelihood of each suffix compared to the baseline poisson distribution. The suffixes [ʧʲik] and [-ok] are less productive than might be expected by chance, and [-ik] is more productive. The rest of the table shows the log-odds of an affix when the base received a score of 0 in the sublexical phonotactic grammar trained on the affixal bases (e.g., *ok-base*), and when the affixed diminutive that people produced received a 0 in the sublexical phonotactic grammar trained on the diminutives containing the affix (e.g., *ok-dim*). All predictors have log-odds that go in the right direction, but they do not always reach significance at $p < 0.05$. Thus, passing the grammar trained on lexical [-ok] diminutives, *ok-dim*, significantly increases the likelihood of a base appearing with the [-ok] suffix in the production study. The grammar trained on [-ok] bases, *ok-base*, trends in the same direction but fails to reach significance for [-ok]; doing well in that grammar does, however, significantly increase the chances of being affixed with [-ik]. The words that did well in the [-ik] diminutive grammar were more likely to end up with [-ik] and less likely to end up suffixed with "other". The predictor for the grammar trained on [-ik] bases trends in the right direction for [-ik], but does not reach significance; being a good [-ik] base is negatively associated with being suffixed with [-ʧʲik]. Getting a good score from either grammar trained on [-ʧʲik] increases the chances of being suffixed with [-ʧʲik]. Good [-ʧʲik] diminutives are less likely to end up in "other", and good [-ʧʲik] bases are less likely to end up suffixed with [-ok].

5.4.2 The handpicked predictor model

This section examines the less successful model, which used handpicked phonological predictors encoding the phonological properties of bases. The purpose of examining this model is 1) to ensure that the phonological variables we manipulated in choosing the stimuli are actually having an effect, 2) to present a human-readable summary of the effects that the phonotactic differences between bases have on allomorph choice, and 3) to show how much variance still escapes the hand-picked predictor model compared to the sublexical phonotactic model in Table 3.

The dependent variable in the model was the choice of suffix ([-ʧʲik], [-ik], [-ok] and "other").

	ʧʲik	-ik	-ok	other
(Intercept)	−0.516	NA	NA	NA
suffix	NA	0.3987	−1.358	(0.01245)
ok-dim	(−24.98)	(−24.11)	**2.589**	(−24.54)
ok-base	(−0.007772)	−0.2988	(0.1333; p=0.22)	(0.1576)
ik-dim	(−0.00141)	**0.5954**	(−21.48)	−1.352
ik-base	−0.2938	(0.2277; p=0.067)	(0.0001345)	(−0.1836)
ʧʲik-dim	**0.6511**	(−0.1542)	(−22.47)	−0.7275
ʧʲik-base	**0.59**	(−0.1964)	−0.3127	(0.2793)

R^2likelihood: 0.4654; AIC: 1719; BIC: 1861

Table 3: Model for affix choice in the production study, with sublexical phonotactic grammar predictors

The predictors were analogous to the model for the lexicon reported in 2, except they also included *rating,* which ranged from 1 to 5. We included rating as a predictor because it is conceivable that people's choice of the suffix was influenced by them having to rate the form they produced (that is, the nature of the task might have affected the choice of suffix).[8] The model also included a random intercept for "word" (base noun) and a random intercept for participant, as well as a random slope for dorsal, labial, final stress, and log frequency of the base noun in the Russian National Corpus. Neither the fully crossed model nor models with more complex random effect structures converged in R.[9] We tested models that included additional predictors (initial.CC, medial.CC, lastC=glide), but they either did not reach significance in the model or prevented it from converging in R.

The model log odds are summarized in Table 4. Positive log odds mean the base is more likely to end up in the suffix's category with the feature encoded by the predictor than one might expect given the baseline likelihoods of each suffix. When the log odds are given in parentheses, their effects did not reach significance at $p < 0.05$. For example, bases with final clusters, such as [mətatsíkl] 'motorcycle', are significantly more likely to be diminutivized with [-ik] than bases without clusters are; on the other hand, they are less likely to be diminutivized with [-ok]. For [-ʧʲik], the effect goes in the right direction but does not reach significance. Other notable effects include stress: both final and medial stress decrease the likelihood of [-ok], consistent with the general lexical pattern, and increase the likelihood of the other two suffixes. The suffix [-ok] is significantly less likely to appear on words with hiatus. The quality of the vowel in the last syllable had an opposite effect than in the lexicon: in the elicitation study, round vowels are less likely to appear with [-ok] and more likely to appear with [-ik]. It is possible that the disparities between the lexicon and the elicitation study are accidental and are due to the comparatively small sample size in the study.

The sublexical phonotactic model and the handpicked predictor model cannot be compared directly based on ANOVA model comparison, but they have visibly different fit: for the sublexical

[8]We attempted to include this predictor in the sublexical model in the previous section, but the model failed to converge in R—most likely due to the high degree of collinearity between people's ratings and the sublexical phonotactic grammar scores.

[9]Fully crossed models that include random slopes for all the predictors and interactions are the most conservative (Barr et al. 2013). Polytomous allows for fitting models with random slopes and intercepts when using the *poisson reformulation* heuristic.

	other/gap	[-tʃʲik]	[-ik]	[-ok]
(Intercept)	(-0.6827)	NA	NA	NA
suffix	NA	-1.382	(0.4216)	1.355
lastC=dor	(-0.04969)	0.8437	-1.006	(0.3125)
lastC=lab	-1.193	0.3955	(-0.08215)	(-0.07677)
lastC=fric	(-0.5948)	(-0.3918)	0.3795	(-0.2776)
lastC=nasal	(0.0808)	1.641	-1.06	-0.7506
lastC=approx	(0.004423)	1.307	-0.4415	-0.3434
final.stress	(-0.2171)	0.6022	0.3345	-0.8424
medial.stress	0.6326	0.5573	0.6696	-0.9839
final.CC	(0.4179)	(-0.1452)	1.043	-1.306
lastV=high	(0.3691)	(-0.1984)	(-0.1116)	0.2876
lastV=round	(0.2513)	(0.2476)	0.389	-0.4183
hiatus	(0.005674)	(0.3691)	(0.1555)	-1.078
rating	-0.2689	(-0.009695)	-0.1595	(0.02)

R^2likelihood: 0.2129; AIC: 2490; BIC: 2753
formula: polytomous(suff ∼ finCC + dor + lab + finstr + midstr + fric + apx + nas + hiatus + lastVrd + lastVhi + rating + (1 + logfreq + dor + lab + finstr | subject) + (1 | word))

Table 4: Model for allomorph choice with handpicked phonological predictors

model, $R^2 = 0.47$, and for the handpicked model, $R^2 = 0.21$. One of the reasons for this is that the handpicked predictor model encodes only the properties of bases, not the resulting diminutives. The sublexical model has both. One could include predictors coding the phonotactic properties of derived words in the handpicked model, as well, but that is not entirely legitimate: the properties a linguist would choose as characterizing the derived words (e.g., stress is final on [-ok], stress is penultimate and palatalization applies in [-ik] words, and so on) are just as likely to be the result of adding the suffixes as be predictors of them; the cause and effect here are difficult to disentangle. In any case, it should be kept in mind that these two models are in part attempting to explain the same thing. They both use wholistic phonological properties characterizing the overall phonotactics of words undergoing affixation, but in the case of the sublexical model, the properties were discovered by a machine learner, whereas in the case of the handpicked model, they were chosen by the analysts.

