Selectional restrictions as phonotactics over sublexicons[*]

Maria Gouskova, Luiza Newlin-Łukowicz, and Sofya Kasyanenko

November 10, 2015

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. In a nonce word study, Russians also chose allomorphs based on sublexical phonotactic well-formedness, even when the phonotactic violations were non-local to the affix itself. These patterns 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.	ángʲil	angʲil-ók	'angel'	*stress-initial /stress-final base, no hiatus, no CC#*
b.	mónstr	mónstr-ʲik	'monster'	*stress-final base, no final dorsal consonants*
c.	ajfón	ajfón-ʨik	'iPhone'	*stress-final base, no final coronal obstruents, no CC#*

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 to appear).

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 sublexicons can be used in certain situations. For example, nonce words and loanwords are not tagged for diminutive formation rules, but people can consult the sublexical phonotactic grammars to decide on the probability that the item will follow a particular rule (this can be modeled as grammar inference; see Becker and Gouskova to appear, 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 controversy 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 some of the restrictions are generalized from base phonological words, not affixed words.

Our theory can be distinguished from alternative approaches to selectional restrictions, which include subcategorization frames, Generalized Alignment, and the 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 described using 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, and a forced choice nonce word study is in section 6. Our interpretation of the results is discussed in section 7, and in section 8 we discuss alternatives such as the emergence of the unmarked, subcategorization frames, and the Minimal Generalization Learner. We discuss acquisition in section 9; section 10 is the conclusion.

2 Proposal: phonotactic grammars for morphological sublexicons

2.1 Assumptions about selection and morpheme insertion

This section briefly outlines the assumptions about morphological theory that underlie 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: 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. 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). Suppletive allomorphy can be conditioned by diacritic features of morphemes such as declension class or by syntactic features such as number and aspect (there are some important typological generalizations about the conditioning of stem vs. affix suppletion, which we discuss in section 8.1). We illustrate stem and affix suppletion 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 (2a–b) have phonologically invariant realization (ignoring phonological alternations such as vowel reduction and so on), 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 [-ɨik], 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"		Realization rules for Russian:
Input:	N / \ — N DIM / \ — √HOUSE N	N / \ — N DIM / \ — √SON N	a. $\sqrt{\text{HOUSE}} \leftrightarrow /\text{dom}_{+ik}/$
Root:	N / \ — N DIM / \ — dom N	N / \ — N DIM / \ — sin N	b. $\sqrt{\text{SON}} \leftrightarrow /\text{sin}_{+ik}/$
Aff 1:	N / \ — N DIM / \ — dom ∅	N / \ — N DIM / \ — sin ∅	c. $\text{DIM} \leftrightarrow /\text{-}^jik/\ N_{[+ik]}\text{—}$ [+ik] | {dom_{+ik}, stolb_{+ik}, …}
Aff 2:	N / \ — N ᶨik / \ — dom ∅	N / \ — N ok / \ — sin ∅	d. $\text{DIM} \leftrightarrow /\text{-ok}/\ N_{[+ok]}\text{—}$ [+ok] | {sin_{+ok}, $\text{p}^j\text{irog}_{+ok}$, …}
To phonology:	/dom-ʲik/ 'house-dim'	/sin-ok/ 'son-dim'	e. $N \leftrightarrow ∅$

4

(3) English: syntactically conditioned stem allomorphy

	English "person"/"people"		Realization rules for English:
Input:	NumP [N PL [√PERSON N]]	NumP [N SG [√PERSON N]]	a. √$\overline{\text{PERSON}}$↔/pipl$_\varnothing$/ / ___PL
Root:	NumP [N PL [pipl N]]	NumP [N SG [pɝsn N]]	b. √$\overline{\text{PERSON}}$↔/pɝsn/
Aff 1:	NumP [N PL [pipl ∅]]	NumP [N SG [pɝsn ∅]]	c. PL ↔ /-z/ N$_{[+z]}$___ [+z] | | {paɪp$_{+z}$, sɑk$_{+z}$, ...}
Aff 2:	NumP [N ∅ [pipl ∅]]	NumP [N ∅ [pɝsn ∅]]	d. PL ↔∅/ N$_{+\varnothing}$___ +∅ | {ʃip$_\varnothing$, gus$_\varnothing$, pipl$_\varnothing$, ...}
To phonology:	/pipl/	/pɝsn/	e. N ↔ ∅
			f. SG ↔ ∅

2.1.2 Morpheme insertion rules refer to diacritic features

In our analysis of (2) and (3), suppletion is never really phonological. It is syntactic or lexical: the rules refer to diacritic features of the root, not to its phonological properties. We assume that in all cases of suppletion, the phonological conditioning is indirect: the lexical representations of affixes and clitics include information about which morphemes they appear with, but not 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 ↔ ə /___ [+A] | {kæt$_{+A}$, dɑg$_{+A}$, teɪbl$_{+A}$, ...}

(5) Alternative: insertion rule that refers to phonological information

INDEF↔ ə / ___ C

The more familiar 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—such cases present the clearest argument for our theory. But our analysis works both for complex and simple 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. There must be a mechanism for identifying this alternation as suppletive,[1] since it lacks a language-wide phonotactic motivation: other morphemes either alternate differently ([ði]/[ðə], Raymond et al. 2002) or do not alternate at all in the same contexts ([ɪn] 'in' is more or less invariant). In the latter 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 2015, and others).

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 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" 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$_A$, dɑg$_A$, teɪ.bl$_A$, bɹæt$_A$, ...}

　　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$_{AN}$, o.ɹənʤ$_{AN}$, i.gl$_{AN}$, ...}

　　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 allows us to explain variation and exceptions. The "a/an" alternation is supposed to be a textbook case of phonologically conditioned allomorphy, but there are exceptions to the "consonant/vowel" pattern. Even the normative statement of the rule allows for [ən] when the following syllable is unstressed and begins with [h], as in "[ən] historical". Raymond et al. (2002) elicited indefinite forms of English words, and they found that words beginning in orthographic "u" differ from those beginning in orthographic "y": "an united" is more likely than "an yellow". Moreover, words beginning in a silent orthographic "h" are significantly more likely to condition "a" than words that begin in orthographic vowels: "a <h>onorable" is more likely than "a octopus". Orthography mattered even though the participants never saw the words written down during the experiment. Further evidence against encoding people's knowledge of the "a/an" alternation via a simple consonant/vowel rule is its late

---

[1]An important question we do not resolve here is how exactly learners connect suppletive allomorphs to the same morpheme. The approach in the learnability literature on morphology suggests that learners form expectations about statistical distributions of morphemes. When these distributions are violated in the learning data, morphemes with complementary distributions are connected to each other. Something like this procedure can also explain how affixes such as comparative "-er" in English are identified as having a restricted distribution. For related ideas in phonological learning of allophones, see Peperkamp et al. (2006), Calamaro and Jarosz (2015), inter alia.

acquisition. Children learn phonologically conditioned suppletion late compared to phonologically conditioned alternations such as [ði]/[ðə], which do not require suppletive analysis (see section 9). This timeline of acquisition makes sense if sublexicons take time to assemble, and if the extraction of the phonotactic generalizations takes time as well.

2.2   Learning phonological generalizations over sublexicons

We propose that phonotactic learning happens in two stages: first, the learner figures out phonotactic generalizations over the whole language, and second, phonotactic learning is restarted for each sublexicon, identifying structures that are un- or under-attested in each morphologically defined sublexicon. The idea that grammars contain constraints stated over subsets of the lexicon is not new—it goes back at least to Morpheme Structure Constraints (Chomsky and Halle 1968; see Booij 2011 for a review). What is new about our proposal is that we reduce both morphophonological knowledge (usually thought to involve rules, see Pinker and Prince 1988, Albright and Hayes 2003) and phonotactic knowledge to the same mechanism—constraints.

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 constraints within each sublexicon will always be more restrictive than the constraints of the language in general, since anything that is allowed in a sublexicon has to be allowed in the language as a whole. Sublexicons themselves can be nested (see Ito and Mester 1995, Fukazawa et al. 1998) or be in non-overlapping sets (Kawahara et al. 2002, Jurgec 2010). But in order to learn restrictions on affixation, 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.

Likewise, 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. If these are the only properties common to the [ə] and [ən] sublexicons, then those will be the only constraints influential when deciding which indefinite allomorph to use on a new item. In our theory, this requires assigning the new item to one of the sublexicons.

This procedure might seem like overkill for a simple (and to a linguist, obvious) generalization, but it does not appear to be an obvious generalization for learners. Learners do not know a

---

[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.

priori what exactly determines the distribution of suppletive allomorphs. The conditions could be morphosyntactic or diacritic (conditioned by word class, say), or they might not align with a simple Onset/NoCoda analysis (see section 8.1 on the latter). While phonological generalizations appear to be accessible to learners early on compared to semantic ones (Gagliardi and Lidz 2014), it remains to be seen whether phonologically natural generalizations are learned more easily than arbitrary ones.

2.2.1 Assigning diacritics to novel items

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.

Our realization rules do not encode any phonological factors—rather, the phonotactic generalizations are learned separately. This procedure has the virtue of being quite general, working for simple and complex generalizations alike. It is also robust against exceptions and variation. Both exceptions and variation weaken phonotactic generalizations over the sublexicons. 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). If a sublexicon trends towards a generalization but there are some exceptions, the generalization merely becomes a statistical tendency, something a phonotactic learner can encode.

When a new morpheme 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 to appear, inter alia). Phonological generalizations about the affix will be more productive 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$_N$", 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. But we would not expect monosyllables to combine with the affix just because they do not end in a strident, 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 many 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. The learning input for [ə]∼[ən] suppletion is both variable ("[ə]∼[ən] historical"), and full of exceptions ("[ə] horrendous" but "[ən] historical"). 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, as in [ən.ʔæ.pl] (Garellek 2012). Since phonotactic sublexical generalizations are probabilistic, we expect variation, and the use of the indefinite is

indeed variable in non-standard dialects, which generalize the [ə] to vowel-initial contexts (Labov and Harris 1986, Gabrielatos et al. 2010). 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, especially when there are many exceptions (Albright 2009b, Albright and Hayes 2003).

2.2.2 Non-local phonological conditioning for suppletive allomorphy

Our theory predicts that some of the phonological conditioning for affixes can be non-local to the affix itself. This is because phonotactic generalizations are learned over the entire phonological word, not just the part of it that is adjacent to the affix. Indeed, there are cases of allomorphy conditioned by factors that are phonologically non-local. For example, Ringen and Heinamaki (1999) report an elicitation study on Finnish vowel harmony, where people's choices of suffix in disharmonic loanwords depend on the sequencing of back and front vowels in the first two syllables of the word: back-front sequences are produced with back vowel suffixes more often than front ones (e.g., *ánalyỳsi-a* ≻ *æ),* and front-back sequences are produced with front vowel suffixes more often than back ones *(hiéroglỳfi-æ* ≻ *a).* Ringen and Heinämäki analyze this as harmony with the main stressed vowel, and this interaction is not local. The suffixes themselves are not necessarily stressed, so their harmonizing with stressed vowels cannot be reduced to representationally local factors, e.g., by appealing to an autosegmental tier for stressed vowels (see Rose and Walker 2004 for related discussion). The well-studied case of Hungarian vowel harmony also involves some non-local conditioning by multiple vowels (Hayes et al. 2009), skipping over transparent [i] and [e]. Suffix vowel harmony in these languages is not suppletive, but it is lexically restricted, with exceptions characterized by several phonological trends; we assume that lexically restricted rules and suppletive allomorphy involve extracting phonotactic generalizations over sublexicons using similar mechanisms (see Gouskova and Becker 2013, Becker and Gouskova to appear).

Another class of non-local cases involves syllable-counting allomorphy (Kager 1996 and others). In Warlpiri, the ergative is realized as [-ŋku] when the stem is disyllabic, as in [ŋápa-ŋku] 'water', but as [-ɭu] on longer stems such as [kúɖujàrra-ɭu] 'child-dual' (Nash 1980). In Kaititj, the ergative is realized as [-ŋ] on vowel-initial disyllables such as [akí-ŋ] 'head', and as [-l] on consonant-initial disyllables and longer words such as [alíki-l] 'dog', [rári-l] 'wind', [ɭúNpiri-ɭ] 'fore-head' (Koch 1980:266). The leftmost syllable with an onset is stressed in Kaititj, so the distribution of [-ŋ] could be described as suffixing to the stressed syllable. But, there are exceptions in both directions, such as [wári-ŋ] 'fire' and [ʈɨí-l-at] 'this', so lexical listing of bases for each allomorph is still necessary. English also has several suffixes that are sensitive to the presence of an un-stressed initial syllable, including the comparative "-er" (*light/lighter* vs. ?*polite/politer, loose/looser vs. ?obtuse/obtuser*). All of these cases would be straightforward for a phonotactic learner that generalizes over bases: "disyllabic", "stress-final", "vowel-initial" and "starts with a primary stress" are simple phonotactic generalizations. Russian diminutive allomorphy also involves several non-local conditioning factors, as we describe next.

3 Case study: Russian diminutives

3.1 Overview

We now turn to a complex case of phonologically conditioned suppletion: Russian diminutive allo-morphy. 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. We then present an elicitation study in which we asked Russian speakers to diminutivize words that do not have commonly used diminutives, and a follow-up study with nonce words, where people chose among three suffixed options. In both studies, 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 alternatives such as subcategorization frames and alignment.

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, Olmsted 1994, Hippisley 1996, Kempe et al. 2003, Steriopolo 2008). We discuss only the masculine nouns that belong to the declension class that is consonant-final in the nominative singular. 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], [-tʃʲik] and
     [-(ə)k]/[-(i)k].

