Source-oriented generalizations as grammar inference in Russian vowel deletion*

Michael Becker
Stony Brook University
michael.becker@stonybrook.edu

Maria Gouskova
New York University
maria.gouskova@nyu.edu

Version of August 21, 2015

abstract

Speakers learn detailed generalizations about the morphophonology of their language, and extend them to nonce words. We propose a theory of this morphophonological knowledge that partitions the lexicon into uniform and productive sublexicons. Each sublexicon has its own phonotactic grammar, which the speaker uses as an inference mechanism to determine the relative productivity of each sublexicon. We report the results of an experiment on the generalization of mid vowel deletion ("yer" deletion) in Russian, showing that speakers encode source-oriented generalizations about the shapes of words that can undergo vowel deletion, as well as product-oriented generalizations about words that result from vowel deletion. An implementation of our model learns the patterns of deletion and captures both source-oriented and product-oriented generalizations.

1   Introduction

It is known that speakers encode detailed generalizations about morphophonological alternations, including those that are only partially productive. A central goal of linguistic theory is the modeling of speaker knowledge of such generalizations, in particular as it determines the treatment of novel words. In this paper, we discuss a partially productive mid vowel deletion process in Russian (e.g., [kast^j<u>o</u>r ∼ kastr-a] 'fire'), and draw attention to a generalization about the kinds of words that allow deletion: they must end in a single consonant. Words that end in a complex coda never undergo deletion, e.g., [as^jotr ∼ as^jotra] 'sturgeon' *[astra]. This is a source-oriented generalization, stated over the bases that undergo deletion, as opposed to a product-oriented generalization, stated over the derivatives that underwent deletion. This generalization calls into question Bybee's claim that "any morphological pattern that can be described by a source-oriented rule can also be described by a product-oriented one" (Bybee 2001:129). The generalization also eludes constraint-based approaches that rely on markedness and faithfulness only.

From the perspective of constraint-based approaches (starting with Optimality Theory, Prince & Smolensky 1993/2004) source-oriented generalizations can be classified into three categories: (i) demanding input-output identity, formalized with faithfulness constraints, (ii) demanding a change from input to output, expressed in terms of the realization of morphemes, or (iii) banning a marked structure from the input. It is this last type of generalization, observed in Russian, that is impossible to express using the theory's standard product-oriented markedness constraints.

Yet we do not propose a return to a source-oriented theory; rather, we show that the generalization is captured by positing a pair of grammars, each of which is product-oriented. In the model we propose, grammars can only capture product-oriented generalizations; the source-oriented generalization is captured by a phonotactic grammar that functions as a grammar inference mechanism, assigning sources to the grammar that is best suited to derive them.

The remainder of the paper is organized as follows: §2 reviews the generalizations about words that undergo vowel deletion in the Russian lexicon. The results of a nonce-word study that tests these generalizations are presented in §3. Our theory of source- and product-oriented generalizations is presented in §4, which also summarizes the results of the learning simulation based on this approach, and §5 concludes.

2   The yer sublexicon

Deletion of the mid vowels [e, o] (traditionally called "yers") in the real nouns of Russian is irregular: in some nouns, the addition of a vowel-initial suffix such as the genitive [a] leaves the nominative base intact (Table 1a), while in other nouns the mid vowel

	nominative	genitive	
a.	nos	nosa	'nose'
b.	rot	rta	'mouth'

Table 1: irregular deletion of mid vowels in Russian

of the base is deleted (Table 1b). The lexically specific, or unpredictable nature of yer deletion in individual nouns has long been recognized (Lightner 1965; Halle 1973a; Melvold 1989; Farina 1991; Yearley 1995; Gouskova 2012; Gouskova & Becker 2013). Much of this literature debates mechanisms of encoding the lexically-specific behavior, or the interactions between yer deletion and other rules of Russian phonology.

Yet more recent work recognizes that there are generalizations about yer deletion that make some predictions possible. For example, when yer deletion creates a triconsonantal cluster, the middle consonant is almost always an obstruent, as in [kastʲor ∼ kas<u>t</u>ra] 'fire'. Examples with a middle sonorant exist, but they are much rarer, e.g., [agnʲets ∼ ag<u>n</u>tsa] 'lamb'; usually a middle sonorant blocks deletion, e.g., [mudrʲets ∼ mudrʲetsa] 'wise', *[mud<u>r</u>tsa] (Yearley 1995; Gouskova 2012; Gouskova & Becker 2013). This restriction means that yer nominatives (i.e. nominatives that lose a vowel in the genitive) and non-yer nominatives (those with a faithful genitive) are different: a post-consonantal sonorant before a final vowel (i.e., CRVC#) is rare among the nominative forms of yer words, but common among non-yer word nominatives. Table 2 provides a quantitative view of this generalization: nominatives that ends in CRVC# account for 20% of the non-yer words, but given a CRVC# nominative, the probability of it losing its final vowel in the genitive is less than 1%.