5.5 Ratings

The experiment asked people to produce a diminutive for each base even if they did not like the resulting form, and this is reflected in their ratings. We present a summary of the ratings graphically here. The ratings are consistent with the affix selection patterns: people tended to chose an affix for a base if it did well in the sublexical phonotactic grammar trained on the bases and the diminutives with that affix. The plots show the ratings for words that people produced with each suffix separately, and they show the density plots for those ratings' distributions along the scale from 1 (worst) to 5 (best). For example, the rightmost upper plot shows that the forms that people produced with [-tʃʲik] got more ratings of "5" when the base passed the grammar trained on lexical [-tʃʲik] bases—the bean is thickest on top. The average rating was lower for the bases that did not pass the [tʃʲik] base grammar, shown on the left bean in that plot. The rightmost bottom plot shows

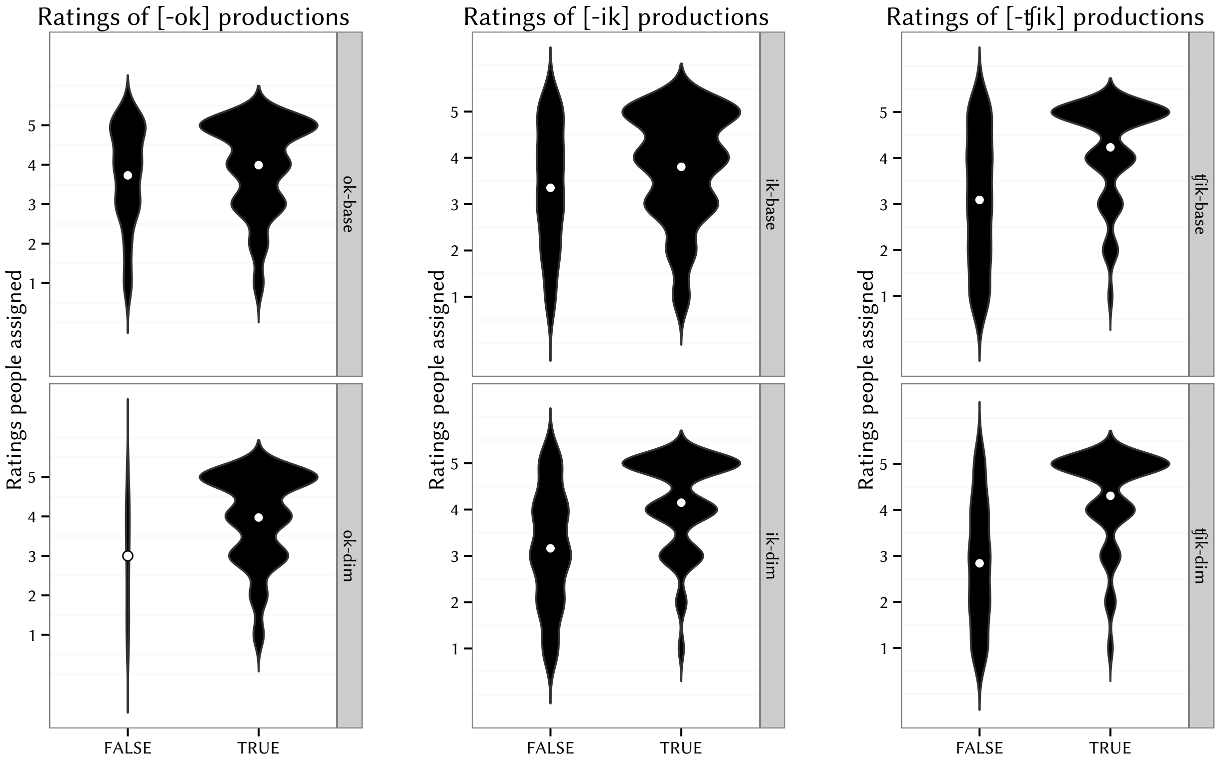


Figure 5: Ratings people gave for the words they produced with [-ok], [-ik], and [-ʧ^{j}ik], grouped by whether the bases and diminutives passed the respective affixes' sublexical phonotactic grammars (true) or failed them (false). The white dot shows the mean.

the ratings for how the same forms did in the grammar trained on [ʧ^{j}ik] diminutives.

Since the ratings tell the same story as the affix choice patterns, we do not analyze them statistically here to save space.

6 Discussion

6.1 How people do the elicitation task

We conclude that the elicitation study supports our hypothesis that people's choices of allomorphs depend on the well-formedness of the bases and the diminutive forms in sublexical phonotactic grammars. People learn these grammars over subsets of the lexicon defined by each allomorph of the diminutive, and when asked to assign a diminutive allomorph to a word that does not already have a lexical specification for diminutive formation, they consult the sublexical phonotactic grammars to make the decision.

We can model what happens in the task as follows. When a word is presented for diminutive formation, the speaker first consults the rules for diminutive formation to check if a diacritic is specified. The stem [barsúk] 'badger' is in the [+ok] sublexicon, so its diacritic is [+ok] and the decision is straightforward:

(22) If the stem is specified for a diacritic, the lexical rule applies:

Input: [barsúk]

Lookup: DIM↔[-ok] /[+ok] | {barsuk$_{+\text{ok}}$, ...}

To phonology: /barsuk-ok/

Phonological output: [bərsutʃʲók]

On the other hand, if the stem is not in a sublexicon, people need to decide on a diacritic for it. To do this, they run each base through the sublexical phonotactic grammars for bases, and they construct diminutives with each allomorph and run them through the sublexical phonotactic grammars for diminutives. In order to do the latter, they need to force the realization rules for each diminutive allomorph to apply to the base and phonologize the resulting stem-affix combinations. There are three possible scenarios: (i) the stem does well in more than one grammar, in which case we expect variation; (ii) the stem does well in the grammars for one allomorph, in which case it will get that allomorph; (iii) the stem does poorly in all of the grammars, in which case expect either no output or an output that people rate poorly. Let us illustrate each scenario with examples.

A word that did well in just one of the subgrammars is [manʲják] 'serial killer'. Of the 8 people who got this prompt, seven produced [mənʲjatʃʲók], rating it an average 4.12. One person produced [manʲjátʃʲik] and rated it a 3. Our subgrammars predict that [-ok] should be the allomorph for this word—it does rather well in both the base and the diminutive subgrammars (simulated by the UCLA Phonotactic Learner; the numbers in the tables below are the violation scores it assigned). The [-ik] form does poorly in the [-ik] base grammar and in the [-tʃʲik] base grammar.[10] The success in the [-ok] grammar determines the course pretty clearly, despite the good performance of [manʲjátʃʲik] in the diminutive grammar for [-ik]. There are diminutives in the [-ik] sublexicon that sound like this (e.g., [mʲátʃʲik] 'ball'), just none that are derived from dorsal-final stems. Our analysis captures this by allowing both the base grammar and the diminutive grammar to contribute to the decision to give this stem the [-ok] diacritic.