	Nom sg	Dim nom sg	Dim gen sg	Gloss
a.	ángʲil	angʲil-ók	angʲil-k-á	'angel'
b.	mónstr	mónstr-ʲik	mónstr-ʲik-ə	'monster'
c.	ajfón	ajfón-tʃʲik	ajfón-tʃʲik-ə	'iPhone'
d.	rʲibʲónək	rʲibʲónətʃʲ-ik	rʲibʲónətʃʲ-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 [mónstrʲik] '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 suffixes that end in [k], including other diminutives. Thus, the double diminutive of [angʲil] is [angʲil-ótʃ-ik], as shown in (10). There are a few monomorphemic exceptions ([karmán] 'pocket' ~[karmáʂ-ək], [mʲiʂók] 'sack' ~[mʲiʂótʃ-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 the remainder of the paper, though for completeness, we report the numbers for it in the elicitation study.[3]

(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-nʲik	'worker'	rabótnʲitʃ-ik	rabótnʲitʃ-k-ə
b.	ángʲil	'angel'	angʲil-ók	'angel-dim'	angʲilótʃ-ik	angʲilótʃ-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. The location of stress is usually characterized as unpredictable. For example, of the 12726 masculine nominatives of 3+ syllables in Zaliznjak's (1977) dictionary, 6241 (49%) have final stress, 899 (7%) have initial stress, and the rest (44%) have stress somewhere in the middle (e.g., [a.ga.rót] 'garden', [ví.mi.sʲil] 'invention', [a.ʂéj.nʲik] 'collar'). Furthermore, stress can alternate when affixes attach, as illustrated in (11) and (12). The [-ok] suffix is stress-dominant: even if the stem is lexically stressed, as in [ángʲil], the diminutive pulls stress onto itself (as in [angʲilók]) or, when its own vowel deletes, onto the following vowel (as in /ángʲil-ok-a/ →[angʲilká]). 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**ngʲil	angʲil-**ók**	g**ó**rət	gərad-**ók**
Gen sg	**á**ngʲil-ə	angʲil-**k-á**	g**ó**rəd-ə	gərat-**k-á**
Nom pl	**á**ngʲil-i	angʲil-**k-ʲí**	gərad-**á**	gərat-**k-ʲí**
Gen pl	**á**ngʲil-əf	angʲil-**k-óf**	gərad-**óf**	gərat-**k-óf**
	'angel'	'angel (dim)'	'city'	'city (dim), town'

The [-ik] and [tʃik] allomorphs pattern differently from [-ok]. For the most part, they attach to stems that have lexical stress on the last syllable, but there are a few examples like [sókəl], where

---

[3]It may be possible to analyze [-ək] as derived from the same UR as [-ok] if we adopt Gouskova and Linzen's (to appear) analysis whereby a suffix can lose stress dominance when another affix with a "regularization factor" is present in the phonological word—this would make [-ok] conditionally dominant, just like Slovenian [-əts]. Double diminutives [-otʃik] could be treated as an unanalyzed suffixes in their own right.

stress moves to the stem-final 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	kró̱t	kró̱t-ʲik	só̱kəl	sakó̱l-ʲik
Gen sg	krat-á	kró̱t-ʲik-ə	só̱kəl-ə	sakó̱l-ʲik-ə
Nom pl	krat-í	kró̱t-ʲik-ʲi	só̱kəl-í, səkal-á	sakó̱l-ʲik-ʲi
Gen pl	krat-óf	kró̱t-ʲik-əf	só̱kəl-əf, səkal-óf	sakó̱l-ʲik-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]	[-tʃʲik]	Gloss
a.	gólupʲ	gəlubʲéj	gəlubók	—	galúptʃʲik	'pigeon'
b.	kavʲór	kavrá	—	kóvrʲik	kavʲórtʃʲik	'carpet'
c.	klʲón	klʲónə	klʲinók	klʲónʲik	—	'maple'
d.	kóləkəl	kələkalá	kələkalók	—	kəlakólʲitʃʲik	'bell'
e.	sókəl	sókəli, səkalá	səkalók	sakólʲik	—	'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 /ts/, which becomes [tʃʲ] (e.g., [pálʲits] 'finger' ~[pálʲitʃʲik] '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 [tʃʲ], [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 [-tʃʲik] allomorph palatalizes /l/ but not other consonants.

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

	Unaffixed	Inst sg	Diminutive	Dim Inst sg	Gloss
a.	utʲúk	utʲugóm	utʲuʐók	utʲuʂkóm	'iron'
b.	kazák	kəzakóm	kəzatʃʲók	kəzatʃʲkóm	'cossack'
c.	lapúx	ləpuxóm	ləpuʂók	ləpuʂkóm	'burdock'

We already mentioned that [-ok] loses its vowel when a vowel-initial suffix follows, and so do some other morphemes (e.g., /kovʲor/ keeps its last vowel in [kavʲór] 'carpet', [kavʲórtʃʲik] 'carpet-dim'

---

[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 [lóbʲik], not [labók] 'pubis'. In cases such as this, either the string is derived from a homophonous root, or it has been reanalyzed as morphologically simple.

but loses it in [kav_rá] 'carpet (gen sg)' and [kóv_rʲik] 'carpet-dim'). Vowel deletion in Russian is complicated and has been the subject of much study (Lightner 1965, Halle 1973a,b, Melvold 1989, Yearley 1995, Hermans 2001, 2002, Gouskova 2012, Becker and Gouskova to appear, Gouskova and Becker 2013). The behavior of vowel deletion stems in our experiment is not discussed in this paper for lack of space.

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 all of the allomorphs resist attaching to words that have medial stress (neither initial nor final). The [-ok] allomorph tends not to attach to nouns that end in a consonant cluster. 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 stress anywhere except on the last syllable. The [-t͡ɕik] allomorph disprefers final clusters, final dorsals, coronal obstruents [t, d, s, z, s̻, z̻, ʃʃ, t͡ɕ], and nouns with non-final stress:

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

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

Recall from (14) that stem-final consonant mutation is affix-specific: the dorsal in /utʲug/ spirantizes in [utʲuz̺ók] but not [utʲug-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 [-t͡ɕik] allomorph have homophones: the borrowed Latin suffix [-ik] '-ic' (see (16)) and the agentive suffix [t͡ɕik] '-er' (see (17)). The latter suffix is an allomorph of [-ʃʃik]. These examples show that it is not the case that the diminutive [-ik] avoids 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 [-t͡ɕik] are attested, and in fact quite common, in agentive nouns.

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

	Unaffixed		Plus /-ik/ '-ic' and /-a/ 'fem nom sg'	Gloss
a.	psʲíx	'psycho'	psʲíx̱ʲikə	'psyche'
b.	pʲidagók	'pedagogue'	pʲidagóg̱ʲikə	'pedagogy'
c.	lʲingvʲist	'linguist'	lʲingvʲistʲikə	'linguistics'

(17) Diminutive suffixes vs. homophones: [-ʃʲik]/[tʃʲik] '-er'

	Unaffixed		With [-tʃʲik] '-er'	Gloss
a.	pulʲimʲót	'machine gun'	pulʲimʲóttʃʲik	'machine gun operator'
b.	danós	'report to the authorities'	danóʃʃik	'informant'
c.	kavʲór	'carpet'	kavʲórʃʃik	'carpeter'

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 noun, but not necessarily the shape of the resulting diminutive. Take velar palatalization: dorsal-final nouns such as [lapúx] become strident-final once [-ok] is attached to them, [ləpuʂ-ók]. The suffix [-ik] attaches to [ʂ]-final nouns, [ʂalaʂ] 'shack' [ʂaláʂ-ik] 'shack-dim', but not if the [ʂ] is derived from /x/: thus, [ləpuʂók] and [ʂalásʲik] are fine, but /lapux-ik/→*[lapúʂik] 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 8). 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 [aftóbusʲik] in the diminutive. She groups under "impossible" words that end in consonant clusters and have a dorsal as their last consonant, such as [lʲémʲink] 'lemming' and [malʲúsk] '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 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 "impossible" [glóbusʲik] has two hits in the corpus, and the "difficult" [aftóbusʲik] 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.

	[-tɕik]	[-ik]	[-ok]	
kazak	kazaktɕik	*kazakʲik	**kazatʃʲok**	'Cossack'
vʲetʲer	vetertɕik	veterʲik	**vʲetʲerok**	'wind'
golubʲ	**golubtɕik**	golubʲik	**golubok**	'pigeon'
pʲenal	**pʲenalʲtɕik**	**penalʲik**	pʲenalok	'pencil box'

Table 1: Examples of constructed diminutives we searched for, in pseudo-orthographic representation (i.e., stress, vowel reduction and devoicing are not shown)

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 (1977). The dictionary lists approximately 20,000 masculine nouns; taking each of them as a base, we constructed three orthographic diminutive forms. Russian orthography marks palatalization on consonants but does not reflect devoicing, vowel reduction, or the location of stress (not reflected at all unless ambiguous in context). We knew where the "dictionary" location of stress would be in the nouns we searched for since it is marked in Zaliznjak (1977), but it is not marked in the corpus we used. Table 1 shows some of our constructed diminutives, which incorporate some stem-final consonant changes (e.g., k→tɕ in diminutives with [-ok], or r→rʲ in diminutives with [-ik]). For some affix combinations, Russian has gaps, so it is not clear what the diminutive forms are supposed to look like (e.g., should the /k/ be spirantized in /kazak+ik/, [kazatɕik], or should it just be palatalized to [kʲ], as other stem-final consonants?). In such cases, we searched for palatalized, non-spirantized variants. One such example is marked with an asterisk in the table. The boldfaced examples got hits in the search for these particular bases.

We then 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. The total number of searches we performed was almost a million (20,000 nouns × 4 diminutive suffixes × 12 case/number forms). We eliminated all hits older than 1950, and before cleaning, we got 771 [-tɕik] 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 [-tɕik], 640 [-ik], and 1197 [-ok]; 17850 nouns did not occur in what we considered to be diminutive forms in the Ngrams corpus. This might seem like a low number for what is described as a very productive affix, but the threshold for inclusion in Google Ngrams is that an orthographic word has to occur in at least 40 books, so hapax legomena are not included (this actually limits the utility of Ngrams for our purposes, since a high number of hapaxes is a good measure of an affix's productivity; see, e.g., Baayen 1994).[5]

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: [-tɕik] is also an occupational suffix (e.g., [pulʲemʲot] 'machine gun', [pulʲemʲottɕik] 'machine

---

[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.

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 [-t͡ɕik], 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 back into 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, 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 [-t͡ɕik]. 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 some of the trends appears at the end of this section.

4.2.1 Descriptive lexical statistics

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 [-t͡ɕik] 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 [-t͡ɕik] sublexicon. Partly this is because final coronal obstruents—including stridents—are completely absent from the [t͡ɕik] set, in an OCP-like pattern of dissimilation. Words that end in [j] most often diminutivize with [-t͡ɕik], 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 [-t͡ɕik] sublexicon (<1%), and they are rare in the [-ok] sublexicon (4%).

   Stress and other non-local properties. There is a strong correlation between syllable count and combining with all three diminutive suffixes: the shorter the noun, the more likely it is to have a diminutive form in the corpus. Masculine noun bases as a whole usually have final stress (55.7%, which includes monosyllables), followed by medial (27.6%) and initial stress (16.6%). Looking at just words of three or more syllables, the proportions are 49% final, 44% medial, and 7% initial (excluding words with secondary stress on the first syllable). Stress is overwhelmingly final on nouns that combine with [-t͡ɕik] 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 not particularly rare—it occurs in 8% of the masculines. It is, however, rare
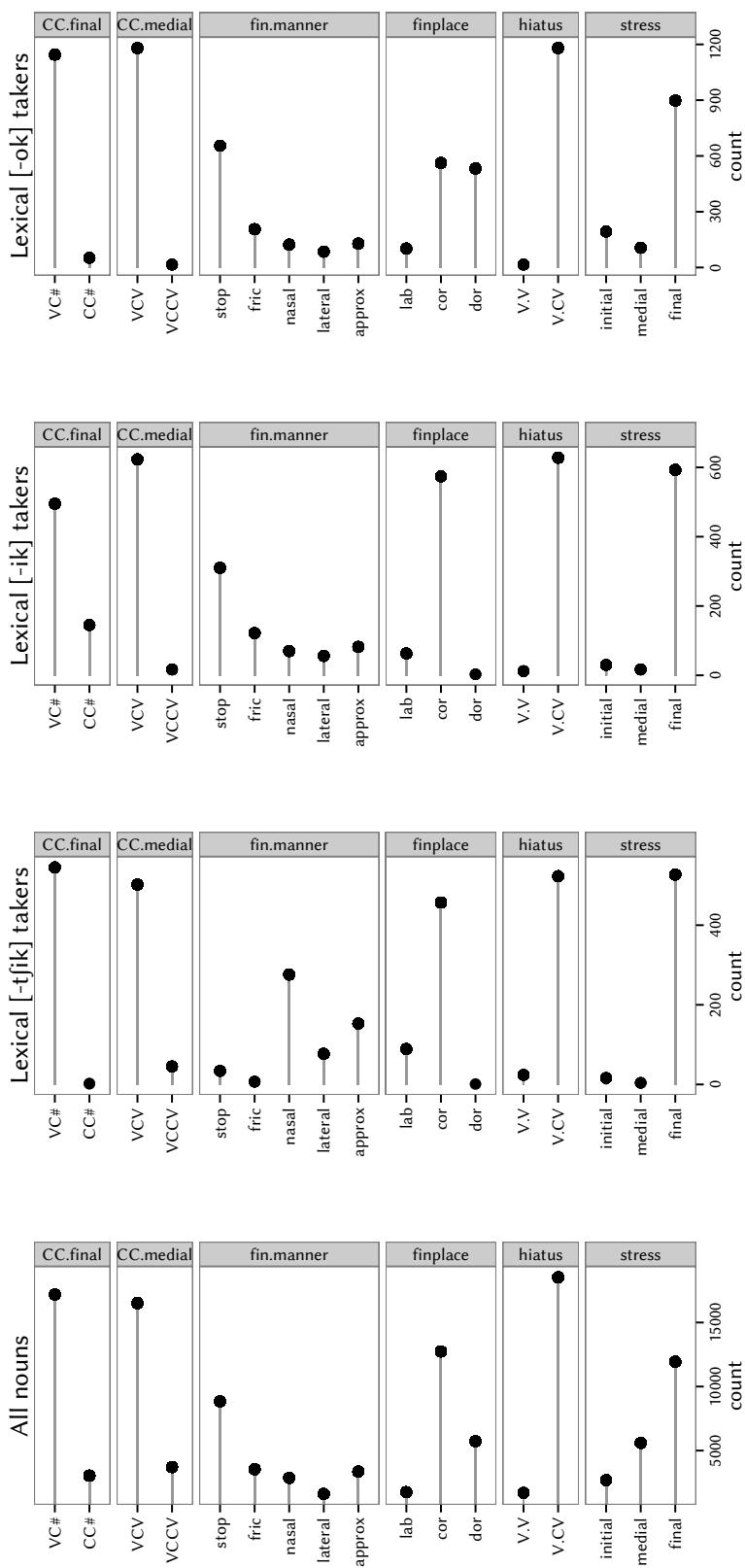
Figure 1: Diminutive sublexicons: properties of bases

17

in the [-ʧʲik] sublexicon (4%) and in the [-ik] and [-ok] sublexicons (2% and 1% respectively). These differences are too small to reach significance in the regression we ran, but when we inspected the items in each sublexicon manually, we saw only a few words we recognized as real diminutives (e.g., [paúk]/[pauʧʲók] 'spider', [idiót]/[idiótʲik] 'idiot', [trotuár]/[trotuárʧʲik] 'sidewalk'). 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).