The purpose of this section is two-fold: first, to survey the yer sublexicon (the set of all yer words), comparing it to the general Russian lexicon. The second goal is to examine the differences between yer words and non-yer words, and to relate these differences to Bybee's notion of source-oriented vs. product-oriented generalizations.

Our discussion of the Russian lexicon largely follows Gouskova & Becker (2013). The survey is based on the 20,563 masculine second declension nouns extracted from an electronic version of Zaliznjak's (1977) dictionary (Usachev 2004). The survey is limited to masculine second declension nouns, in the nominative singular and genitive singular, to match the items used in the experiment reported in §3 below. The reason for this limitation is that feminine and neuter paradigms sometimes lack a vowel-deleted form altogether ("paradigm gap"), and thus add a further complication (Halle 1973b; Hetzron 1975; Pertsova 2005). Of these 20,563 nouns, 1,902 nouns (9%) were identified as yer nouns, and the remaining 18,661 nouns were identified as non-yer nouns.

Structure in NOM	yer words	non-yer words	prob. of deletion	type of generalization
Complex coda	0%	17%	0%	source-oriented (§2.1)
CRVC#	<1%	20%	<1%	product-oriented (§2.2)
TVR#	6%	22%	3%	product-oriented
CTVR#	1%	7%	3%	product-oriented
Monosyllabic	1%	8%	1%	ambiguous (§2.3)
Mid vowel	>99%	37%	22%	source-oriented or faithfulness
Total	1,902	18,661	9%	

Table 2: Generalizations about yer deletion in the Russian lexicon, with the probability of vowel deletion in the genitive given the presence of a structure in the nominative.

2.1 Source-oriented: complex coda blocking

A final consonant cluster (complex coda) is not rare in Russian, found in 17% of the non-yer words, e.g., [pamost] 'stage'. But complex codas are completely absent from yer words, i.e. a paradigm such as *[pamost ∼ pamsta] is unattested (see Table 2). The unattested genitive form [pamsta] is a possible word of Russian, since triconsonantal clusters abound, and they are also a possible product of yer deletion (cf. [kastʲor ∼ ka<u>str</u>a] 'fire'), but they are not a possible product of yer deletion from a nominative form that ends in a complex coda. Therefore, this is a source-oriented generalization, stated over the nominative base. While many source-oriented generalizations can be analyzed as effects of faithfulness (see §4.3 below), this source-oriented generalization is defined in terms of markedness, or an offending structure in the base.

2.2 Product-oriented: sonority sequencing preference

There are generalizations about the sonority profile of the clusters that result from yer deletion: sonorants are almost never trapped in the middle of a CCC cluster. More nuanced statements about preferred and dispreferred sonority profiles are missing from the literature, because Russian is famous for allowing highly marked clusters (Blevins 1995; Clements 1990; Davidson & Roon 2008), and even when judging clusters in non-words, Russian speakers are extremely good at dealing with a wide range of clusters (Berent et al. 2007, 2009). And yet, as we show in §3, when clusters are created by vowel deletion, Russian speakers prefer those that end in an obstruent followed by a sonorant (i.e., TR, RTR, and TTR) over all other clusters. This means, for example, that the TR cluster in the nonce [ʂom ∼ ʂma] is preferred over the RT cluster in the nonce [moʂ ∼ mʂa]. We claim that this is a product-oriented generalization. Interestingly, the preference for

(C)TR products is not clearly seen in the lexicon study; as Table 2 shows, given a (C)TVR# nominative, deletion is not particularly probable.

One might object that the relevant consonants are found in the same order both in the nominative source and in the genitive product, and that the generalization is therefore compatible with either orientation, e.g., the nonce [ʂom] is preferred because it has an obstruent followed later in the word by a sonorant, unlike the nonce [moʂ]. There are three reasons, however, to think that speakers do not track sonority profiles of consonants that are separated by a vowel: an argument from typology, an argument from learnability, and an argument internal to Russian.

Cross-linguistically, restrictions on sonority profiles within consonant clusters are ubiquitous, for example, the difference in English between the attested [pl] in [pleɪ] 'play' and the unattested [pt] in *[pteɪ] (Clements 1990, inter alia). On the other hand, restrictions on sonority sequencing across vowels are unattested: there is no known language that allows [papa], [nana], and [pana] but rules out *[napa], i.e. there is no language that requires later consonants to be more sonorous or less sonorous than earlier consonants, all else being equal.[1] Regardless of whether such long-distance restrictions on consonant sonority are harder for learners to discover or whether they are ruled out by universal markedness principles, they are not typical of human language phonologies.