(23) A stem that does well in just one subgrammar: [manʲják]

	[-ok]	[-ik]	[-tʃʲik]
Base subgrammars	0	14.211	9.302
forms evaluated:	[manʲják]	[manʲják]	[manʲják]
Derived subgrammars	0	0	0
forms evaluated:	[manʲjatʃʲók]	[manʲjátʃʲik]	[manʲjáktʃʲik]
Cumulative	0	14.211	9.302

The most likely rule: DIM ↔[-ok] | [+ok]

The stem [alʲiándr] 'oleander' exhibits a bit more variation. Of the seven people who got it as a prompt, five people diminutivized it with [-ik] (mean rating 4), and two gave it the [-tʃʲik] allomorph (mean rating 3). This is in line with how the sublexical phonotactic grammars evaluated the bases and corresponding diminutives that people produced. The [-ok] base grammar assigned it a high violation score; the base grammar for [-ik] liked it, but it got some violations in the [tʃʲik]

[10]In the [-ik] base grammar, [manʲják] fails several medial cluster constraints and *[−low]#, which penalizes final dorsals. It fails very similar constraints in the [tʃʲik] base grammar.

base grammar.[11] Since people did not produce any [-ok] diminutives for [alʲiándr], we do not have a subgrammar score for it, but both of the other derived subgrammars assigned good scores to the corresponding diminutives. The end result is that the most likely diacritic for this base is [-ik]:

(24) A stem that does well in one of the grammars: [alʲiándr] 'oleander'

	[-ok]	[-ik]	[-ʧʲik]
Base subgrammars	11.515	0	4.292
forms evaluated:	[olʲeándr]	[olʲeándr]	[olʲeándr]
Derived subgrammars	3.29	0	0
forms evaluated:	[olʲeandrók]	[olʲeándrʲik]	[olʲeándrʧʲik]
Cumulative	14.805	0	4.292

The most likely rule: DIM ↔[-ik]/[+ik]

Let's now consider bases that people found difficult to diminutivize. One such base is [lʲémʲink] 'lemming'. It has all the properties that Polivanova (1967) associates with "impossible" diminutives: a word-final consonant cluster that ends in a dorsal and non-final stress. Ten people got this prompt, and they produced a variety of forms, all of which they rated poorly (mean rating 2.4). There was one gap, and the remaining productions were one each of [lʲémʲingʲik] and [lʲemʲingók], two [lʲémʲingʧʲik], and a number of "other" productions ([lʲemʲingaʧʲók], [lʲémʲingʲiʂkə], [lʲemʲingóʧʲik], [lʲémʲinguʂkə], [lʲimʲingʧʲók]). The subgrammars for the actual productions with these affixes generally give this base poor scores, with the exception of the [-ok] base grammar.

(25) A stem that has no diminutive diacritic and does poorly in all the sublexicons: [lʲémʲink] 'lemming'

	[-ok]	[-ik]	[-ʧʲik]
Base subgrammars	0	5.138	6.854
forms evaluated:	[lʲémʲink]	[lʲémʲink]	[lʲémʲink]
Derived subgrammars	4.439	0.905	2.167
forms evaluated:	[lʲimʲingók]	[lʲémʲingʲik]	[lʲémʲinkʧʲik]
Cumulative:	4.439	6.043	9.021

A form Polivanova predicts to be "difficult" rather than impossible is [aftóbus] 'bus'. Our participants mostly converged on [aftóbusʲik] (7 people, mean rating 4.375), with one person contributing [aftəbusók] (rated 1). This outcome is not surprising, considering how poorly the base does in the sublexicons for [-ok]. Nobody produced [aftóbusʧʲik], so we did not train the UCLA Phonotactic Learner on this form—but it would do poorly in the [-ʧʲik] derived subgrammar, which contains constraints against stress lapses.

[11] The key violations in the [-ok] grammar: *[−high][−high] for the [ea] vowel sequence, and *[−approx][+cont]# and *[−syll][+son] for the final cluster. The one violation in the [-ʧʲik] grammar is for the vowel sequence; the constraint is also *[−high][−high].

(26) A stem that has no diminutive diacritic and does poorly in all the sublexicons: [aftóbus] 'bus'

	[-ok]	[-ik]	[-ʧʲik]
Base subgrammars	8.738	4.05	5.2
forms evaluated:	[aftóbus]	[aftóbus]	[aftóbus]
Derived subgrammars	19.509	0.905	2.167
forms evaluated:	[aftobusók]	[aftóbusʲik]	[aftóbusʧʲik]
Cumulative:	28.247	4.955	7.367

It should be kept in mind that the numbers generated by the UCLA Phonotactic Learner subgrammars cannot be compared or interpreted in absolute terms. The numbers merely indicate that there are violations—but not whether these violations are critical to the success of an individual base or diminutive form. Thus, the violations of [aftobusók] in the [-ok] subgrammars, at 28.247, dwarf the violations of [lʲemʲingók], at 4.439. Conversely, the violations of [lʲémʲingʲik] are only slightly higher than those of [aftóbusʲik], and yet the latter form is clearly far less problematic for people. We believe this is a side effect of this particular implementation rather than a problem for the theory, however. The first possible implementational bug is that our training sets are based on automated searches of a large corpus—not actual sublexicons that people have in their heads. Partly, we did this to minimize our own biases; it is apparent that the intuitions of linguists can diverge from what people actually do (cf. Polivanova's labeling of [aftóbusʲik] as "bad" compared to the people's mean rating of 4.375 out of 5). Thus, our sublexicons are not a perfect model of what people learn. The second source of implementational artifacts is the UCLA Phonotactic Learner itself, which is based on a specific theory of phonotactic learning. We did not assist the learner in any way, so it actually missed some of the generalizations about the sublexicons that we identify in our statistical analyses in section 4.2. This problem is connected to the first problem—the grammars are only as good as the sublexicons they are trained on, and if the sublexicons are small, some generalizations are going to go unnoticed by this learner even if they are psychologically real to people.

We note, however, that quite a lot of what happened in the experiment suggests that our analysis is on the right track. People produced many odd forms, some that arguably should not even be generated by the grammar of diminutive formation if it uses rules with phonological contexts. If rules of the form DIM $\leftrightarrow$[ok]/Vk__ really existed, then [manʲjá ĩk] should never have been generated. We interpret this as a stage in the diacritic assignment procedure: in order to decide how to diminutivize a word that isn't listed in a diacritically defined sublexicon, people have to generate some forms by forcing the rules to apply. The productions were particularly diverse with forms that fail multiple subgrammars, like [lʲémʲink] and [súfʲiks]; more often than not, people produced each of [-ik], [-ok], and [-ʧʲik] in addition to gaps and forms with other suffixes for these stems. This creativity is partly due to the high-pressure nature of the elicitation task—but our theory of generalization of phonological selectional restrictions accounts for this behavior without additional mechanisms. We now turn to some alternative approaches to phonologically selective allomorphy.

6.2 A comparison with alternatives

There are many approaches to phonological selectional restrictions, phonologically conditioned suppletive allomorphy, and phonologically predictable morphology—both in formal/generative traditions and in connectionist and language use-based approaches. We focus on formal theories here.