4.2.2   Statistical model for the lexicon

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 900 nouns from the "none" category for this regression. (We also fit a polytomous regression to the entire dataset, without sampling, and the results were qualitatively similar to the regression we report.) Table 2 shows the log odds for each predictor, by suffix. 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. As for manner, 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 also has an effect: high vowels decrease the likelihood of [-ik], and round vowels increase the likelihood of [-ok], in a kind of vowel harmony pattern (cf. Becker 2009 for Hebrew plurals in [-ot]). 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. When stress is final, affixation with [-ʧʲik] and [-ik] is more likely than the baseline, and when stress is medial, affixation with [-ok] is less likely. Monosyllables are more likely to diminutivize with [-ik] and [-ok] and less likely to diminutivize with [-ʧʲik]; polysyllabic words are in general more likely to have no diminutives at all. 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.

Collinearity is always a concern with modeling any lexicon, since there are correlations between

	"none"	[-tʃik]	[-ik]	[-ok]
Intercept	−0.6815	−4.736	−1.366	(−0.1948)
lastC=dor	−0.4743	−3.379	−4.653	2.176
lastC=glide	(0.4075)	(0.3255)	−1.816	(0.03966)
lastC=nas	−0.6685	2.955	−1.32	−0.3231
lastC=approx	−0.2752	2.341	−0.7298	(−0.1582)
final.CC	1.078	−3.593	0.5944	−1.345
medial.CC	1.278	(0.2861)	−1.068	−2.095
hiatus	1.021	(−0.3324)	−0.7363	−1.017
final.stress	(−0.1879)	2.068	0.733	−1.05
medial.stress	1.362	(−0.9343)	(0.2376)	−2.143
monosyllabic	−1.266	−1.139	0.7426	0.9465
lastV=high	0.4137	(−0.1201)	−0.4734	(−0.146)
lastV=round	−0.2382	(0.2181)	(−0.1535)	0.2845

R^2 likelihood: 0.28, AIC=6187, BIC=6502, Collinearity measures: $6.66 \leq \kappa \leq 12.695$, $1.01 \leq VIF \leq 2.954688$

Table 2: Model for the lexical bases' likelihoods of appearing with various diminutive allomorphs

phonological properties (though see Wurm and Fisicaro 2014 for a different perspective on this). For example, Ryan (2014) finds that Russian words are more likely to have initial stress the more consonants they have in their first-syllable onset; correlations between stress location and word-final codas would also be unsuprising. Thus, we report the collinearity measures for the models summarized in the polytomous regression. Since the coefficient table summarizes four separate regressions, we show the range of kappa (condition number) values and the variance inflation factors for the models. The values indicate a relatively low degree of correlation between the predictors (Belsley et al. 2004).

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 *#[+nasal, Dorsal]. In order to help the learner identify generalizations about stress phonology, we supplied it with a stress tier.

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). Since our affixed diminutives come from an orthographic corpus, we had to add 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

[-t͡ʃʲik] 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., [kovʲór] 'carpet' vs. [kóvrʲik] '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. An example of the constraint set generated for one of the grammars, the base grammar for [-ik], is given in Table 3. Each constraint is accompanied with a brief description. Some constraints worth noting are the constraint against final lapses and the constraint against initial stress followed by a lapse, which is part of the requirement that [-ik] bases have final stress, and the constraints against final dorsals—these correspond to Polivanova's (1967) observations about [-ik]. 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 predicted 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. If a word receives a high probability from more than one subgrammar, the choice of diminutive affix should vary. The ratings people assign to their own diminutives should reflect the phonotactic well-formedness of the diminutive in its sublexical grammar.

Our theory makes a prediction that distinguishes it from approaches to phonologically predictable morphology that 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/middles 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.

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.

Constraint	Description	Weight
*[-stress][-stress]]$_{Wd}$, Stress tier	No final stress lapses	5.192
*[+lab,-back]]$_{Wd}$	No final palatalized labials	4.822
*[-syll][+voice,-syll][-approx]	Specific CCC constraint	4.623
*[-son,+voice][-voice]	Voice agreement	4.567
*[+syll]]$_{Wd}$	No final vowels	4.387
*[$_{Wd}$ [+son,-syll][-lateral]	Specific initial CC constraint	4.24
*[+cons,-back][-syll]	*CⁱC	4.161
*[-back,-syll][-high,-low,+back][-son]	Specific sequence constraint	4.13
*[+delrel][+son]]$_{Wd}$	Specific final cluster constraint	4.089
*[+stress][-son][+high]	Specific stress sequence constraint	4.05
*[+stress][+stress], Stress tier	No stress clash	4.049
*[-voice][+cont,-ant]	Specific CC constraint	3.967
*[+son,+back][+voice,-low,-back][+son]	Specific sequence constraint	3.954
*[+delrel,-voice][-son,+voice]	Specific voicing sequence constraint	3.923
*[+back,+syll][+back,+syll]	Specific VV constraint	3.898
*[+syll][-low,+syll][+cont]	Specific VVC constraint	3.888
*[+stress][+syll]	*V́ V	3.809
*[$_{Wd}$ [+stress] V V, Stress tier	No initial stress in longer words	3.426
*[+lab][+high,-syll]	Specific CC constraint	2.985
*[-low]]$_{Wd}$	No final dorsals/low vowels	2.918
*[-nasal,+lab][-approx,+voice]	Specific CC constraint	2.904
*[-lateral,+voice,-back][-syll]	Specific sequence constraint	2.696
*[-approx,-retro][+cons,+lab]	Specific sequence constraint	2.666
*[-back,-syll][-nasal,+lab]	Specific sequence constraint	2.455
*[+voice,-back,-syll][+voice,-strid]	Specific sequence constraint	2.348
*[+lab][-cont,-back]	Specific cluster constraint	2.22
*[+dor,-back]]$_{Wd}$	No final palatalized dorsals	2.133
*[+voice,-back,-syll][+cont,-back,-syll]	Specific cluster constraint	2.088

Table 3: An example sublexical phonotactic grammar for [-ik] bases, generated by the UCLA Phonotactic Learner

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]~[kəblutʃ-ók] 'heel', [kóləkəl]~[kəlakólʲ-tʃik] 'bell', [mʲátʃ]~[mʲátʃ-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]~[kəzatʃ-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.). Cells that are either impossible (e.g., Russian has no dorsal nasals) or have very few members in the lexicon are marked with "NA."

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

	Stop	Fricative	Nasal	Lateral	Rhotic
cor	s̪pagát (19)	kamís̪ (25)	bʲidón (14)	tʃixól (31)	abazúr (30)
dor	kazák (21)	kóʐux (4)	NA	NA	NA
lab	pʲirʲiskóp (12)	rukáf (1)	astranóm (10)	NA	NA

	No cluster	Cluster
cor	bʲiérkut (83)	ts̪imʲiént (37)
dor	bátʲik (18)	tʃʲitvʲérk (7)
lab	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	utʃítʲilʲ/utʃitʲilʲ-á (1)

5.2.4 An aside on using real words

We used 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.
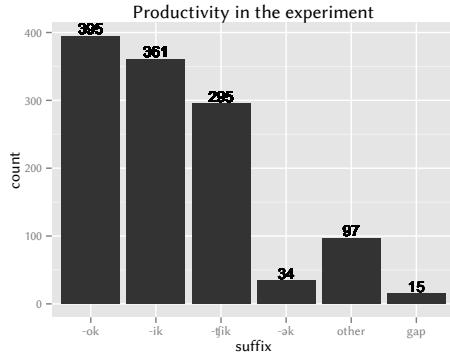
Figure 2: Overall productivity of each suffix in the experiment

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-tʃⁱik] 'iPhone', [manⁱják]→[mənⁱjitʃʲ-ók] 'maniac, serial killer' and [kənⁱibál]→[kənⁱibálⁱ-tʃʲik] 'cannibal', but had trouble with [klóun] 'clown' and [ótpusk] 'vacation'. Among established lexical diminutives, we find [pənⁱidⁱélⁱnⁱik]→[pənⁱidⁱélⁱnⁱitʃʲ-ək] "Monday", but there is no diminutive for [tʃʲitvⁱérk] "Thursday", which is phonologically not compatible with any of the allomorphs (ends in a cluster so no [-ok] or [-tʃʲik], and it is dorsal-final so no [-ik]). Thus, we believe that semantic/stylistic factors play at best a minor role in this experiment. The second experiment (section 6) uses nonce words, which eliminate semantic factors altogether.

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 [-tʃʲik], and 34 with [-ək]. They also produced 97 forms with other expressive suffixes (such as [-uʂkə] or [-iʂkə], or [-óniʂ] 'offspring of'). 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 top left plot in Fig. 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 allomorph. These plots allow one to compare the relative proportion of, say, cluster-final nouns that were suffixed with [-ok] vs. [-ik] or [tʃʲik], 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 [-ok] than with the other two suffixes. The position of stress in the base (initial, medial, final) also varies by allomorph: initial bases combined with [-ok] far more often than with [-ik] or [-tʃʲik]. Medial stress bases are less likely to combine with [-ok] than with [-ik] or [-tʃʲik], however. With respect to manner, sonorants preferentially combine with [-tʃʲik], 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 [-tʃʲik], and almost never with [-ik].

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

	-tʃʲik	-ik	-ok
Cluster	kambájn/kambájn-tʃʲik	mətatsíkl/mətatsíklʲ-ik	pójist/pəjizd-ók
Dorsal	gʲértsək/gʲértsək-tʃʲik	patsólnux/patsólnuxʲ-ik	manʲják/mənʲjátʃʲ-ók
Coronal	astiróit/astiróidtʃʲik	ʂpagát/ʂpagátʲ-ik	xóbət/xəbat-ók
Labial	pʲirʲiskóp/pʲirʲiskóptʃʲik	jipʲískəp/jipʲískəp-ʲik	tʃʲérʲip/tʃʲirip-ók
Hiatus	klóun/klóun-tʃʲik	antʃʲóus/antʃʲóusʲ-ik	ópʲium/apʲium-ók

(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-tʃʲik]
final	[tʃʲirtʲóz] vs. [tʃʲirtʲiz-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 holistic 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. 4. This figure was generated as follows. We took all the bases and all the (non-gap) diminutives that people produced in the experiment, and we tested the phonotactic grammars for each sublexicon obtained from the Ngrams corpus on these bases and diminutives (recall sections 4.1 and 4.3). The resulting violations are summed and plotted in the barplots: for example, in the [-tʃʲik] subplot on the left, the shortest column is the one that shows the violations of the bases and diminutives that people produced with the [tʃʲik] suffix. The base+diminutive sublexical grammars for [tʃʲik] assigned the worst violations to those bases and those diminutives that people produced with [-ok]. The taller
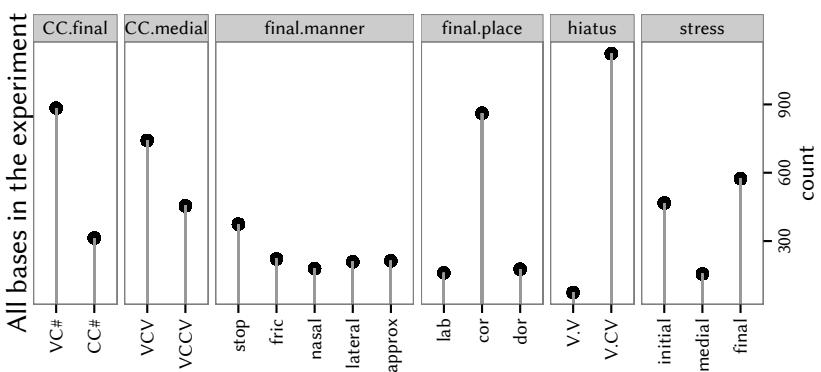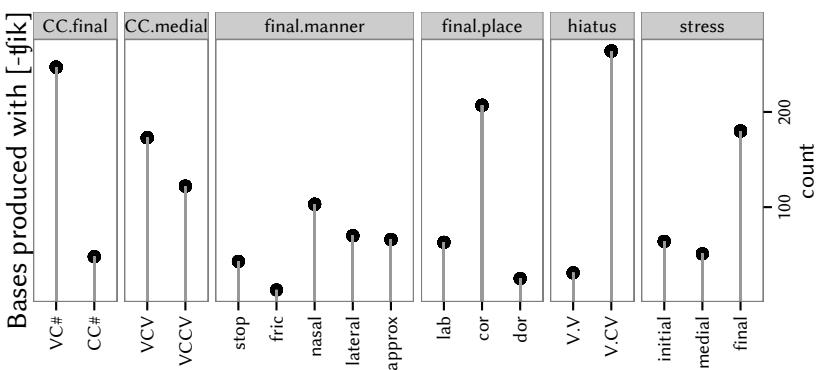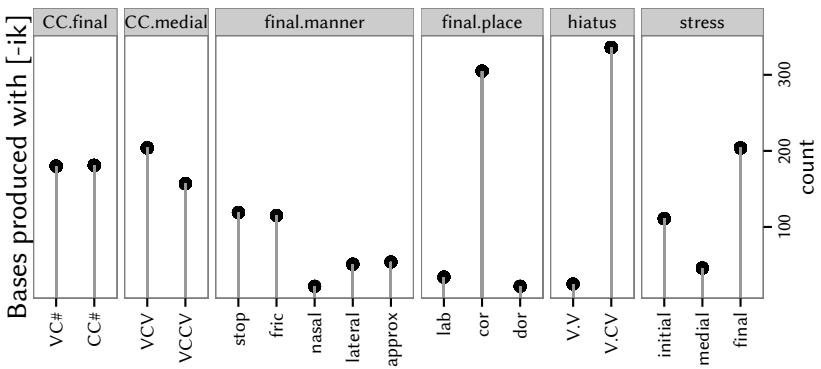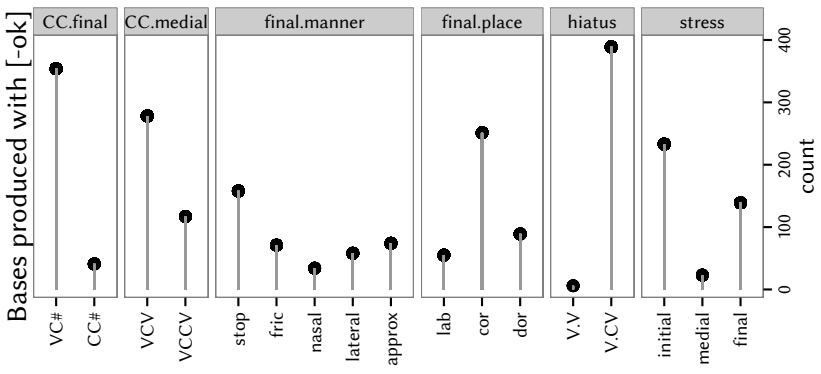
Figure 3: Number of bases people produced with each suffix, organized by phonological property of the base
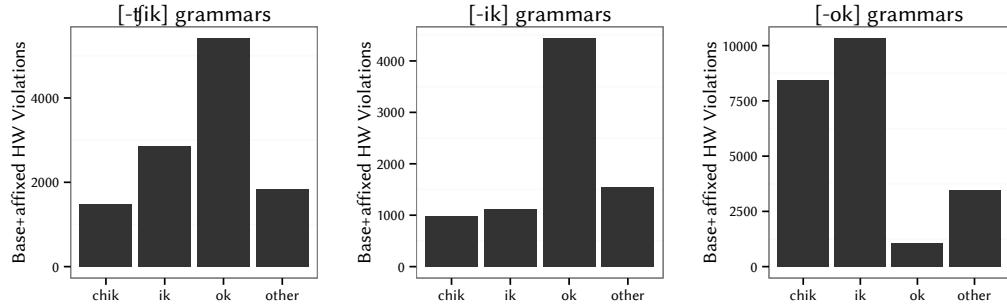
Figure 4: Summed violation scores assigned by grammars trained on base and diminutive sublex-icons, grouped by people's production patterns.