From the perspective of learning, there is no known mechanism that tracks the relevant sonority profiles in the nominative, where they are separated by a vowel, and also finds broader generalizations about sonority profiles in the language. This will be shown in §4 below, using two different source-oriented learners. On the other hand, available theories of phonotactics can and do learn the relevant preference for (C)TR clusters in the genitive products, as will be shown for the Russian data in §4. The preference for (C)TR clusters, then, is properly stated over the clusters in the genitive product of deletion, i.e., it is a product-oriented generalization.

Thirdly, the experimental results in §3 show that Russian speakers prefer the TR sonority profile in the genitives only. In the nominatives, they give the highest ratings to words whose last two consonants are sonorants.

2.3 Ambiguous generalizations

Two additional generalizations about Russian mid vowel deletion can be characterized as either source-oriented or product-oriented. We suggest that both may be due to faithfulness effects.

Monosyllables are common in the non-yer lexicon, making up 8% of it (see Table 2). In the yer lexicon, monosyllables exist (e.g., [son ∼ sna] 'dream'), but they are rare, making

---

[1]Obviously, there are many other ways in which nonadjacent consonants can interact, e.g., via nasal harmony, the limiting of coda consonants to sonorants, disallowing word-initial sonorants, etc., but these patterns are radically different from the patterns seen within consonant clusters.

up less than 1% of all yer words. This generalization can be interpreted as either source-oriented or product-oriented, since both the base and the derivative have just one syllable. The deletion in a monosyllable like [son] affects the initial (and only) syllable of the base, and thus can be seen as violating the special faithfulness that protects initial syllables (Becker et al. 2011, 2012; Beckman 1997, 1998; Gouskova & Becker 2013).

Finally, of the five vowels of Russian [a, e, i, o, u], only mid vowels delete. This is virtually exceptionless in the lexicon (Gouskova 2012; Gouskova & Becker 2013). Gouskova & Becker (2013) show that speakers extend the limitation to novel words in a rating study, finding the deletion of a mid vowel in, e.g., [riṣon ∼ riṣn-a] more acceptable than the deletion of the high vowel in, e.g., [karut ∼ kart-a]. This generalization can be seen as source-oriented, since the quality of the vowel is defined only in the base. Alternatively, however, the generalization can be seen as general faithfulness (all base vowels must be present in the derivative) overridden by a product-oriented markedness constraint against mid vowels in the derivative.

2.4 Summary

The distribution of vowel deletion in Russian is partially predictable. One predictor is clearly source-oriented: deletion is blocked by a complex coda in the base. Another predictor is product-oriented: derivatives are preferred if the cluster created by deletion ends in an obstruent followed by an a sonorant. Two additional generalizations were ambiguous: only mid vowels can delete, and deletion is dispreferred when it impacts monosyllables.

3 Experiment: wug-testing yer deletion

The following nonce word task ("wug test", Berko 1958) checked whether Russian speakers extend the lexical trends discussed in §2 above to nonce words. In particular, we test the source-oriented generalization about the syllabic profile of the nominative base: deletion is blocked from a base like [poṣm] due to its final complex coda, while deletion from a simple-coda base like [pṣom] is allowed, even though both give rise to the same genitive form [pṣma]. Looking at the sonority profile within clusters in the product of deletion, we test obstruents followed by a sonorant (TR), e.g., [ṣl], [mṣl], [pṣl], compared to clusters that end in two sonorants (RR) or two obstruents (TT), e.g., [xmn], [xṣt]. Monosyllabicity (e.g., [kipoṣm] vs. [poṣm]) and the choice of mid vowel (e.g., [poṣm] vs. [pʲeṣm]) were also counterbalanced.

3.1 Participants

The experiment was built using `Experigen` (Becker & Levine 2012), and posted online. All participants volunteered their time and effort. The website was accessed by hundreds of people, but most of them only responded to a few items. We report the data from a total of 115 participants who responded to at least 73 of the 74 items they saw, for a total of 8,505 data points, of which 5,516 were responses to target items. Data from other participants was discarded. Participants were recruited on Russian social network websites (`odnoklassniki.ru`, `vkontakte.com`, `livejournal.com`) and through word of mouth.