6.2.1 The emergence of the unmarked

The first approach we consider is cast in Optimality Theory and uses UNIVERSAL MARKEDNESS CONSTRAINTS in an emergence of the unmarked ranking (TETU, McCarthy and Prince 1994, Mascaró 1996, Kager 1996). For English indefinite allomorphy, the choice would be handled by ONSET, which disprefers "a" before "apple", and NOCODA, which disprefers "an" before "pear". These constraints dominate no faithfulness in English but still have an effect in the grammar by selecting between allomorphs. The dispreference of the Russian allomorph [-ik] against dorsal-final stems could be explained in these terms, possibly by invoking a constraint against dorsal-[i] sequences—this constraint might even be independently motivated in Russian, since other [i]-initial suffixes cause dorsal mutation (see Padgett 2003, Kapatsinski 2010). The main problem with this approach is well-known—there are cases that are markedness-neutral or go contrary to markedness (see Paster 2006, Wolf to appear, Embick 2010, Nevins 2011). An example of the latter type is Haitian Creole definite allomorphy, conditioned in the opposite way of the English indefinite: [papa-a] 'the father' but [pitit-la] 'the child' (Klein 2003). ONSET and NOCODA cannot make the right decision in such a case. Equally problematic are cases of homophonous morphemes that impose distinct selectional restrictions. Consider the English comparative in "tall-er", agentive in "work-er", and the demonym suffix in "New York-er". The comparative suffix prefers monosyllabic or disyllabic trochees and disfavors bases that end in consonant clusters (Hilpert 2008, Mondorf 2009, LaFave 2012, inter alia). On the other hand, the demonym "-er" does not avoid cluster-final bases and does not impose a size maximum on the base (cf. "New England-er", "Budapest-er" vs. "Ohio-an", "San Franciscan").[12] Accounting for this allomorphy and selectivity in the same constraint hierarchy would be a challenge for this approach.

The markedness approach to allomorph selection can generate unattested cases of the kind we discussed in 2.1.1, whereby phonological markedness constraints select stems based on the affixes that follow: allowing markedness to select stem allomorphs in parallel with suffix allomorphs predicts the *go-ed/went-ing* allomorphy pattern. A markedness constraint such as ONSET will prefer the *went-* allomorph over *go-* when the following affix is vowel-initial. In our theory, the overall phonotactic well-formedness of the derived word matters only in deciding whether the stem can be assigned a particular diacritic—since the stem is already given, its shape cannot be retroactively altered after consulting the sublexical grammar for derived phonological words containing a particular suffix.[13]

Another difficulty for this markedness-based approach, noted by Paster (2006) and others, is that surface constraints cannot capture input-oriented selectional restrictions. In Haitian Creole, a glide is inserted between a tense vowel and [-a]: [papje]/[papjej-a] 'the paper', but stem-final glides condition [-la], like other consonants ([bagaj]/[bagaj-la] 'the thing'). In Russian, the [-ik] allomorph's avoids dorsal-final bases, and [-ok] avoids medial stress bases—even though they are not stress-medial once the suffix is attached. Even in English "a/an", the surface syllabification does not follow the ONSET/NOCODA generalization (Garellek 2012). The properties relevant for affix insertion hold only of bases and are bled by affixal phonology, so some generalizations need

[12] Unlike the comparative suffix, English demonym allomorphy does not appear to be nearly as well studied. Gordon (2014) and Raffelsiefen (1999) do identify some generalizations.

[13] Suppose, however, that the learner is exposed to a language with morphosyntactically conditioned stem allomorphy of the *person/people* variety, and they start forming sublexical phonotactic grammars over the types of items that display this allomorphy. We speculate that this is how people extract generalizations that have been captured by stem readjustment rules in Distributed Morphology: if there are phonological generalizations to be found and the context for the alternations is morphosyntactic, sublexical grammars should be able to capture them. Readjustment rules have come under criticism recently (Harley and Tubino Blanco 2012, Haugen and Siddiqi 2013) so alternatives to them are a welcome result.

to be stated over bases. This is why in our theory, base sublexical phonotactics can play a role just as derived sublexical phonotactics, and insertion rules really only make reference to specific lexical items, via diacritics. The phonological generalizations are important for extending the rules to new items, but the morphological grammar is robust even when the phonological generalizations are opaque.

6.2.2 Subcategorization frames and alignment

Another approach to selectional restrictions casts morpheme realization rules in terms of SUBCATEGORIZATION FRAMES that reference phonological and lexical information (Lieber 1980, Paster 2006, Bye 2007, Embick 2010). When the realization of an affix is phonologically predictable, the rule specifies the context. For the Haitian Creole definite allomorphy, the rule would look as in (27); for English indefinite allomorphy, the contexts would be similar (recall (5)). This approach can be recast in OT using Generalized Alignment constraints (McCarthy and Prince 1993), which may state that an affix prefers to be adjacent to a particular structure such as a stressed syllable, or a consonant. Klein (2003) uses alignment for Haitian Creole, noting that in [liv-la], the stem can be aligned to the syllable boundary, whereas in *[liv-a], ONSET would favor resyllabification and alignment to syllables would be lost.

(27) SUBCATEGORIZATION FRAMES (following the discussion of Haitian Creole in Embick 2010)

a. DEF$\leftrightarrow$ a / V ___

b. DEF$\leftrightarrow$ la / C ___

(28) GENERALIZED ALIGNMENT for Haitian Creole (Klein 2003)

R-ALIGN-STEM-SYLL: 'Align the right edge of the stem with the right edge of a syllable'

The virtue of this approach is that it divorces phonological selection from universal markedness. But it has problems, as well. They mainly come down to dealing with lexical variability and stating complex phonological contexts. The issue of lexical variability arises even in the simple case of English "a"/"an" allomorphy: there are exceptions, which may or may not be phonologically systematic depending on speaker and dialect. To deal with sporadic exceptions, the subcategorization approach could add item-specific rules (e.g., DEF$\leftrightarrow$ən/___*historical*) and assume that these rules take precedence over the more general rules (see Halle and Marantz 1993, Halle and Idsardi 1997).

But when such exceptions form a phonological class, the rules become harder to state. The rule for English [ən] would refer to a disjunctive context: before vowel-initial words and [h]-initial words. The second problem arises when such rules are attempted for complex conditioning environments. The environment for Russian diminutive allomorphs [-ik] is relatively easy to state in rule format: it occurs after stressed syllables that do not end in dorsals (although there are, of course, exceptions to the stress generalization). The environment for the diminutive allomorphs [-ok] and [-ʧʲik], on the other hand, is hard to state in this form. Recall that [-ok] tends to favor initial or final stress but not medial stress; it also imposes restrictions on vowel hiatus. The unwieldy rule in (29b) does not even succeed in capturing these environments, since they are constraints on what cannot occur in these bases rather than a list of what can occur. To wit, specifying that the penultimate vowel must be unstressed is actually wrong, since it could be the first vowel of the word, in which case it can be stressed (there are lexical diminutives such as [sókəl]/[səkalók] 'falcon'). Likewise, capturing the hiatus restriction would require listing the entire string, since it can be somewhere other than the last two vowels of the stem. The rule for [ʧʲik] fares no better. We can capture the avoidance of final consonant clusters and the preference for final stress, but the prohibition against stridents and coronal obstruents is difficult, since final consonants can be coronal ([lʲimóntSik] 'lemon-dim', [pʲinʲálʲʧʲik] 'pencil box-dim') as long as they are not obstruent. Thus, subcategorization frames have a hard time capturing negative constraints on insertion (see Wolf 2008:106–7 for discussion).