the column, the more violations the experimental bases and diminutives got from the respective sublexical grammar.[6] 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 allomorph: [-ik] vs. [-ok] vs. [-tʃik] vs. "none/other". We used the poisson reformulation heuristic, which accommodates random slopes and intercepts for predictors. 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" (0 violations) and "fail" (violations>0). The model structure is summarized as an R formula under the model coefficient table.

The log-odds from the model are summarized in Table 4. The first two rows show the overall likelihood of each suffix allomorph compared to the baseline poisson distribution.[7] Thus, [-tʃik] and [-ok] are less productive than might be expected by chance, and [-ik] is more productive. (This reflects our choice of stimuli, not the productivity of these allomorphs in the language, however.) The rest of the table shows how the log-odds of choosing a particular allomorph correlate with the base passing the sublexical grammar trained on bases, and how they correlate with the diminutive passing the grammar trained on diminutives. Predictors that are significant at $p < 0.05$ are given without parentheses; positive log-odds indicate increased likelihood, and negative log-odds indi-

---

[6]For example, take the base [slovák] 'Slovak' and the diminutives people produced for this form, which included [slovatʃók], [slováktʃik], and [slovátʃiʂka]. Each of the base subgrammars for [ik], [-ok] and tʃik was then tested on the experimental base [slovák], and each diminutive subgrammar was tested on all the diminutives people produced. The leftmost column in the [tʃik] subgraph of Fig. 4 includes the summed scores that the [-tʃik] grammar assigned to [slovák] and [slováktʃik], and the rightmost column in the same subgraph shows the scores this grammar gave to [slovák] and [slovatʃók].

[7]This polytomous regression summary is really the summary of a large hierarchical (mixed effects) poisson regression that includes the individual predictors and their interactions with suffix type. Thus, when "ok-dim" is shown as having log-odds of 2.589, it means that the likelihood of a base that passed the "ok-dim" grammar with a score of 0 and had been chosen with the "ok" suffix is 2.589 times higher than the baseline poisson distribution. The significance is assessed based on the z-value. We are grateful to Antti Arppe for discussing the polytomous package with us.

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

polytomous(allomorph $\sim$ ok-dim + ok-base + ik-dim + ik-base + ʧik-dim + ʧik-base + (1 + ok-dim + ok-base + ik-dim + ik-base + ʧik-dim + ʧik-base|participant) + (1|base), heuristic= "poisson.reformulation")

R^2likelihood: 0.4654; AIC: 1719; BIC: 1861; VIF=15.493, κ=9.245

Correlation of model's prediction with people's choices	
Kendall's τ	0.817
Spearman's ρ	0.77
Pearson's product-moment	0.69

Table 4: Model for allomorph choice in the elicitation study, with sublexical phonotactic grammar predictors

cate decreased likelihood. We also report the correlations between the statistical model's predicted allomorph choices and people's productions.

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 [-ʧik]. Getting a good score from either grammar trained on [-ʧik] increases the chances of being suffixed with [-ʧik]. Good [-ʧik] diminutives are less likely to end up in "other", and good [-ʧik] bases are less likely to end up suffixed with [-ok].

Both the base and the diminutive grammars have an effect in the model, but diminutive grammars carry the lion share of the work. This can be seen from the magnitude of the estimates—diminutive grammars consistently have larger log-odds estimates than corresponding bases. But the base grammars are still doing something in the model—the R^2 likelihood of the full model is larger than that of the model including only the diminutive grammars (R^2=0.456, AIC=1722, BIC=1803; higher R^2 and lower AIC/BIC numbers indicate better fit).

	"none"	[-t͡ɕik]	[-ik]	[-ok]
Intercept	−0.9156	(0.05384)	1.539	2.668
final.CC	(−0.3778)	−1.163	(0.2913)	−2.148
lastC=dor	(−0.711)	(−0.2564)	−2.361	(−0.3146)
lastC=son	(0.4413)	1.606	−0.7354	(−0.2002)
final.stress	(0.2288)	0.9206	0.6684	−0.6351
medial.stress	(0.1663)	(0.001719)	(0.1441)	−1.953
hiatus	−1.64	−1.571	−1.805	−3.022

R^2likelihood=0.18, AIC=2446, BIC=2587; VIF=20.08, κ=11.10

polytomous(final.CC + lastC=dor + lastC=son + final.stress + medial.stress + hiatus + (1+ final.CC + lastC=dor + lastC=son+ final.stress+ medial.stress+ hiatus| participant), heuristic= "poisson.reformulation")

Correlations between model prediction and people's productions	
Kendall's τ	0.38
Spearman's ρ	0.41
Pearson's product-moment	0.39

Table 5: Polytomous regression table for the elicitation experiment, handpicked predictors

5.4.2 The handpicked predictor model

This section models the elicitation study results with handpicked phonological predictors that encode 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 4.

The dependent variable in the model was the choice of suffix ([-t͡ɕik], [-ik], [-ok] and "other"). The predictors were analogous to the model for the lexicon reported in section 4.2, except that it also includes a random intercept for "word" (base noun) and a random intercept for participant, as well as random slope for all the predictors.

The model log odds are summarized in Table 5. 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 less likely to be diminutivized with [-ok] and [t͡ɕik]. Final dorsals decrease the likelihood of [-ik]. Final sonorants increase the likelihood of [-t͡ɕik], but decrease the likelihood of [-ik]. Final stress increases the likelihood that [-ik] or [t͡ɕik] would be chosen (and perhaps as a result also decreases the chances of [-ok], albeit unexpectedly). The suffix [-ok] is significantly less likely to appear on words with medial stress, and it is much less likely to appear on words with hiatus compared to the other suffixes (although hiatus is in general unlikely). This regression model has some problems, due to the high correlations between lexical properties of the bases we could choose from, but it still manages to achieve a modest match with people's choices. Still, compared to the success of the full model in the previous section (correlations around 70-80%), the handpicked model is inferior.

		[tʃʲik]	[-ik]	[-ok]
a.	Final syllable	no final CC ✓	not restricted✓	no final CC✓
b.	Final C place	no final dorsals	no final dorsals✓	not restricted
c.	Stress location	must be final✓	must be final✓	not medial✓
d.	Final C manner	prefers sonorants✓	not restricted	not restricted

Table 6: Support for Polivanova's observations. A check mark indicates that the effect was supported in the elicitation study

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 a problematic move: 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 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.

Finally, our study confirmed many of Polivanova's (1967) observations, as summarized in the table below.[8] The effects for which we found quantitative support are marked with a check mark.

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 leftmost 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 leftmost bottom plot shows the ratings for how the same forms did in the grammar trained on [tʃʲik] diminutives.

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

---

[8]Polivanova devotes a lot of attention to final [ts], which is supposed to disfavor [-ik]. We did not investigate this.
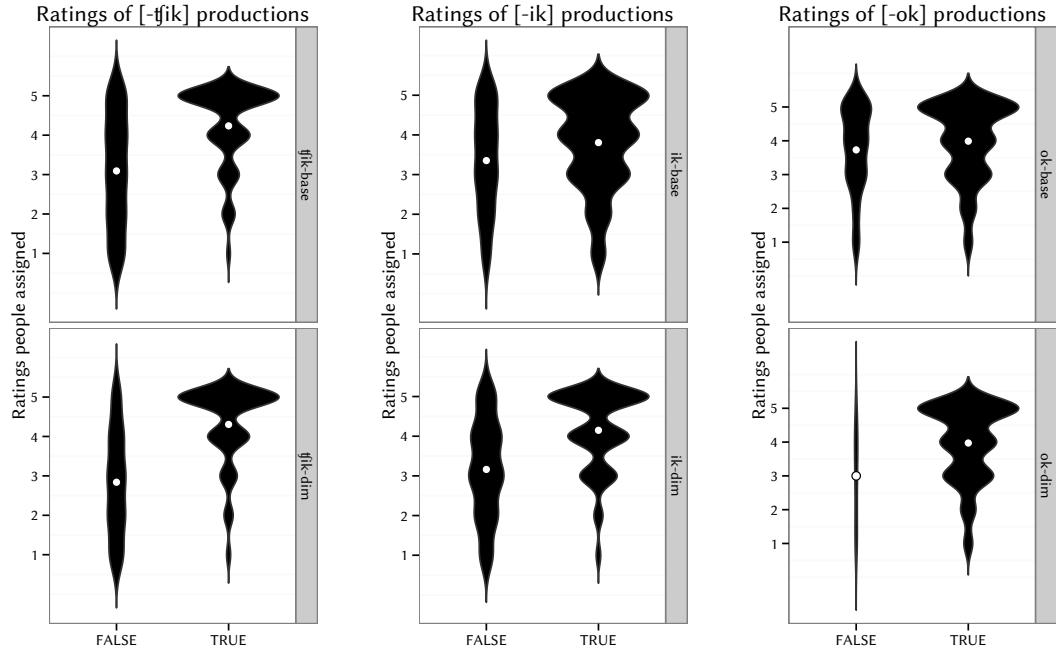
Figure 5: Ratings people gave for the words they produced with [-tʃik], [-ik], and [-ok], grouped by whether the bases and diminutives passed the respective affixes' sublexical phonotactic grammars (true) or failed them (false). The white dot shows the mean.

6 A Nonce Word Study

We next conducted a nonce word study (wug study, Berko 1958) that was designed to test the effects of nonlocal properties of bases on allomorph choice. The elicitation study identified hiatus and medial stress as possible nonlocal properties of interest. If this feature of diminutive formation is productive, we have support for our model that differentiates it from certain alternatives. For example, the Minimal Generalization Learner (Albright and Hayes 2003) cannot find or characterize such generalizations about middles of words, as we show in 8.3. We further expect that both the base and diminutive well-formedness will affect allomorph choice, which distinguishes our theory from approaches that allot all of the selection to the input or the output, such as input-oriented subcategorization frames (Paster 2006, Embick 2010), output filter/Control theory (Bye 2007), and product-oriented schemata (Bybee 2001, Kapatsinski 2011). Some of these models are discussed in more detail in section 8.

In this experiment, we use a large number of items that are phonotactically diverse, and we expect that the ones that do better in allomorph-defined sublexicons will be more likely to be chosen with those sublexicons. We also manipulate the stimuli in a targeted way to compare characteristics such as the presence or absence of hiatus, the location of stress, and the presence of medial consonant clusters.

	Hiatus	No hiatus
Medial stress	pudúal (30)	pudúlal (30)
Final stress	ɕilʲiár (30)	ɕilʲinár (30)

Table 7: Stimulus design for the experiment: hiatus

6.1 Methods

6.1.1 Materials

The experiment tested a total of 300 nonce masculine nouns. To study the effects of vowel hiatus, we used 120 items. Half of these nouns included vowel-vowel sequences, and the other half had the same vowels separated by a consonant, as shown in table 7. We also manipulated stress: in half of the nouns, the first of the two vowels was stressed, and in the other half—the second.