The participants were asked for demographic information about their place of birth, age, and other languages that they had studied. There were 68 females, 34 males, 13 didn't say. Age was given by 111 participants with a range of 18–64, mean 34, median 32. As to place of birth, 98 said they were from Russia, 6 were from Ukraine, 6 were from some other country of the former Soviet Union, 5 didn't say or said they were born in a non-Russian-speaking country. Most participants claimed to have knowledge of some other language, with 85 mentioning English, 26 French, 20 German, and other European languages. As for languages with Russian-like qualities (mid vowel deletion and non-English-like clusters), 9 participants said they had some knowledge of Ukrainian, 6 Polish, and 2 Hebrew.

The server log indicates that the 115 participants took on average 16 minutes to complete the survey (range 6–48, median 14).

3.2 Procedure

The experiment was conducted online, with participants free to choose their favorite browser. Each participant rated a randomly selected set of 74 items: 48 target paradigms (one each of the forms in Table 3, each with a randomly chosen unique consonant combination) and 25 filler paradigms, plus the sample item [kudat]. The 74 items were randomized and randomly combined with 74 unique frame sentences.

The sample item [kudat] was presented with a suggested rating of 5. Its two genitives were the faithful [kudata] and unfaithful [kudasa], and it was suggested that the first would be acceptable while the second would not be. Word stress or other alternations were not mentioned in the instructions; Gouskova & Becker (2013) did not find any effect of stress position on ratings of spoken nonce words.

Both target items and fillers were presented as schematized in Figure 1.[2] The unaffixed base form was presented first, in a random frame sentence where the base form would be

---

[2]The original Russian instructions were as follows: В этой реке водится длинный шер // Оцените подчеркнутое слово. Может ли оно быть словом русского языка? // не может 1 2 3 4 5 может // Иван поймал длинного шра. // А может это слово быть склоненным вариантом слова "шер"? // нет да // Иван поймал длинного шера. // А может это слово быть склоненным вариантом слова "шер"? // нет да.

In this river, there lives a long <u>ṣer</u>

Rate the underlined word. Can it be a word of Russian?

it cannot [1] [2] [3] [4] [5] it can

Ivan caught a long <u>ṣra</u>
Can this word be a declined variant of the word <u>ṣer</u>?
[no] [yes]

Ivan caught a long <u>ṣera</u>
Can this word be a declined variant of the word <u>ṣer</u>?
[no] [yes]

Figure 1: Translation of an example trial showing a nominative base and two genitive derivatives. The order of the two genitive forms was randomized, shown here with the unfaithful, vowel-deleted genitive first.

interpreted as a nominative singular noun. The participant was asked to rate it on a scale of 1 to 5 as a possible word of Russian.[3] Once the nominative was rated, an additional sentence appeared, this time leading to an interpretation as a genitive singular noun. The genitive form was either faithful (identical to the nominative form, plus the genitive suffix) or vowel-deleted (suffixed but missing the base's final mid vowel), presented in randomized order. The participant was asked to decide whether or not the genitive form was acceptable. Once the first genitive form was judged, a second genitive form was shown, again asking for a yes-or-no judgment.

The 74 frame sentences had slots that required a nominative singular masculine noun and a genitive singular masculine noun. Some of the contexts called for an animate accusative singular masculine noun; the suffix -a is the same for both genitives and animate accusatives. We refer to the suffixed context as "genitive" throughout the following discussion.

3.3 Materials

Based on a total of 403 different consonant combinations, a pool of over 50,000 potential nonce paradigms was created. Each paradigm consisted of three words: a nominative base with a mid vowel in its final syllable, e.g., [ṣox], and two genitive forms made by adding the suffix [a], e.g., [ṣoxa] and [ṣxa]. Similarly, the vowel [e] was used to form the

---

[3]We used a scale of 1–5 because Russians are familiar with it from secondary school grading. Weskott & Fanselow (2011) show that numerical scales are no less informative than magnitude estimation, and they reduce undesirable variance in the rating data.

			monosyllables		disyllables	
simple coda	(CV)CVC	TR	ToR	TeR	CVToR	CVTeR
		RT	RoT	ReT	CVRoT	CVReT
		RR	RoR	ReR	CVRoR	CVReR
		TT	ToT	TeT	CVToT	CVTeT
	(CV)CCVC	TTR	TToR	TTeR	CVTToR	CVTTeR
		RTR	RToR	RTeR	CVRToR	CVRTeR
		TRR	TRoR	TReR	CVTRoR	CVTReR
		TTT	TToT	TTeT	CVTToT	CVTTeT
complex coda	(CV)CVCC	TTR	ToTR	TeTR	CVToTR	CVTeTR
		RTR	RoTR	ReTR	CVRoTR	CVReTR
		TRR	ToRR	TeRR	CVToRR	CVTeRR
		TTT	ToTT	TeTT	CVToTT	CVTeTT