(29) Russian subcategorization frames for diminutives (straw man)

a. DIM↔[-ik]/$V_{[+\text{stress}]}C_{[-\text{dor}]}$——

b. DIM↔[-ok]/C+$V_{[-\text{stress}]}$C+VC___

c. DIM↔[ʧʲik]/$V_{[+\text{stress}]}C_{[\text{-str, -cor}]}$——

Alignment constraints do not fare much better, since they are inherently edge-oriented in their formalization. What is the prosodic constituent that [-ok] is aligning itself to? There is no generalization that would unify $\sigma\sigma\acute{\sigma}$and $\acute{\sigma}\sigma\sigma$ to the exclusion of $\sigma\acute{\sigma}\sigma$. The same problem comes up in other cases where prosodic constituency is thought to be important, like the English comparative affix. If the comparative affix is aligning itself to a trochaic foot, then an initial unstressed syllable or an extra foot before the main stress foot should not prevent suffixation. Alignment to a foot is satisfied equally well in the impossible **concreter (kàn)(kɹíɹ̩)ɹ̩* and in the attested *bigger (bíg)ɹ̩*. On the other hand, the generalization about adjectives that combine with "-er" is easy to state phonotactically: they may not have unstressed initial syllables, and they may not have more than one stress per word.

6.2.3 The Minimal Generalization Learner

The final alternative we consider is The Minimal Generalization Learner (Albright and Hayes 2003). This learner shares a few features with our approach. It learns by keeping track of lexical examples of affixation, and it looks for phonological generalizations over bases that follow the same rule. The learner also relies on type frequencies within each class to decide how reliable the generalizations are. Unlike our approach, the learner induces not only phonological generalizations about bases but also the changes that each base undergoes in affixation. Thus, given learning data such as (30), it should notice the palatalization of final consonants and unify the changing substrings in a single rule, since the consonants can be sonorant or obstruent, and coronal or labial. The learner is "minimal" in that it looks for the smallest environment that will suffice for a generalization; the intuition behind this is that factors conditioning allomorphy are generally local.

(30) Some learning data for the Minimal Generalization Learner

a.	kúst	kústʲik	'bush'	The rule:
b.	klʲón	klʲónʲik	'maple'	
c.	klóp	klópʲik	'bedbug'	$C_{[-\text{dor}]}\rightarrow C_{[-\text{back}]}$ik
d.	komár	komárʲik	'mosquito'	
e.	gnóm	gnómʲik	'gnome'	

This local approach works well for cases such as the English past tense suffix, but not for Russian diminutives. We ran some Minimal Generalization Learner simulations on our sublexicons, and the learner posited several rules for [-ik] and [ʧʲik] (though none as general as the one above; most were specific to individual consonants). It failed to notice any generalizations about the [-ok] suffix, and it is not hard to see why. The string that changes in the mapping from [sókol] to [sokolók] is quite large—everything after [s] is changing. Correspondingly, the learner posited rules such as [ókol] →[okolók]/s___, and failed to make generalizations over the various words in which stress moves long-distance. The learner missed even the exceptionless generalization that dorsal-final words diminutivize with [-ok] as long as the dorsal not in a cluster.

To give the Minimal Generalization a better chance of finding generalizations, we removed stress from the learning data. It found the dorsal-final pattern for [-ok] this time, but of course it missed the stress generalization, which is a major factor in how Russian words diminutivize (clear both from Polivanova's 1967 observations and our own). The learner also had no way of mimicking the difficulty that people had in diminutivizing words with hiatus, such as [klóun] 'clown'. Once it finds a local generalization about the word-final consonant, it does not look any further to the left. Thus, there was not much correspondence between the Minimal Generalization Learner's affix choices for our experimental bases and people's choices.

The sublexical phonotactic approach is a more powerful approach than the Minimal Generalization Learner: a phonotactic learner will find generalizations about word-final consonants, since that is what bases for suffix allomorphs usually have in common. But it will also find generalizations about overall word shape, which is needed for Russian diminutivization.

7 Conclusion

We argued that phonologically predictable affixation arises when learners make phonotactic generalizations over subsets of the lexicon defined by morphological rules. The morphological rules refer to sets of bases via diacritics, and people learn what words in each diacritically defined subset sound like. People extend morphological rules by checking whether the base and the derived word could belong to the diacritically defined set. They do this by using sublexical phonotactic grammars to assign probabilities to bases and derived words. This approach is powerful and can handle cases in which there is variation between allomorphs, as well as cases of ineffability, where no good suffixed forms can be constructed for a base. The approach also predicts that generalizations about affix distribution can be made over the locus of the affix—for example, phonologically selective suffixes will be picky about word ends, not necessarily about word beginnings—but it is also possible for the phonological predictors of affixation to be non-local.

We demonstrated this approach on a case that has not received much attention from western phonologists, Russian diminutive allomorphy. We showed that people can generalize the suffixes to bases that lack a lexical specification for diminutive formation, and that people's decisions about affix formation mirror the generalizations that the UCLA Phonotactic Learner extracted over sublexicons of Russian bases and derived diminutives. We discussed alternatives to our theory that rely on the emergence of the unmarked and subcategorization frames, which have difficulties with certain aspects of this complex pattern. We also demonstrated that the Minimal Generalization Learner fails for cases such as Russian, where the generalizations are not minimal.

References

Adriaans, Frans, and René Kager. 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language* 62:311–331.

Albright, Adam. 2008. How many grammars am I holding up? Discovering phonological differences between word classes. In *Proceedings of the 26th West Coast Conference on Formal Linguistics*, ed. Charles B. Chang and Hannah J. Haynie, 1–20. Somerville, MA: Cascadilla Proceedings Project.

Albright, Adam. 2009a. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26:9–41.

Albright, Adam. 2009b. Lexical and morphological conditioning of paradigm gaps. In *Modeling ungrammaticality*, ed. Curt Rice and Sylvia Blaho, 117–164. London: Equinox Publishing.

Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90:119–161.

Albright, Adam, and Bruce Hayes. 2006. Modeling productivity with the gradual learning algorithm: The problem of accidentally exceptionless generalizations. In *Gradience in grammar: Generative perspectives*, ed. Gisbert Fanselow, Caroline Fery, Matthias Schlesewsky, and Ralf Vogel, 185–204. Oxford: Oxford University Press.

Alderete, John. 1999. Morphologically-governed accent in Optimality Theory. Doctoral Dissertation, University of Massachusetts, Amherst, Amherst, MA.

Arppe, Antti. 2013. Polytomous: Polytomous logistic regression for fixed and mixed effects. *R package version 0.1.6* URL `http://CRAN.R-project/org/package=polytomous`.

Baayen, R Harald, Richard Piepenbrock, and Rijn van H. 1993. The {CELEX} lexical data base. {CD-ROM}.

Barr, Dale, Roger Levy, Christoph Scheepers, and Hal Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255–278.

Becker, Michael. 2009. Phonological trends in the lexicon: The role of constraints. Doctoral Dissertation, University of Massachusetts, Amherst, MA.

Becker, Michael, and Maria Gouskova. 2012. Source-oriented generalizations as grammar inference in Russian vowel deletion. Ms. Indiana University and NYU. Available on LingBuzz at `http://ling.auf.net/lingbuzz/001622`.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14:150–177.

Bobaljik, Jonathan. 2000. The ins and outs of contextual allomorphy. In *University of Maryland working papers in linguistics*, ed. K. K. Grohmann and Caro Struijke, volume 10, 35–71. College Park, MD: University of Maryland.