We also tested the effects of medial and final clusters, with 180 items. Of these, 90 had a CC sequence in the middle (e.g., *sʲubmár*), 30 ended in CC (e.g., *xʲirásn*), and 60 had no consonant clusters (e.g., *sunʲés*). The no-cluster group included 30 dorsal-final stems (e.g., *raʐóx*) and 30 stems that ended in palatilized consonants (nʲ,mʲ, rʲ, sʲ/tɕʲ/ʂ/ʃʲ). The main manipulation in this stimulus set was to bias the choice toward certain allomorphs: thus, dorsal-final bases are supposed to combine only with [-ok], and strident-final bases should combine with either [-ok] or [-ik] but not with [-tɕʲik]; CC-final bases should combine with [-ik] only. The nonce words with medial clusters had final consonants that were compatible with any of the allomorphs, but the medial clusters themselves were underattested in the sublexicon for at least one of the allomorphs but were allowed by the sublexical grammar of one of the other allomorphs. For example, the nonce word [bʲatsxál] gets a violation score of 4.725 from the [-tɕʲik] base grammar and a score of 10.69 from the [-ok] base grammar, but its violation score in the [-ik] grammar is 0. Thus, we would expect this word to be chosen with [-ik] most often (and indeed it was). For the nonce word [bʲulʲbʲólʲ], the scores were [tɕʲik]: 0, [-ok]: 24.012, [-ik]: 6.938. We expected people to choose the suffix [tɕʲik] for this word most often. To factor out the effects of stress position, we recorded all 180 of these nonce words with final stress, since all three allomorphs combine with final stress nouns. Finally, we verified that our stimuli were passable nouns of Russian in two ways: by inspecting them using our native intuitions, and by training the UCLA Phonotactic Learner on the full list of Russian masculines and having the resulting grammar rate the stimuli. All of the stimuli received a perfect score (0 violations).

We checked the stimuli for lexical neighbors, using Levenshtein edit distance 1 as our criterion (1 substitution, 1 deletion, or 1 insertion; see Luce 1986). Two of the 300 items had neighbors in the [tɕʲik] sublexicon: *ʐulʲipan*∼*tʲulʲipan* 'tulip', *kuvain*∼*kuvʂin* 'pitcher'). Four of the items had neighbors in the [ok] sublexicon: *xurʲak*∼*xrʲak* 'boar', *zafut*∼*zakut* 'small barn (dialectal)', *ʂazʲok*∼*ʂaʐok* 'step (dim)', *bʲizʲok*∼*bitok* 'cutlet'. There were no neighbors in the [-ik] sublexicon. This small number of neighbors meant that neighborhood density did not ultimately correlate with affix choice in the analysis.

A male native speaker of Russian recorded each word in its base form and with each of the three diminutive suffix allomorphs, [-ok], [-ik], and [-tɕʲik], so the total number of audio files was 1200. Consistent with the morpho-phonology of the suffix, the [-ok] diminutives were recorded with stress shifted to the suffix (e.g., [pudúal] ∼[pudualók]). On the other hand, [-ik] and [-tɕʲik] diminutives were

31

Начальная форма слова: Base form of the word:

чунуа́п ►   **tʃʲunuáp** ►

Варианты уменьшительно-ласкательной формы: Variants of diminutives:

1. **чунуапо́к** ►   1. **tʃʲunuapók** ►
2. **чунуа́пчик** ►   2. **tʃʲunuáptʃʲik** ►
3. **чунуа́пик** ►   3. **tʃʲunuápʲik** ►

Какой вариант Вам больше нравится?   Which variant do you prefer?

[1] [2] [3]   [1] [2] [3]

Трудно было сделать выбор?   Was it hard to choose?

[да][нет]   [yes] [no]

Figure 6: Example of a trial in the experiment

recorded with stress on the same syllable as the base (e.g., [pudúalʲik], [pudúalʲitʃʲik]). The audio files and a list of the stimuli are available at https://www.nyu.edu/projects/russian/dim_nonce/.

6.1.2 Design

The experiment was conducted on the web using Experigen (Becker and Levine 2010). The participants used their own browsers and speakers/headphones. The stimuli were presented in audio as well as in Russian orthography with stress marked. Each audio file had to be played before advancing to the next stage of the experiment. The base was always presented first, but diminutive forms were presented in a randomized order, so that none of the three suffixes consistently appeared first, second, or last. We asked people to select one form of three as their preferred diminutive and to indicate whether the choice was difficult (yes/no). An example of a trial is shown below, along with an English translation.

To keep the experiment tolerably short, we showed each person a structured subset of the 300 nonce words, which included 35 items per participant. These 35 items included seven subgroups: 5 trisyllables without hiatus, 5 trisyllables with hiatus, 5 disyllables that ended in VC, 5 disyllables that ended in CC, and 5 of each of medial cluster words that were expected to combine with [-tʃʲik], [-ik], and [-ok], respectively.

6.1.3 Participants

We recruited anonymous participants through various Russian social network sites and word of mouth. Participants volunteered their time and information about their sex, age, and linguistic background. A total of 219 participants completed 35 trials; we excluded 4 outlier participants who never selected the [-ok] allomorph, so our data include the responses of 215 participants.

6.2 Results

People overall chose the [-tʃʲik] suffix more often than [-ik], which was in turn chosen more often than [-ok]. This pattern does not necessarily reflect the productivity of these allomorphs in Russian, since we chose our stimuli to test specific hypotheses rather than based on overall phonotactic type

**Productivity of each suffix
in the experiment**
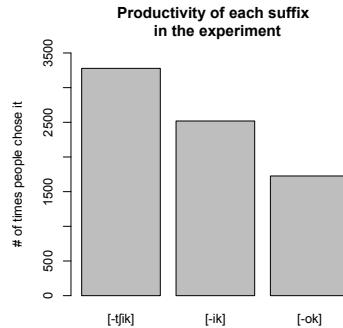
Figure 7: Overall productivity of suffix allomorphs in the experiment

frequencies in the language. But it is important that the distribution is skewed in the direction of [-t͡ʃik] and away from [-ok] because our statistical analysis of people's affix choices builds on the baseline distribution of these affixes in the experiment.

As in the elicitation study, we analyze the data in two ways: first, we use the sublexical phonotactic grammar violations produced by the UCLA Phonotactic Learner, trained on the sublexicons we obtained from Google Ngrams and tested on the base and diminutive stimuli we designed for this experiment. Second, we use handpicked phonological variables. As we will show, the first method is more successful.

In figure 8 are the Cleveland plots for the nonce words, first shown with all the bases used in the experiment, and then broken down by affix.

6.2.1 Analysis with sublexical phonotactic grammar predictors

The relationship between the sublexical phonotactic well-formedness of the stimuli and people's allomorph choices is shown in Figure 9. This is a bar plot of violation scores assigned by the sublexical grammars trained on the bases and derived diminutives that people saw in the experiment. As can be seen from the graph, the sublexical base and diminutive grammars trained on [-ok] lexical bases and diminutives assign the lowest violation scores to those bases that people chose with [-ok]; the same is true for [-ik] and [t͡ʃik]. The pattern is most clear in the [-ok] plot but diminishes in the [t͡ʃik] plot, because [-t͡ʃik] acts as a default in the study—people seem to opt for it when no other choices seem good. Thus, the overall pattern of the [-t͡ʃik] grammar preferring the [-t͡ʃik] bases and diminutives still holds, but it is less pronounced.

To analyze the effects of sublexical phonotactic violation scores on people's choices of allomorphs, we used a polytomous poisson logistic regression model (as in the previous experiment described in 5.4.1). The model reported in Table 8 includes six predictors: the violation scores of the bases and the diminutives for each of the three allomorphs' sublexical phonotactic grammar. Each cell summarizes the interaction of the predictor with suffix type. The model includes a random by-participant intercept and slopes for all independent variables, and a random by-base intercept.

33
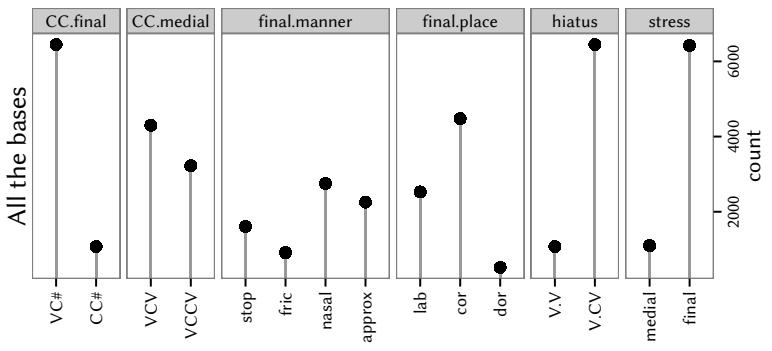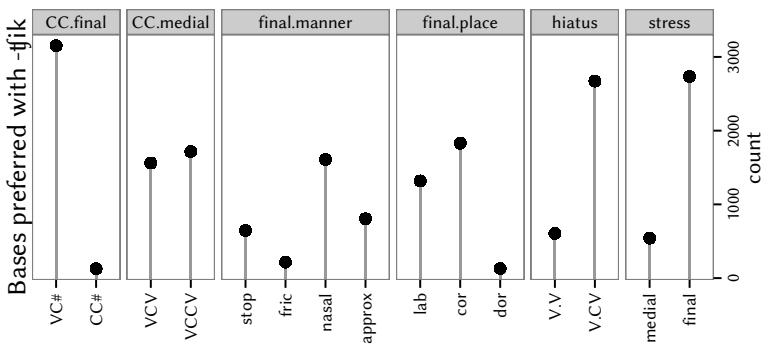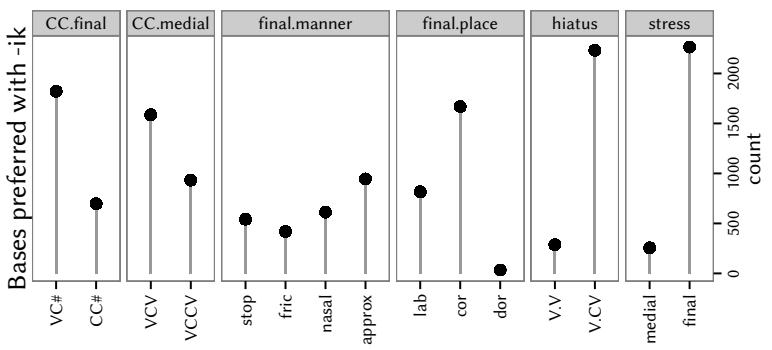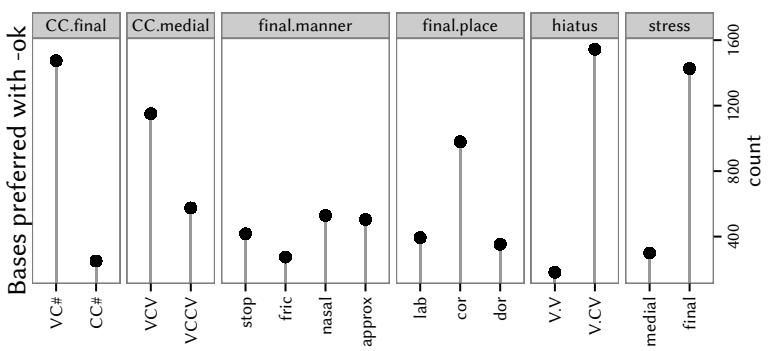
Figure 8: Phonological features of bases chosen with each affix in the nonce word study
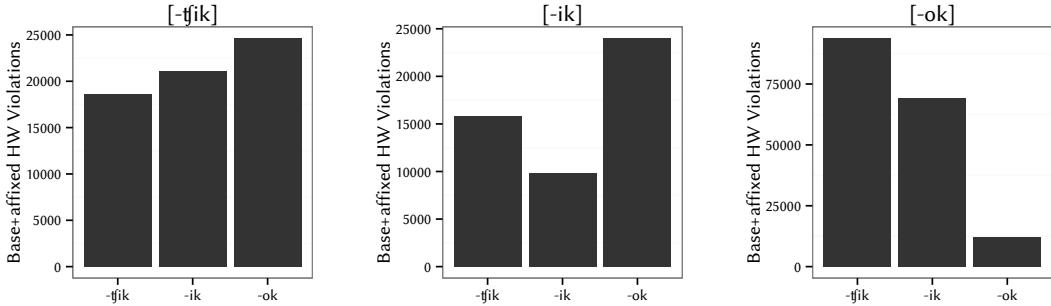
Figure 9: Summed violation scores assigned by sublexical phonotactic grammars trained on [-tʃik] bases/diminutives (left), [-ik] bases/diminutives (center), and [-ok] bases/diminutives (right)—plotted against people's choices (x-axis groupings within each plot). The sublexical phonotactic grammars trained on [-ok] assign the lowest scores to the bases that people chose with [-ok]; ditto for [-ik] and [-tʃik].

The absolute violation scores produced by the UCLA Phonotactic Learner are not comparable between sublexical grammars, so we transformed the scores into pass (score=0)/fail (score>0) to allow the model to converge.

The baseline likelihood of the allomorphs is in the first two lines of Table 8. The [-tʃik] allomorph is significantly more likely than the other two, which do not differ from chance (i.e., the baseline poisson distribution). As for the sublexical predictors, they have the expected effect: for example, if a base passes the [-ik]-base sublexical phonotactic grammar, it is more likely (log odds 0.11) to be chosen with the [-ik] suffix, whereas its likelihood of ending up in the [-ok] or [-tʃik] categories is not different from the baseline poisson distribution. The same is true to an even larger extent for the [-ik]-diminutive sublexical phonotactic grammar—if a diminutive passes that grammar, its likelihood of being chosen is increased by 0.77, whereas its chances of ending up with [-ok] are unaffected, and its chances of being chosen with [-tʃik] are significantly decreased (-0.23). All six phonotactic grammars have an expected positive effect on the choice of their allomorphs, and they affect the choice of other allomorphs either negatively or not at all. The only small exception is the [-ok]-base grammar, passing which slightly increases a base's chances of being chosen with the [-tʃik] allomorph. This is unsurprising, since this allomorph serves as a kind of default in the experiment, and the effect itself is rather small.

We report three measures of model fit. First, the R^2 likelihood, which is a measure of model fit for a logistic regression model; second, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), measures of entropy unaccounted for by the model, and third, the correlation between people's choices of allomorphs and the statistical model's predictions of these choices. For the latter method, we used the *predict* function in the *polytomous* package.

We next examine the role of source-oriented and product-oriented phonotactic generalizations. We cannot compare the full model with a model that contains just base or just derived sublexical predictors using ANOVA model comparison, since the models would not be properly nested. We can, however, fit models with only base or only derived predictors, using the same random effect structure as for the full model. When we did this, we found that the fit of the full model was

	-tʃʲik	-ik	-ok
allomorph	**1.099**	(0.05169)	(-0.0201)
-tʃʲik-base	**0.1726**	(-0.02213)	-0.2547
-tʃʲik-dim	**0.7363**	-0.5391	(-23.13)
ik-base	(-0.0241)	**0.1126**	(-0.0331)
ik-dim	-0.2252	**0.7697**	(-23.48)
ok-base	0.07312	-0.3123	**0.2439**
ok-dim	(-23.6)	(-23.02)	**1.113**

R^2 likelihood: 0.392; AIC: 9795; BIC: 9940, VIF=5.95, κ=7.64

polytomous(allomorph∼tʃʲik-base+ tʃʲik-dim+ ik-base+ ik-dim+ ok-base+ ok-dim+ (1+ tʃʲik-base+ tʃʲik-dim+ ik-base+ ik-dim+ ok-base+ ok-dim|participant) + (1| base), heuristic=poisson.reformulation)

Correlation of model prediction vs. people's choice	
Spearman's ρ	0.63
Kendall's τ	0.59
Pearson's product-moment	0.68

Table 8: Polytomous regression model with sublexical phonotactic predictors for allomorph choice in the nonce word study

better based on R^2 likelihood and AIC/BIC than the fit of the other two models: for the base-only model with the predictors *ok-base, ik-base, tʃʲik-base*, the numbers were 0.017, 15780, 15870 respectively; for the derived-only model with the predictors ok-dim, ik-dim, tʃʲik-dim, the numbers were 0.378, 10010, 10090 respectively. While the contribution of the derived diminutive sublexical phonotactic grammars is larger than the contribution of base grammars, both types of grammars matter.

6.2.2 Regression model with phonological predictors

The other way to analyze the results would be to test the individual phonological predictors that we manipulated in designing the stimuli—these are the properties that appeared to be relevant for allomorph choice in the production study. We report a model that includes a number of binary predictors, all characterizing the bases: *finalC=dor*: [tʲanóx] vs. [sunʲés], *finalC=son*: [sʲipírʲ] vs. [sunʲés]*, final.CC:* [xʲirásn] vs. [sunʲés], *medial.CC:* [sʲubmár] vs [sunʲés] (residualized against *finalC=son*), *hiatus:* [pu.dú.al] vs. [pudú.lal], *medial.stress*: [ba.vʲí.lup] vs. [tʃʲu.nu.ráp] (residualized against *hiatus*).

As can be seen from this model summary, each predictor has an effect on the choice of at least one allomorph. For example, final clusters are more likely to combine with [-ik] and less likely to combine with [-tʃʲik] compared to the baseline poisson distribution. There is also an unexpected small increase in the likelihood of [-ok] for final cluster words, although this could be due to the forced choice nature of the study (the dispreference of [-tʃʲik] against clusters might be stronger than that of [-ok], in a trade-off). Medial clusters are less likely to combine with [-ik]. Following the lexical trends, final dorsals combine most freely with [-ok], and final sonorants with [-tʃʲik]. Medial stress increases the chances of [-ok]—unexpectedly, given the lexical trends. But bases with hiatus are less likely to be suffixed with [-ok] and more likely to be suffixed with [-tʃʲik]. One pos-

	- t͡ʃik	-ik	-ok
allomorph	0.945	(0.05451)	-0.8178
final.CC	-1.39	0.6474	0.2444
medial.CC	(0.1098)	-0.1456	(-0.03824)
finalC=dor	-0.4044	-1.936	1.452
finalC=son	0.2907	-0.4047	0.1722
medial.stress	(-0.09861)	-0.3935	0.7198
hiatus	0.2429	(-0.14)	-0.409

R^2 likelihood=0.09633, AIC=14540, BIC=14680, VIF=6.349, κ=8.241454
polytomous(allomorph$\sim$final.CC + medial.CC + finalC=dor +finalC=son+ medial.stress +hiatus + (1 + final.CC + medial.CC + finalC=dor +finalC=son+ medial.stress +hiatus | participant) + (1|word), heuristic=poisson.reformulation)

Table 9: Model for allomorph choice in nonce word study, with handpicked phonological features for the bases

sible explanation for the medial stress effect is that the other two suffixes have an even stronger dispreference for medial stress.

We were unable to calculate correlations between this model's predictions and people's choices for this model, but for a slightly pared down analogous model, we got correlations from .27 to .29—hovering around chance, given a 1 in 3 likelihood of any allomorph being chosen randomly. Thus, the fit of this model is poor compared to the model we reported in the previous section. (This is also visible from its low R^2 and high AIC/BIC compared to the sublexical phonotactic predictor model.) The handpicked predictor model performs worse on all measures of fit, although it does verify that people are sensitive to the properties we manipulated in the stimuli—they just appear to be using more information than our model includes.

6.3 "Was the choice difficult?"

Forced choice experiments artificially limit people's options, and they might actually prefer an option that is not available to them (such as "no diminutive would be better than any of these"). Therefore, we asked our participants whether the choice was difficult, using "yes/no". A choice can be difficult because there are multiple good options, not just because there are multiple bad ones. The results suggest, however, that people tended to find the choice difficult when all the options were phonotactically bad. The following plot shows the number of difficulty ratings for each suffix. People hit the "yes" button about the same number of times for each suffix, but they hit the "no" button much more often for the [-t͡ʃik] allomorph:

We analyzed the difficulty response in a hierarchical logistic regression model (*glmer* function in *lme4* in R, Bates 2005, R Development Core Team 2013), which is reported in Table 10. The model includes the pass/fail scores from sublexical phonotactic grammars for each suffix and random intercepts for participant and base. As can be seen from the negative intercepts, people are less likely to find the choice difficult when the base and the diminutive pass the respective sublexical phonotactic grammars—if the base or the diminutive passes the grammars, the choice is easy. The only exception is the diminutive grammar for [-ok], passing which meant a more difficult choice for participants. The reasons for this are not entirely clear—it is possible that people find it harder to
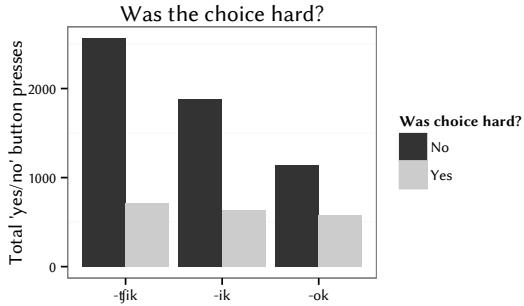
Figure 10: Was the choice difficult?

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.02555	0.11193	-9.162	$< 2e-16$***
ʧik-base	0.04731	0.07699	0.614	0.5389
ʧik-dim	-0.40887	0.07833	-5.220	$1.79e-07$***
ok-base	-0.33170	0.08067	-4.112	$3.93e-05$***
ok-dim	0.47566	0.09594	4.958	$7.13e-07$***
ik-base	-0.08010	0.07740	-1.035	0.3007
ik-dim	-0.23527	0.08493	-2.770	0.0056 **

VIF=1.35, κ=3.97

Table 10: Logistic regression model for people's answers about choice difficulty

choose between [-ok] and [-ʧik], which appears to be the default at least in this study's context. Finally, two of the base grammars, [ik] and [ʧik], did not have a significant effect in the regression.

7 Discussion

7.1 How people do the elicitation task

We investigated the hypothesis that people's choices of allomorphs depend on the well-formedness of the bases and the diminutive forms in sublexical phonotactic grammars. Both the elicitation study and the nonce word study support this hypothesis. People learn the sublexical phonotactic 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 the elicitation 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$_{+ok}$, . . .}

To phonology: /barsuk-ok/

Phonological output: [bərsutʲʃ́ók]

On the other hand, if the item 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 from the elicitation study that did well in just one of the subgrammars is [manʲják] 'serial killer'. Of the 8 people who got this prompt, seven produced [mənʲjatʲʃ́ók], rating it an average 4.12. One person produced [manʲjátʲʃik] 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 [-tʲʃik] base grammar.[9] The success in the [-ok] grammar determines the course pretty clearly, despite the good performance of [manʲjátʲʃik] in the diminutive grammar for [-ik]. There are diminutives in the [-ik] sublexicon that sound like this (e.g., [mʲátʲʃik] '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]	[-tʲʃik]
Base subgrammars	0	14.211	9.302
forms evaluated:	[manʲják]	[manʲják]	[manʲják]
Derived subgrammars	0	0	0
forms evaluated:	[manʲjatʲʃók]	[manʲjátʲʃik]	[manʲjáktʲʃik]
Cumulative	0	14.211	9.302

The most likely rule: DIM ↔[-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 [-tʲʃik] 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 [tʲʃik] base grammar.[10] 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]:

---

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

[10]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 [-tʲʃik] grammar is for the vowel sequence; the constraint is also *[−high][−high].

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

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

The most likely rule: DIM ↔[-ik]

Let's now consider bases that people found difficult to diminutivize in elicitation. 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). One participant produced no form at all, and the remaining productions were one each of [lʲémʲingʲik] and [lʲemʲingók], two [lʲémʲingtʃʲik], and a number of "other" productions ([lʲemʲingatʃók], [lʲémʲingʲiʂkə], [lʲemʲingótʃʲik], [lʲémʲinguʂkə], [lʲimʲingtʃók]). 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]	[-tʃʲik]
Base subgrammars	0	5.138	6.854
forms evaluated:	[lʲémʲink]	[lʲémʲink]	[lʲémʲink]
Derived subgrammars	4.439	0.905	2.167
forms evaluated:	[lʲimʲingók]	[lʲémʲingʲik]	[lʲémʲinktʃʲik]
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].

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

	[-ok]	[-ik]	[-tʃʲik]
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]	[aftóbustʃʲik]
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 written 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" even though it is attested in the Russian National Corpus, and our participants rated it 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.

A reviewer suggests that the comparative distance between the competitors within each base set might be a factor in deciding between allomorphs.

Note, however, that quite a lot of what happened in the elicitation 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 ↔[ok] / Vk__ really existed, then [manʲjátʃʲik] should never have been generated. We interpret such productions 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] 'lemming' and [súfʲiks] 'suffix'; more often than not, people produced each of [-ik], [-ok], and [-tʃʲik] 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 rather than writing it off as metalinguistic or uninteresting.

7.2   Non-local conditioning in the nonce word study

The main prediction we tested in the nonce word study was that well-formedness in each sublexical phonotactic grammar for bases and affixed diminutives should correlate with people's choices of allomorph in a forced choice task. This prediction is confirmed: people were more likely to choose nonce words with a particular affix if the base and the diminutive pass the respective sublexical grammars.

We were more specifically interested in testing non-local phonotactic conditioning of suppletive allomorphy in this experiment. The lexicon of Russian provides some mixed support for the effects of hiatus and medial stress placement on allomorphy. In Zaliznjak's dictionary (1977), only one word is listed with the [-ok] suffix and a hiatus sequence: [paúk]/[pautʃók] 'spider'. We found somewhat more words with hiatus in Google Ngrams, but most of them are not recognizable diminutives. However, both the elicitation study and the nonce word study turned up a result of hiatus: [-ok] was significantly less likely to combine with hiatus words than the other two suffixes.

The results for medial stress are more complicated. According to Polivanova's (1967) generalizations, medial stress is supposed to make diminutivization difficult in general—in part this is because [-ik] and [tʃʲik] just have a strong preference for final stress. The lexical trend is for [-ok] to combine with initial and final stress words but not medial stress words. In the elicitation study, this is exactly what we found, but in the nonce word study, the medial stress effect went in the

41

opposite direction. Our explanation for this is that the preferences of [-ik] and [ʧik] for final stress are stronger than than the dispreference of [-ok] against medial stress, at least for the stimuli we used. We believe that the best explanation of people's behavior here is sublexical phonotactic well-formedness; it may be a mistake to focus on individual phonological properties when it is the overall shape of the words that matters.

7.3 Is there anything the model cannot do?

The model we propose is powerful: it can identify conditioning factors in either the source or the product of affixation, and these factors may even be non-local to the affix itself. We have argued that this power is necessary to handle the Russian case. A fair question, however, is whether there are any patterns that this model cannot characterize.

One of the patterns that would be impossible to characterize is covert conditioning based on underlying representations that are cobbled together from multiple surface forms. For example, consider a language in which vowels delete word-finally (e.g., /badu/→[bad], but /bag/→[bag]) and are only seen in affixed forms (e.g., /badu-k/→[baduk]). Suppose also that this language has syncope: medial vowels delete, as in /badu-nik/ →[badnik] (a language that has both of these rules is Tonkawa). We would not be able to account for suppletive allomorphy between [-nik] and [-lik] where the [-nik] suffix selects for underlyingly vowel-final words, and [-lik] selects for underlyingly consonant-final words. This is because the learner would presumably assume that [bad] and [bag] are the bases for [badnik] and [baglik], and they both end in consonants in the unaffixed form. The allomorphy in such cases would have to be treated as completely lexical and arbitrary, and the "underlyingly vowel-final" generalization would not be extended to new forms. Some purported cases of this sort have been discussed in the literature (Paster 2006, inter alia), but we are skeptical of the productivity of the generalizations. In our theory, source-oriented phonological generalizations about output bases can be captured, but input-oriented ones cannot.

Another type of pattern that our model does not predict is extension through variegated similarity, as in analogical models (see Albright and Hayes 2003 for an in-depth discussion). In these models, words that share a similar substring can pattern together: for example, because "render" has "end" as its substring, "render" would form its past tense with [-əd] (incorrectly). In our model, it is possible for middles of words to matter, but similarity is more structured: patterns are defined by underattestation of various phonological strings and structures, not on shared substrings. Thus, there is an exceptionless generalization that all verbs that combine with [əd] end in a coronal stop in the unaffixed form, and a phonotactic learner will notice it; just having a shared string is not enough to motivate a phonotactic constraint.