Booij, Geert. 2011. Morpheme structure constraints. In *The blackwell companion to phonology*, ed. Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice, volume 4, xxx–xxx. Wiley-Blackwell.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10:425–455.

Bye, Patrik. 2007. Allomorphy: Selection, not optimization. In *Freedom of Analysis?*, ed. Sylvia Blaho, Patrik Bye, and Martin Krämer, Studies in Generative Grammar. Berlin: Mouton de Gruyter.

Calamaro, Shira, and Gaja Jarosz. 2014. Learning general phonological rules from distributional information: A computational model. *Cognitive science* .

Carstairs, Andrew. 1988. Some implications of phonologically conditioned suppletion. In *Yearbook of morphology 1988*, ed. Geert Booij and Jaap van Marle, 67–94. Dordrecht: Kluwer.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Coetzee, Andries, and Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 31:47–89.

Coetzee, Andries, and Joe Pater. 2011. The place of variation in phonological theory. In *The Handbook of Phonological Theory*, ed. John Goldsmith, Jason Riggle, and Alan Yu, 401–434. Oxford: Blackwell, 2nd edition edition. Available as ROA-946 on the Rutgers Optimality Archive, http://roa.rutgers.edu.

Crosswhite, Katherine. 1999. Vowel reduction in Optimality Theory. Doctoral Dissertation, UCLA, Los Angeles, CA.

Dmitrieva, Olga. 2012. Geminate typology and the perception of consonant duration. Doctoral dissertation, Stanford University.

Embick, David. 2010. *Localism versus globalism in morphology and phonology*. Cambridge, MA: MIT Press.

Embick, David, and Alec Marantz. 2008. Architecture and blocking. *Linguistic Inquiry* 39:1–53.

Fukazawa, Haruka, Mafuyu Kitahara, and Mitsuhiko Ota. 1998. Lexical stratification and ranking invariance in constraint-based grammars. In *Cls 32, part 2: The panels*, ed. M. Catherine Gruber, Derrick Higgins, Kenneth Olson, and Tamra Wysocki, 47–62. Chicago, IL: Chicago Linguistic Society.

Gabrielatos, Costas, Eivind Nessa Torgersen, Sebastian Hoffman, and Susan Fox. 2010. A corpus-based sociolinguistic study of indefinite article forms in London English. *Journal of English Linguistics* 38:297–334.

Garellek, Marc. 2012. Word-initial glottalization and voice quality strengthening. In *UCLA Working Papers in Phonetics*, volume 110, 1–23. UCLA.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the stockholm workshop on variation within Optimality Theory*, ed. Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–120. Stockholm: Stockholm University.

Gordon, Samantha. 2014. What do you call a person from . . . . In *Annual Conference on Phonology*. MIT, Cambridge, MA.

Gouskova, Maria. 2012. Unexceptional segments. *Natural Language and Linguistic Theory* 30:79–133.

Gouskova, Maria, and Michael Becker. 2013. Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language and Linguistic Theory* 31:735–765. Available on Lingbuzz at `http://ling.auf.net/lingBuzz/001456`.

Gouskova, Maria, and Tal Linzen. to appear. Morphological conditioning of phonological regularization. *The Linguistic Review* URL `http://lingauf.net/lingbuzz/002246`.

Halle, Morris. 1973a. The accentuation of Russian words. *Language* 49:312–348.

Halle, Morris. 1973b. Prolegomena to a theory of word formation. *Linguistic Inquiry* 4:3–16.

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

Halle, Morris, and William Idsardi. 1997. r, hypercorrection, and the Elsewhere Condition. In *Derivations and constraints in phonology*, ed. Iggy Roca, 331–348. Oxford: Clarendon Press.

Halle, Morris, and Alec Marantz. 1993. Distributed Morphology. In *The view from building 20. Essays in honor of Sylvain Bromberger*, ed. Kenneth Hale and Samuel Jay Keyser. Cambridge, MA: MIT Press.

Harley, Heidi, and Mercedes Tubino Blanco. 2012. Cycles,vocabulary items and stem forms in Hiaki. Ms. University of Arizona.

Haugen, Jason D, and Daniel Siddiqi. 2013. Roots and the derivation. *Linguistic inquiry* 44:493–517.

Hay, Jennifer, and Harald Baayen. 2002. Parsing and productivity. *Yearbook of morphology (2001)* 35:203–236.

Hayes, Bruce. 2009. *Introductory phonology*. Malden, MA, and Oxford, UK: Wiley-Blackwell.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.

Hermans, Ben. 2001. Yer vocalization in suffixed forms; the case of Russian. *Progress in Grammar: Articles at the 20th Anniversary of the Comparison of Grammatical Models Group in Tilburg* .

Hermans, Ben. 2002. Overapplication of yer vocalization in Russian. In *Linguistics in the Netherlands*, ed. Hans Broekhuis and Paula Fikkert, volume 19, 85–95. John Benjamins.

Hilpert, Martin. 2008. The English comparative—language structure and language use. *English Language and Linguistics* 12:395–417.

Hippisley, Andrew. 1996. Russian expressive derivation: A network morphology account. *The Slavonic and East European Review* 74:201–222.

Iosad, Pavel, and Bruce Morén-Duolljá. 2010. Russian palatalization: The true(r) story. Talk presented at the Old World Conference in Phonology 7, University of Nice, Nice, France. Available at `http://www.anghyflawn.net/static/pdf/ocp7.slides.pdf.pdf`.

Ito, Junko, and Armin Mester. 1995. The core-periphery structure of the lexicon and constraints on reranking. In *Papers in Optimality Theory II (University of Massachusetts occasional papers in linguistics)*, ed. Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 181–210. Amherst, MA: GLSA Publications.

Jurgec, Peter. 2010. Disjunctive lexical stratification. *Linguistic Inquiry* 41:149–161.

Kager, René. 1996. On affix allomorphy and syllable counting. In *Interfaces in phonology*, ed. Ursula Kleinhenz, 155–171. Berlin: Akademie Verlag.

Kapatsinski, Vsevolod. 2010. Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Laboratory Phonology* 1:361–393.

Kawahara, Shigeto, Kohei Nishimura, and Hajime Ono. 2002. Unveiling the unmarkedness of Sino-Japanese. *Japanese/Korean Linguistics* 12:140–151.

Kempe, Vera, Patricia Brooks, and Natalija Mironova. 2003. Diminutivization supports gender acquisition in Russian children. *Journal of Child Language* 30:471–485.

Klein, Thomas B. 2003. Syllable structure and lexical markedness in creole morphophonology: Determiner allomorphy in haitian and elsewhere. In *The phonology and morphology of creole languages*, ed. Ingo Plag, 209–228. Tubingen: Max Niemeyer Verlag.

LaFave, Nathan. 2012. A more experimental investigation of English adjective gradation. Paper presented at NWAV 41, Indiana University, Bloomington, IN. URL `https://files.nyu.edu/nbl225/public/LaFave%20English%20adj%20gradation%20poster.pdf`.

Lieber, Rochelle. 1980. On the organization of the lexicon. Doctoral Dissertation, MIT, Cambridge, MA.

Lightner, Theodore. 1965. Segmental phonology of Modern Standard Russian. Doctoral Dissertation, MIT, Cambridge, MA.