Relating to this, consider the following hypothetical example of an implausible pattern, suggested to us by a reviewer: an "allomorph A is suffixed to stems that have a vowel [i] or [a] in the second syllable, whose base begins with a voiced fricative only if it ends in labial stop, whose initial stress in the base changes to penultimate stress in the derivative in two syllable words, and remains the same in three syllable words, otherwise allomorph B is prefixed to the stem." This type of case cannot exist given our assumptions about phonotactic learning, which are shared with Hayes and Wilson (2008). Computationally plausible phonotactic learners can easily discover constraints on feature co-occurrence in segments (e.g., "no voiced stops"), or bigram constraints (e.g., "no voiceless stops after nasals", or "no voiced stops word-finally"), or trigram constraints (e.g., "no voiceless stops between two vowels"). Past trigrams, the search space becomes much harder to explore. Hayes and Wilson demonstrate that certain generalizations, such as the distribution of vowels in Shona, require representational tiers, so that constraints on vowel harmony can be stated in a quasi-local way. Now, two MaxEnt constraints can gang up because constraint

weights are additive, but these types of interactions would not result in conditional affixation of the type whereby, say, words that start in voiced-fricatives and end in labials exclusively combine with a special allomorph. In our system, the co-occurrence of phonotactic violations in a single base makes affixation less likely, not more—and this accords with informal observations by linguists such as Polivanova (1967), who observe that dorsal-final bases ending with consonant clusters that have medial stress are virtually impossible to diminutivize in Russian. Thus, another type of pattern that would be difficult for our theory to characterize is one in which the mere co-occurrence of phonological properties in a base defines the affixation pattern. This is in stark contrast with the updated version of the Minimal Generalization Learner (Albright and Hayes 2006), which can indeed discover at least some accidental generalizations of this sort (see section 8.3).

Another reviewer asks how our theory would account for phonological conditioning of suppletive allomorphy by bound stems.[11] Even output-oriented theories do not need to preclude bases from being bound (see Benua 1997 for a particularly clear discussion). We do not want to make a firm commitment on this issue without a deeper exploration of at least one clear case of such conditioning. A marginal example of this might be English learned plurals (*syllabus/syllabi*), which people seem to generalize based on some phonological properties.[12] Presumably, people notice the atypical plural and group such items into sublexicons, at which point they can notice the Latinate stress patterns and other phonological features such as the extremely restricted final consonant inventory. If the presence of a derived form with a special affix is enough to trigger the creation of sublexicons, then bases can be bound, in a way that is completely consistent with our theory. Bound bases do not stymie implemented morphophonological learners such as the Minimal Generalization Learner or the Sublexical Learner (Becker and Allen submitted), the latter of which shares our assumptions about the nature of phonological generalizations in phonologically conditioned affixation.

A natural question to ask is whether learners really do search for generalizations exhaustively after a relatively simple one has been found. We argued that they do even in simple cases such as *a/an*, and we cite some evidence from the timeline of their acquisition in section 9. But it appears that learners can be oblivious to statistical trends in the lexicon when a phonologically relevant generalization suffices to figure out an irregular alternation (Becker et al. 2011, 2012). We want to emphasize that this issue is far from settled and requires further scrutiny: recall that vowels in affixes in Finnish and Hungarian can be conditioned by factors other than vowels (section 2.2.2, and Ringen and Heinamaki 1999, Hayes et al. 2009). We can speculate that cases like that, as well as the Russian case, represent the results of reaching the later stages in a stepwise search procedure. This procedure might start with the simplest and most local conditioning environments (consonants vs. vowels, say) and proceed to non-local enviroments that can be reduced to local via abstract representations (autosegments, feet, etc.). If none of these suffice, then full-on phonotactic learning over sublexicons is deployed. Even such learning is not guaranteed to find a useful generalization in cases where the allomorphy is not phonologically conditioned, and in such

---

[11]The same reviewer asks how our theory would deal with 'process-based morphology'. This includes patterns such as umlaut, stress/accent shift without an overt affix, and featural changes. We do not have much to say about them here, but there is no shortage of analyses of such patterns from a formalist phonological perspective that assumes item-based morphology (Benua 1997, Alderete 2001, Wolf 2007, Gouskova and Linzen to appear, inter alia). There are also accounts of the generalization of such patterns to novel items in the literature on inductive morphophonological learning (Albright and Hayes 2003, Becker and Allen submitted) and some heated debate in the morphological literature about analyzing such cases with readjustment rules (Gribanova 2014, Siddiqi 2009, Haugen and Siddiqi 2013).

[12]Some evidence for this comes from the treatment of s-final Greek words, which English speakers have some uncertainty about pluralizing. Witness "octopus", prescriptively "octopodes" in the plural, but often "octopi" in normal usage. The same is true for "nemesis": "*We've always been each other's greatest nemesises. . . uh, [səsaɪ]. . . neme[si]. . . wh-what's the plural on that?*" (*Mystery Men*, Universal Pictures, 1999).

cases, sublexical phonotactic grammars would not improve the fit to the learning data and would be discarded. This is not the model we justified in this paper, but it is a plausible refinement of the learning procedure that is the subject of our future work.

8   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 in this section.

8.1   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 main problem with this approach is well-known—there are many 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. On the other hand, in our approach, the generalization about vowel-finality for [-a] and consonant-finality for [-la] should be quite easy to state in terms of phonotactic constraints on bases.

Equally problematic for the emergence of the unmarked approach are cases of homophonous morphemes that impose distinct selectional restrictions. Consider the English comparative in "taller", 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").[13]

Accounting for allomorphy and morpheme-specificity in the same constraint hierarchy would be a challenge for this approach. One of the technical problems is that the crucial violation sites are somewhere other than in the morpheme that needs to be singled out for special treatment. For example, Onset is violated by both "a" and "an", but the violation that matters is the one incurred by the following morpheme. The learner must figure out that the crucial violation is contributed by the following noun (i.e., the second syllable of *[ə.əmi.bə]), but that this volation is only crucial when the indefinite appears nearby. There is no learning algorithm that works for indexing constraints with this locality profile—worked out proposals such as Pater (2008) rely on violations being contributed by the morphemes that require indexation, not the morphemes that are adjacent to them.

The dispreference of the Russian allomorph [-ik] against dorsal-final stems raises this issue, as well. It could be explained 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 problem is that the suffix itself is routinely a

---

[13]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.

part of such sequences in plurals: if [kʲi] is prohibited in *[kulakʲ-ik] 'fist-dim' (actually [kulatʃók]), why is it allowed in [kʲirpʲitʃ-ikʲ-i] 'brick-dim-pl'? In general, it is not clear how to capture the various restrictions seen in Russian diminutive formation in an emergence of the unmarked account. For example, the [-ok] suffix tends to avoid cluster-final bases—and yet the significant effect of medial cluster avoidance that we found was for the [-ik] suffix, not for [-ok]. Likewise, there is no markedness-based reason for why [-ok] should avoid bases with medial stress; stress always moves in [-ok] diminutives, so it shouldn't matter where it moved from.

This highlights a general difficulty for this markedness-based approach, noted by Paster (2006) and others: surface constraints cannot capture source-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 English "a/an", the surface syllabification does not necessarily 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 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.

Yet another problematic aspect of the Emergence of the Unmarked as a theory of suppletive phonologically conditioned allomorphy arises from combining this theory with parallelism (see Bobaljik 2000, Embick 2010, Nevins 2011). TETU with parallelism predicts that the roots can condition the choice of affixes, but the properties of affixes can also condition the allomorphy of roots—if both types of morphemes have multiple URs. For example, suppletive allomorphy of stems such as *go/went* can be conditioned by whether the following suffix begins in a vowel or a consonant, as in **go-ed* vs. **went-ing*. But it appears that cases like this are unattested.[14] The explanation for this typological gap in our version of Distributed Morphology is sequential inside-out insertion: roots are realized first, and their diacritic features can condition the insertion rules of outer morphemes. But the phonological shapes of outer morphemes cannot determine the shapes of inner ones. In our theory of phonological conditioning, the well-formedness of the entire derived word in its sublexicon can determine the likelihood of an allomorph being chosen for a novel item, but 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.[15]

---

[14]This generalization, due to Bobaljik (2000), is occasionally questioned—for example, Wolf (2013) identifies some cases that he argues involve outer conditioning. But these cases could be analyzed as involving morphosyntactic conditioning, or restricted morphophonological rules. See also Myler (to appear) for a case of apparently outward conditioning that actually arises because the conditioning environment is obscured by syntactic movement.

[15]Suppose, however, that the learner is exposed to a language with ubiquitous morphosyntactically conditioned stem allomorphy of the *person/people* variety, and she forms 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, Gribanova 2014) so alternatives to them are a welcome result.

8.2 Subcategorization frames and alignment

Another approach to selectional restrictions casts morpheme realization rules in terms of subcat-egorization 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↔ a / V ___

   b. DEF↔ 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 sys-tematic depending on speaker and dialect. To deal with sporadic exceptions, the subcategorization approach could add item-specific rules (e.g., DEF↔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 [-tʃik], 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 [tʃik] 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 ([lʲimónt͡ʃik] 'lemon-dim', [pʲinʲálit͡ʃik] '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$_{[+stress]}$C$_{[-dor]}$___

    b. DIM↔[-ok]/C+V[−stress]C+VC‗‗

    c. DIM↔[ʧʲik]/V[+stress]C[-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)(kɹír)ꜰ and in the attested bigger (bíg)ꜰ. 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.

8.3   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 pattern. The learner further relies on type frequencies within each class to decide how reliable the generalizations are. Unlike our approach, the learner induces rules that codify changes from bases to affixed forms, and it has no way of generalizing about sequences that are absent from certain sublexicons. When we trained the learner on Russian lexical diminutive formation and tested it on our nonce words, it failed to generalize in the way that people do, for reasons that bear investigating.

    The basic workings of the learner are sketched in (30). Given a list of base-derivative pairs, it identifies the changes in the string based on an alignment algorithm (e.g., the addition of substring "ik", and the palatalization changes). The learner posits individual rules summarizing the changes for each pair, and then generalizes over these rules if the contexts have phonological features in common. Here, there are some labial-final stems and coronal-final stems; the learner will spot the generalization and posit a palatalization+suffixation rule for each natural class. The learner is "minimal" in that it looks for the smallest environment that will suffice for a generalization about where the change happens.

(30)   How the Minimal Generalization Learner works

a.	skelét	skeletʲik	'skeleton'	[−hi, −lo, −back][cor, −voice, −cont, +back]→[−back]ik
b.	klʲón	klʲónʲik	'maple'	[−hi, −lo,+back][cor, +nas, −cont, +back]→[−back]ik
c.	klóp	klópʲik	'bedbug'	[−hi, −lo, +back][lab, −voice, −cont, +back]→[−back]ik
d.	komár	komárʲik	'mosquito'	[−hi, +lo,+back][cor, +son, −nas, −lat, +back]→[−back]ik
e.	krém	krémʲik	'cream'	[−hi, −lo, −back][lab, +nas, −cont, +back]→[−back]ik

More general rules: [cor, +back]→[−back]ik

                               [lab, +back]→[−back]ik

    This approach works well for English past tense formation in Albright and Hayes (2003). In this case, [voʊt]/[voʊtəd], [nɪt][nɪtəd], and [nid]/[nidəd] are unified under a rule that inserts [əd] after [+coronal, −continuant, −nasal]. This rule covers a large number of cases in English, but not [hɪt]/[hɪt], [pʊt]/[pʊt], or [sɪt]/[sæt]; these would motivate their own rules ("do nothing after a lax vowel followed by [t]" and "change [ɪ] to [æ] before [t]", respectively). Each rule is evaluated based

on a confidence score—rules that apply to a large proportion of eligible forms have higher raw confidence scores, and rules that apply to a large proportion and a large absolute number of forms also have higher adjusted confidence scores. These scores can be used to decide which rules to apply when generalizing the grammar to new words; rules that are more general are generalized more readily. This matches people's behavior in wug tests fairly well.

When trained on Russian diminutives, the Minimal Generalization Learner encounters problems, some of which are due to its inherent design, and others—to the implementation of the rule component. At a minimum, the learner should be able to discover rules for final consonant palatalization and mutation for all three suffixes and stress assignment in diminutives. To be fully successful, the learner also should match people's fine-grained discrimination of base phonotactic properties that are not local to the suffix: the role of hiatus, medial clusters, and final clusters. It has trouble with all of these tasks.

We trained the Minimal Generalization Learner on our sublexicons and tested it on the stimuli from the elicitation and forced choice experiments. The learner posited many rules for [-ik] and [ʧik]. None of the rules were as general as the one given in (30); most were specific to individual consonants. One of the problems for the learner is the stress shift in [-ok] diminutives. For a human linguist, the rule is rather simple to state: move stress to the final syllable. The rule is equally easy to capture as a product-oriented generalization over words suffixed with [-ok]—they always have final stress. The UCLA Phonotactic Learner identifies this generalization via two constraints formulated on the Stress tier, *[−stress]# and a constraint against stress anywhere except word-finally. The constraint against unstressed final syllables has the highest weight in the [-ok] diminutive sublexicon. On the other hand, the Minimal Generalization Learner defines its sublexicons operationally by treating each stress shift operation as its own rule; it is so focused on the source-oriented part of diminutive formation that it fails to capture the simple product-oriented generalization that stress is always final in [-ok]-derived diminutives. In part, this problem is implementational, since the learner treats the entire changed portion of the string as undergoing the rule—thus, for [sókol]/[sokolók], it posits a rule whereby [ókol]→[okolók]/s__, and the entirety of [úzʲel]/[uzʲelók] is encoded in a separate rule. If the learner had the rule vocabulary to state a stress flop operation more elegantly, it might do better. Nonetheless, a significant part of the problem stems from the focus on how stress is moved rather than on where it ends up—the sublexicon approach succeeds here precisely because it is given the product of [-ok] affixation and can find the phonotactic generalization about stress quite easily.

Final consonant changes present a problem of a different sort: it is easier to summarize the negative constraints on palatalization than to encode them in positive rules. In [-ik] diminutives, all stem-final consonants that aren't already palatalized become palatalized, except for the dorsals [k]/[g]/[x], which do not occur in the sublexicon, and the retroflex sibilants [ʂ] and [ʐ], which lack palatalized counterparts and do not alternate. (The affricate [ts] becomes [ʧ] in this morphological context.) The UCLA Phonotactic learner captures the lack of dorsals with a constraint *[−low]# in the base grammar, and it captures the lack of palatalized retroflex sibilants with *[+retroflex, −back] in the diminutive grammar. On the other hand, the Minimal Generalization Learner never manages to formulate general rules for palatalization before [-ik], because the complement of [k g x ʂ ʐ] in the Russian consonant inventory is not a natural class that can be easily defined in terms of phonological features. What the learner does instead is posit several consonant-specific palatalization rules, missing a fairly simple generalization.

In [-ok] diminutives, the problem of final consonant alternations is of a different nature. Quite generally, the dorsals [k], [g], [x] become [ʧ], [ʐ], [ʂ] (see Łubowicz 2002, Padgett 2003, and others). The Minimal Generalization Learner would be able to deal with [g] and [x] were stress out of the picture, and indeed, when we removed stress from the learning data, it found rules for this subclass

of dorsals. But it still mistreats the change affecting [k] as infixation: for [bík]/[bɨtʃók] 'bull', it posits a rule whereby [ɨ] becomes [ɨtʃó], since only the underlined part of the string is changing. This implementational issue can be addressed by biasing morphological learners against infixation (see Becker and Allen submitted on hypothesis reduction).

The biggest problem with the learner, however, is that it has a hard time identifying non-local conditioning environments. Albright and Hayes (2006) address the problem in the case of Navajo sibilant harmony, where the anteriority of an affix strident depends on the anteriority of stridents in the stem—whether adjacent to the affix or not. Their solution is to generalize over the entire string, looking for featural commonalities in all the sounds of the base. Given a sufficiently diverse training set, the learner will home in on stridents being the predictor of harmony. Unfortunately, this solution also identifies many other, unwanted generalizations that Albright and Hayes argue to not have any psychological reality for speakers. It should be apparent that, psychological reality nonwithstanding, this solution would not find the right predictors for Russian diminutives—for the simple reason that all Russian words have vowels and stress in them, but it isn't the presence of vowels and stress that predicts which affix appears but rather their positions. It is not clear how the learner can ever discover that words combining with [-ok] tend to not have hiatus in them—there is no feature-based generalization covering strings of consonants and vowels as opposed to vowels and vowels in the words that do combine with [-ok]. In order to notice that sequences of two [−consonantal][−consonantal] segments are underattested in the [-ok] sublexicon, the learner would have to generalize over sequences of [−consonantal][+consonantal] and [+consonantal][−consonantal]; it is impossible to find such a generalization in the learner's setup. It does not look at n-gram constraints because it is not equipped to look for constraints in the first place. Ultimately, Albright and Hayes conclude that constraints are a necessary part of the learning process, implemented in the UCLA Phonotactic learner (albeit for phonotactic learning only, not for morphophonemic learning).

Unsurprisingly, there was no correlation between the Minimal Generalization Learner's predictions for the Russian three-way forced choice experiment and people's behavior (0, meaning it was right about 1/3 of the time). The learner never generalized [-ok] to any of the nonce words. It did generalize the suffix to some elicitation stimuli, if only because they were lexical words of Russian and some of them ended in the same sequences as the words in the sublexicons. But the correlation between its choices and people's productions was again poor, and the same held for confidence ratings.

9   Acquisition of diminutive formation and other suppletive alternations

The last topic we discuss before concluding is the acquisition of sublexicons. Research on the acquisition of morphophonology suggests that it takes children a long time to attain adult mastery (Berko 1958 et seq.). It is well-known that younger children often overgeneralize a rule or an allomorph that is restricted in the adult language to items that are not supposed to undergo the rule (see, for example, Graziano-King and Cairns 2005 on English comparatives, or Köpcke 1998 on German plurals). Only fairly late do children start using adult-like forms correctly, and being able to apply rules to nonce forms comes even later. This late acquisition of adult-like use patterns receives a ready explanation if we assume that it takes time to assemble sublexicons of sufficient size, and that it takes time to extract phonotactic generalizations from them.

If the procedure we assume in our theory is applied to all suppletive allomorphy, then we would expect even the "simple" patterns like "a"/"an" suppletion to take a long time to learn correctly. And

indeed, Newton and Wells (1999) find that English-learning three-year olds produce "an" before vowel-initial words less than 30% of the time; the proportion increases to just 80% for seven-year-olds.[16] By comparison, the alternation [ði]/[ðə], whose conditioning environment is similar, increases from 35% at age 3 to 80% by age 5—adult-like patterns are learned much earlier for this alternation (which, while variable, is more robustly phonological in English; see Raymond et al. 2002).

When it comes to Russian diminutives, there are no studies that we know of that investigate the phonological aspects of their acquisition. The focus has been on their morphosyntax. What we do know is that diminutives are far more ubiquitous in child-directed Russian speech than in adult-directed speech (Kempe et al. 2007). This is consistent with the pragmatics of diminutive use, but it also has the happy side effect of simplifying certain properties of the Russian gender/noun classes and its stress system (this is an insight originally due to Olmsted 1994; see also Kempe et al. 2003, 2005, who demonstrate that diminutives actually facilitate the acquisition of gender). One longitudinal diary (Kempe et al. 2003, Protassova and Voeikova 2007) shows that the parent often refers to an object using both non-diminutive and diminutive nouns (e.g., the feminine noun [nagá]/[nóʂkə] 'leg (+dim)'), supplying the material for the sublexicons that the child would need to learn the generalizations.

Because the focus has so far been on the morphosyntax of diminutive formation rather than on the morphophonology, not much is known about the process by which children acquire the phonological generalizations about diminutive formation. Olmsted (1994) collects a few anecdotal examples suggesting that children's diminutives are not entirely adult-like, even though some rules and generalizations may already be in place. For example, children may expect words that end in a sequence analyzable as a diminutive to have a base (see (31)), opting for the simplest morphological analysis. Likewise, in forming their own diminutives, children apply some of the generalizations that we identified: thus, the medial-stressed [palkóvnⁱik] follows the pattern of words ending in [k]-final suffixes and is affixed with [-ək] (see (32a)). More interestingly, an anecdotal child diminutivizes /ovrág/ [avrák] 'ravine' with [-ok] (see (32b)) and mutates the final consonant to [tʃ], assuming it is [k]-final, even though it is really /g/-final (cf. [avrágu] 'ravine-dative'). This noun happens to belong to the small sublexicon of [-ək] bases, which contains both [k] and [g]-final nouns (recall section 3.2), but presumably the child has not heard or noticed the relevant data yet. This example reinforces the point that learners need to be exposed to examples such as [avrágu] and [avrázək] to know that the base consonant is supposed to mutate to [z] and that the more productive [-ok] suffix does not attach to this particular base.

(31) Some anecdotal backformations in Russian child speech

	"Diminutive" form	Child's backformation	Correct adult base	Age	
a.	[aduvántʃⁱik]	[aduván]	— (pseudodiminutive)	5;0	'dandelion'
b.	[agúrtʃⁱik]	[agúr]	aguréts	—	'cucumber'
c.	[lunátⁱik]	[lunát]	— (-ic loanword)	—	'lunatic'

[16]Marjorie Pak reports similar findings in an unpublished handout from an LSA talk, http://www.marjoriepak.com/LSA-pak-allomorphy-phonology.pdf.

(32) Some anecdotal diminutive formations in child speech

	Base form	Child's diminutive	Correct adult diminutive	Age	
a.	palkóvnʲik	palkóvnʲitʃʲək	palkóvnʲitʃʲək	—	'colonel'
b.	avrák	avratʃʲók	avrázək (/ovrag+ək/)	4y3mo	'ravine'
c.	zvʲérʲ	zvʲérʲik	zvirʲók	6y6mo	'animal'
d.	glás	glázʲikʲi (pl)	gláski, glazók (irregular stress)	6y5mo	'eye'

What has not yet been done is a systematic investigation of how children treat words that present phonological difficulties for diminutive allomorph choice, or whether children can generalize the rule to new items in the way that adults do.

10 Conclusion

We argued that knowledge about phonologically conditioned suppletion is represented via 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.

Alderete, John. 2001. Dominance effects as transderivational anti-faithfulness. *Phonology* 18:201–253.

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. 1994. Derivational productivity and text typology. *Journal of quantitative linguistics* 1:16–34.

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

Bates, Douglas M. 2005. Fitting linear mixed models in R. *R News* 5:27–30.

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

Becker, Michael, and Blake Allen. submitted. Learning alternations from surface forms with sublexical phonology. *Phonology,* URL `http://ling.auf.net/lingbuzz/002503`.

Becker, Michael, and Maria Gouskova. to appear. Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry* 47.

Becker, Michael, Nihan Ketrez, and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish devoicing neutralization. *Language* 87:84–125.

Becker, Michael, and Jonathan Levine. 2010. Experigen: An online experiment platform. URL `https://github.com/tlozoot/experigen`.

Becker, Michael, Andrew Nevins, and Jonathan Levine. 2012. Asymmetries in generalizing alternations to and from initial syllables. *Language* 88:231–268.

Belsley, David, Edwin Kuh, and Roy E. Welsch. 2004. *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: Wiley.

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Doctoral Dissertation, University of Massachusetts, Amherst, Amherst, MA.

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, 2049–2069. Chichester, UK & Malden, MA: Wiley-Blackwell.

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

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Oxford University Press.

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. 2015. Learning general phonological rules from distributional information: A computational model. *Cognitive science* 39:647–666.

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.

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 *Proceedings of the 32nd meeting of the chicago linguistic society CLS, 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.

Gagliardi, Annie, and Jeffrey Lidz. 2014. Statistical insensitivity in the acquisition of Tsez noun classes. *Language* 90:58–89.

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`.

Graziano-King, Janine, and Helen Smith Cairns. 2005. Acquisition of English comparative adjectives. *Journal of Child Language* 32:345–373.

Gribanova, Vera. 2014. Exponence and morphosyntactically triggered phonological processes in the Russian verbal complex. *Journal of Linguistics* .

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.

Hayes, Bruce, Kie Zuraw, Péter Siptár, and Zsuzsa Cziráky Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85:822–863.

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.

Kapatsinski, Vsevolod. 2011. What statistics do learners track? rules, constraints or schemas in (artificial) grammar learningules, constraints or schemas in (artificial) grammar learning. In *Frequency effects in language: Learning and processing*, ed. S. Th. Gries and D. Divjak. Mouton de Gruyter.

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 Steven Gillis. 2005. Diminutives in child-directed speech supplement metric with distributional word segmentation cues. *Psychonomic Bulletin and Review* 12:145–151.

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

Kempe, Vera, Patricia Brooks, Natalija Mironova, Angelina Pershukova, and Olga Fedorova. 2007. Playing with word endings: Morphological variation in the learning of Russian noun inflections. *British Journal of Developmental Psychology* 25:55–77.

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. Tuebingen: Max Niemeyer Verlag.

Koch, Harold. 1980. Kaititj nominal inflection: some comparative notes. In *Papers in australian linguistics*, ed. B. Rigsby and P. Sutton, volume 13, 259–276. Canberra: Australian National University.

Köpcke, Klaus-Michael. 1998. The acquisition of plural marking in English and German revisited: Schemata versus rules. *Journal of child language* 25:293–319.

Labov, William, and Wendell Harris. 1986. De facto segregation of black and white vernaculars. In *Diversity and diachrony*, ed. David Sankoff. Amsterdam: Benjamins.

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.

Łubowicz, Anna. 2002. Derived Environment Effects in Optimality Theory. *Lingua* 112:243–280.

Luce, Paul A. 1986. Neighborhoods of words in the mental lexicon. Doctoral Dissertation, Indiana University, Bloomington, IN.

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.

Nash, David. 1980. Topics in Warlpiri Grammar. Doctoral Dissertation, MIT, Cambridge, MA. Published by Garland Press, New York,1986.

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

Newton, Caroline, and Bill Wells. 1999. The development of between-word processes in the connected speech of children aged between 3 and 7 years. In *Pathologies of speech and language: Advances in clinical phonetics and linguistics*, ed. Ben Maassen and Paul Groenen, 67–75. London: Whurr Publishers.

Olmsted, Hugh. 1994. Diminutive morphology of Russian children: A simplified subset of nominal declension in language acquisition. In *Alexander Lipson: In memoriam*, ed. Charles Gribble, Robert Rothstein, Edythe Haber, Hugh Olmsted, Robert Szulkin, and Charles Townsend, 165–209. Columbus, OH: Slavica Publishers.

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.

Pater, Joe. 2008. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In *Phonological argumentation: Essays on evidence and motivation*, ed. Steve Parker, 123–154. London: Equinox.

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.

Pinker, Steven, and Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28:73–193.

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 nat-sional'noj shkole. reprinted in Anna K. Polivanova, 2008. Obscheee russkoe jazykoznanie: Izbrannye raboty*, volume 4, 8–22. Moscow: RGGU.

Protassova, Ekaterina, and Maria Voeikova. 2007. Diminutives in Russian at the early stages of acquisition. *Language acquisition and language disorders* 43:43.

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.

Ringen, Catherine O., and Orvokki Heinamaki. 1999. Variation in Finnish vowel harmony: An OT account. *Natural Language and Linguistic Theory* 17:303–337.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80:475–532.

Ryan, Kevin. 2014. Onsets contribute to syllable weight: Statistical evidence from stress and meter. *Language* 90:309–341.

Siddiqi, Daniel. 2009. *Syntax within the word: Economy, allomorphy, and argument selection in distributed morphology*, volume 138. Amsterdam: John Benjamins Publishing Company.

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. 2007. For an autosegmental theory of mutation. In *UMOP 32: Papers in Optimality Theory III*, ed. Leah Bateman, Michael O'Keefe, Ehren Reilly, and Adam Werle, volume 239–258. Amherst, MA: GLSA.

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`.

Wurm, Lee H, and Sebastiano A Fisicaro. 2014. What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language* 72:37–48.

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.