Mascaró, Joan. 1996. External allomorphy as emergence of the unmarked. In *Current trends in phonology: Models and methods*, ed. Jacques Durand and Bernard Laks, 473–483. Salford, Manchester: European Studies Research Institute, University of Salford.

Mascaró, Joan. 2007. External allomorphy and lexical representation. *Linguistic Inquiry* 38:715–735.

McCarthy, John J., and Alan Prince. 1993. Generalized alignment. In *Yearbook of morphology*, ed. Geert Booij and Jaap van Marle, 79–153. Dordrecht: Kluwer.

McCarthy, John J., and Alan Prince. 1994. The emergence of the unmarked: Optimality in prosodic morphology. In *Proceedings of the North East Linguistic Society 24*, ed. Merce Gonzalez, 333–379. Amherst, MA: GLSA Publications.

Melvold, Janis. 1989. Structure and stress in the phonology of Russian. Doctoral Dissertation, MIT, Cambridge, MA.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331:176–182.

Mondorf, Britta. 2009. *More support for more-support: The role of processing constraints on the choice between synthetic and analytic comparative forms*, volume 4. Amsterdam: John Benjamins Publishing.

Myler, Neil. to appear. Exceptions to the mirror principle and morphophonological "action at a distance": The role of "word"-internal phrasal movement and spell out. In *The structure of words at the interfaces*, ed. Heather Newell, Maire Noonan, Glyne Piggot, and Lisa Travis. Oxford University Press.

Nevins, Andrew. 2011. Phonologically-conditioned allomorph selection. 2357–2382. Wiley-Blackwell.

Padgett, Jaye. 2002. Russian voicing assimilation, final devoicing, and the problem of [v] (or, The mouse that squeaked). Ms. Santa Cruz.

Padgett, Jaye. 2003. Contrast and postvelar fronting in Russian. *Natural Language and Linguistic Theory* 21:39–87.

Paster, Mary. 2006. Phonological conditions on affixation. Doctoral Dissertation, UC Berkeley, Berkeley, CA.

Peperkamp, Sharon, Rozenn Le Calvez, Jean-Pierre Nadal, and Emmanuel Dupoux. 2006. The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition* 101:B31–B41.

Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation*, volume 28. Berlin: Walter de Gruyter.

Polivanova, Anna K. 1967. Obrazovanie umen'shitel'nyx suschestvitel'nyx muzhskogo roda. In *Russkij jazyk v natsional'noj shkole. reprinted in Anna K. Polivanova, 2008. Obscheee russkoe jazykoznanie: Izbrannye raboty*, volume 4, 8–22. Moscow: RGGU.

R Development Core Team. 2013. R: A language and environment for statistical computing. Vienna, Austria. URL `http://www.R-project.org`.

Raffelsiefen, Renate. 1999. Phonological constraints on English word formation. In *Yearbook of morphology 1998*, 225–287. Springer.

Raymond, William D., Julia A. Fisher, and Alice F. Healy. 2002. Linguistic knowledge and language performance in English article variant preference. *Language and Cognitive Processes* 17:613–662. URL `http://dx.doi.org/10.1080/01690960143000380`.

Revithiadou, Anthi. 1999. *Headmost accent wins: Head dominance and ideal prosodic form in lexical accent systems*. The Hague: Holland Academic Graphics.

Siegel, Dorothy. 1974. Topics in English Morphology. Doctoral Dissertation, MIT, Cambridge, MA.

Stankiewicz, E. 1968. *Declension and gradation of Russian substantives*. Berlin: Mouton.

Steriopolo, Olga. 2008. Form and function of expressive morphology: a case study of Russian. Doctoral Dissertation, University of British Columbia, Vancouver, BC.

Usachev, Andrei. 2004. *Fully accented paradigms from Zaliznjak's (1977) grammatical dictionary*. `http://dict.buktopuha.net/all_forms.rar`.

Wolf, Matthew. 2008. Optimal interleaving: Serial phonology-morphology interaction in a constraint-based model. Doctoral Dissertation, University of Massachusetts, Amherst, Amherst, MA.

Wolf, Matthew. 2013. Candidate chains, unfaithful spellout, and outwards-looking phonologically-conditioned allomorphy. *Morphology* 23:145–178.

Wolf, Matthew. to appear. Lexical insertion occurs in the phonological component. In *Understanding allomorphy: Perspectives from optimality theory*, ed. M. Eulalia Bonet, Maria-Rosa Lloret, and Joan Mascaró. London: Equinox. Available as ROA-912 on the Rutgers Optimality Archive, `http://roa.rutgers.edu`.

Yearley, Jennifer. 1995. Jer vowels in Russian. In *Papers in Optimality Theory II (University of Massachusetts occasional papers in linguistics)*, ed. Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 533–571. Amherst, MA: GLSA Publications.

Zaliznjak, Andrej Anatoljevich. 1977. *Grammatičeskij slovar' russkogo jazyka. [A grammatical dictionary of the Russian language]*. Moscow: Russkij Jazyk.

Zaliznjak, Andrej Anatoljevich. 1985. *Ot praslavjanskoj akcentuacii k russkoj. [From Proto-Slavic to Russian accentuation.]*. Moscow: Nauka.

Appendix I: Items for the Elicitation Study

Stimuli used in the elicitation study. Stress, stress type, and yer status are shown with subscripts: "y" for "yer", "A" for fixed stress, "B" for final stress, and "C" for mobile (collapsing over Zaliznjak's (1977) types C:F). Voicing alternations and vowel reduction are not transcribed; we show pseudo-orthographic representations similar to the Cyrillic words seen by our participants. The learning simulations were run on similar pseudo-orthographic representations.

abaz̨úr$_A$	'lampshade'	vólos$_C$	'hair'	klúb$_A$	'club'
abor^{j}ig^{j}én$_A$	'aborigine'	víbor$_A$	'choice'	koz̨úx$_B$	'casing, cover'
aváns$_A$	'advance'	vímp^{j}el$_A$	'pendant'	koktéjl^j$_A$	'cocktail'

av^{j}iátor$_{A}$	‘aviator’
avtóbus$_{A}$	‘bus’
adápt^{j}er$_{A}$	‘adapter’
adjutánt$_{A}$	‘aide-de-camp’
áist$_{A}$	‘stork’
ajfón$_{A}$	‘iPhone’
akvár^{j}ium$_{A}$	‘aquarium’
anátom$_{A}$	‘anatomist’
antrákt$_{A}$	‘intermission’
anʧóus$_{A}$	‘anchovy’
arxáng^{j}el$_{A}$	‘archangel’
ast^{j}eróid$_{A}$	‘asteroid’
astronóm$_{A}$	‘astronomer’
átom$_{A}$	‘atom’
bár^{j}in$_{C}$	‘landowner’
barsúk$_{B}$	‘badger’
basséjn$_{A}$	‘pool’
bát^{j}ik$_{A}$	‘batik’
b^{j}érkut$_{A}$	‘golden eagle’
b^{j}idón$_{A}$	‘can’
b^{j}il^{j}járd$_{A}$	‘billiards’
b^{j}ind^{j}úg$_{B}$	‘lever in a mill’
b^{j}inókl^{j}$_{A}$	‘binoculars’
b^{j}ís^{j}er$_{A}$	‘small beads’
b^{j}ít͡seps$_{A}$	‘bicep’
boks^{j}ór$_{A}$	‘boxer’
burgom^{j}ístr$_{A}$	‘burgomaster’
valúx$_{B}$	‘wether’
vál^{j}dʂnep$_{A}$	‘woodcock’
v^{j}ést^{j}ern$_{A}$	‘western’
v^{j}ík^{j}ing$_{A}$	‘viking’
v^{j}ixór$_{yB}$	‘cowlick, tress’
fl^{j}íg^{j}el^{j}$_{A}$	‘weathervane’
fórum$_{A}$	‘forum’
xóbot$_{A}$	‘elephant trunk’
xoz^{j}ájin$_{A}$	‘host, owner’
xómut$_{A}$	‘yoke’
t͡sem^{j}ént$_{A}$	‘cement’
ʧabán$_{B}$	‘shepherd’
ʧér^{j}ep$_{A}$	‘skull’
ʧert^{j}óz$_{B}$	‘technical sketch’
ʧetv^{j}érg$_{B}$	‘Thursday’
ʧexól$_{yB}$	‘cover’
ʧuvák$_{B}$	‘dude’
ʂat^{j}ór$_{yB}$	‘tabernacle’
ʂatún$_{B}$	‘connecting rod’
ʂél^{j}ep$_{C}$	‘whip’
ʂkol^{j}ár$_{A}$	‘scholar’
garáʐ$_{B}$	‘garage’
g^{j}epárd$_{A}$	‘cheetah’
g^{j}ért͡sog$_{A}$	‘duke’
golk^{j}íp^{j}er$_{A}$	‘goalie’
gólub^{j}$_{C}$	‘pigeon’
golíʂ$_{B}$	‘pebble’
gusár$_{A}$	‘hussar’
gusl^{j}ár$_{A}$	‘gusli player’
d^{j}év^{j}er^{j}$_{C}$	‘husband’s bro’
dʐémp^{j}er$_{C}$	‘jumper, sweater’
d^{j}il^{j}iʐáns$_{A}$	‘stagecoach’
d^{j}inozávr$_{A}$	‘dinosaur’
d^{j}ir^{j}iʐábl^{j}$_{A}$	‘blimp’
dúpl^{j}eks$_{A}$	‘duplex’
jep^{j}ískop$_{A}$	‘bishop’
ʐólob$_{C}$	‘chute, trough’
ʐólud^{j}$_{C}$	‘acorn’
ʐémʧug$_{C}$	‘pearl’
ʐér^{j}ex$_{C}$	‘asp (fish)’
ʐurávl^{j}$_{B}$	‘crane (bird)’
ídol$_{A}$	‘idol’
kadík$_{B}$	‘Adam’s apple’
kazák$_{B}$	‘cossack’
kamzól$_{A}$	‘camisole’
kamíʂ$_{B}$	‘cattail, reed’
kann^{j}ibál$_{A}$	‘cannibal’
káp^{j}ers$_{A}$	‘caper’
karás^{j}$_{B}$	‘crucian’
kaskad^{j}ór$_{A}$	‘stuntman’
k^{j}is^{j}él^{j}$_{B}$	‘kissel’
kl^{j}év^{j}er$_{C}$	‘clover’
klóun$_{A}$	‘clown’
ótpusk$_{C}$	‘vacation’
pal^{j}indróm$_{A}$	‘palindrome’
parab^{j}éllum$_{A}$	‘parabellum’
párus$_{C}$	‘sail’
pásport$_{C}$	‘passport’
p^{j}enál$_{A}$	‘pencil box’
p^{j}ér^{j}ep^{j}el$_{C}$	‘quail’
p^{j}er^{j}iskóp$_{A}$	‘periscope’
p^{j}it^{j}ekántrop$_{A}$	‘Pithecanthropus’
pógr^{j}eb$_{C}$	‘cellar’
podsólnux$_{A}$	‘sunflower’
pójezd$_{C}$	‘train’
pól^{j}us$_{C}$	‘pole’
pon^{j}ed^{j}él^{j}n^{j}ik$_{A}$	‘Monday’
posól$_{yB}$	‘ambassador’
psalóm$_{yB}$	‘psalm’
kombájn$_{A}$	‘combine harvester’
kóm^{j}iks$_{A}$	‘comic book’
kómpl^{j}eks$_{A}$	‘complex’
kóndor$_{A}$	‘condor’
kónsul$_{A}$	‘council’
kónus$_{C}$	‘cone’
kórob$_{C}$	‘basket’
korotíʂ$_{B}$	‘shorty’
kostíl^{j}$_{B}$	‘crutch’
kóʧet$_{C}$	‘cock’
krokod^{j}íl$_{A}$	‘crocodile’
lázer$_{A}$	‘laser’
l^{j}émm^{j}ing$_{A}$	‘lemming’
mal^{j}ár$_{B}$	‘painter’
mámont$_{A}$	‘mammoth’
man^{j}ják$_{B}$	‘maniac’
mást^{j}er$_{A}$	‘master’
m^{j}ér^{j}in$_{A}$	‘stallion’
moll^{j}úsk$_{A}$	‘mollusc’
monókl^{j}$_{A}$	‘monocle’
mosól$_{yB}$	‘hip bone’
motot͡síkl$_{A}$	‘motorcycle’
miʂják$_{B}$	‘arsenic’
ob^{j}el^{j}ísk$_{A}$	‘obelisk’
ov^{j}ós$_{yB}$	‘oats’
óvoʃʃ^{j}$_{C}$	‘vegetable’
ókorok$_{C}$	‘ham’
okt^{j}ábr^{j}$_{B}$	‘october’
ol^{j}eándr$_{A}$	‘oleander’
óp^{j}ium$_{A}$	‘opium’
órd^{j}en$_{C}$	‘order’
ork^{j}éstr$_{A}$	‘orchestra’
súff^{j}iks$_{A}$	‘suffix’
tábor$_{A}$	‘band of Gypsies’
tabún$_{B}$	‘horse herd’
tarántul$_{A}$	‘tarantula’
tr^{j}úf^{j}el^{j}$_{C}$	‘truffle’
uʧít^{j}el^{j}$_{C}$	‘teacher’
fák^{j}el$_{A}$	‘torch’
f^{j}én^{j}iks$_{A}$	‘phoenix’
f^{j}it^{j}íl^{j}$_{B}$	‘wick’
edel^{j}v^{j}éjs$_{A}$	‘edelweiss’
ekv^{j}il^{j}ibr^{j}íst$_{A}$	‘equilibrist’
ekzám^{j}en$_{A}$	‘exam’
eksp^{j}ért$_{A}$	‘expert’
etáʐ$_{B}$	‘floor’
jur^{j}íst$_{A}$	‘attorney’
s^{j}ímvol$_{A}$	‘symbol’

ʂómpol$_C$	‘skewer’	psʲixólog$_A$	‘psychologist’	slovák$_A$	‘Slovak’
ʂpagát$_A$	‘split’	púdʲing$_A$	‘pudding’	soávtor$_A$	‘coauthor’
ʂúlʲer$_C$	‘cheater at cards’	rʲituál$_A$	‘ritual’	sókol$_C$	‘falcon’
ʃʃʲegól$_{yB}$	‘goldfinch’	rukáv$_B$	‘sleeve’	spósob$_A$	‘method’
		osʲótr$_{yB}$	‘sturgeon’	stolʲár$_B$	‘carpenter’