Table 3: Nonce word types. Each participant saw one each of the 48 word types.

paradigm [ʂex ∼ ʂexa, ʂxa].[4] These nonce paradigms covered a wide range of consonant clusters in a variety of phonological environments, as schematized in Table 3. We attached randomly chosen CV sequences (made of a consonant and a non-mid vowel) such as [pi] to create CVCVC bases such as [piʂex] and [piʂox], again with two faithful genitives and the vowel-deleted genitive [piʂxa].[5]

Consonant triplets were used to make CCVC and CVCC bases, e.g., the consonant triplet [ʂxl] was used to make the monosyllables [ʂxel], [ʂxol], [ʂexl], and [ʂoxl], a faithful genitive for each, and the single vowel-deleted genitive [ʂxla]. Randomly chosen initial CV sequences were attached as above.

Obstruents (T) and sonorants (R) from Table 4 were chosen randomly, except for [j] and [ʃʃʲ], which do not combine freely with other consonants, and [k kʲ] (excluded for morphological reasons, see below).

For (CV)CVC nominative bases, all possible combinations of sonorants and obstruents were used: TR, RT, RR, TT. For (CV)CCVC and (CV)CVCC bases, the combinations TTR, RTR, TRR, and TTT were used. Since the distribution of sonorants in Russian CCC clusters is rather limited, we used only one of the three combinations that puts a sonorant before another sonorant (namely T̠RR) and one of the four combinations that puts a sonorant before an obstruent (namely R̠TR). It has already been established experimentally in Gouskova & Becker (2013) that T̠RT is quite bad. Some consonant

---

[4]The vowel [e] was spelled with the Cyrillic letter ⟨e⟩, leading the consonant letter before it to be read palatalized, while elsewhere consonants were to be read velarized (see §3.5).

[5]We used the following 37 CV combinations: [pi pu bi ba bu mi ma mu ti ta tu nu si su zi zu ʂi ʂa ʂu ʐi ʐa ʐu ri ra ru li la lu ki ka ku gi ga gu xi xa xu]. We excluded CV combinations such as [pa] and [ni], which sound like existing prefixes of Russian.

		labial	dental	retroflex	(pre)palatal	velar
T	plosives	p pʲ b bʲ	t tʲ d dʲ			(k kʲ) g gʲ
	affricates		ʦ		ʧʲ	
	fricatives	f fʲ	s sʲ z zʲ	ʂ ʐ	(ʃʃʲ)	x xʲ
R	nasals	m mʲ	n nʲ			
	liquids		l lʲ r rʲ			
	glides	v vʲ			(j)	

Table 4: The consonants of Russian, showing the obstruents (T) and sonorants (R). Consonants in parentheses were not used in the target items.

profile	*n*	examples
TR	33	bv, xm, fm, ʂn, gm, xv, ʐl, pm, ʧʲr, ʧʲn, …
RT	25	rs, lʂ, np, mb, rʂ, nx, mʂ, ng, rb, nb, …
RR	5	mv, lr, nr, rn, mn.
TT	39	tf, ʂʧʲ, px, ʐd, bd, zg, pʂ, gʐ, …
RTR	61	npr, mzm, mgv, lfm, mʂl, mʂm, lzr, ldv, rʂm, nzr, …
TTR	98	fpv, pʂl, ʦʂr, ʐdr, xʂl, ʂxm, ʦn, txl, pʂm, ʧʲʂl, …
TRR	23	ʦnr, zmv, pmv, fmn, xmn, ʦmn, ʦmv, ʧʲnr, zmn, ʦrn, …
TTT	119	ʧʲfp, xʂx, bdz, ʐdz, fʧʲs, ʧʲfx, txs, ʧʲʂp, ʦʂx, xʂt, …

Table 5: examples of consonant combinations used in target items

combinations were excluded for phonotactic reasons (e.g., voicing disagreement). Others were excluded for morphological reasons, e.g., [k]-initial clusters were excluded to prevent an interpretation as the preposition [k-] 'toward', and [k]-final clusters were excluded due to the the diminutive suffix [-(o/e)k]. We thus ensured that stimuli are interpreted as monomorphemic. Additionally, some clusters were eliminated for appearing in real words, e.g., [dl] was excluded because it appears in the word [dol] 'valley'.

These criteria yielded a list of 102 consonant pairs. Of the possible consonant triplets, those that began and ended with the eligible pairs were chosen, e.g., the triplet [ʂxl] was chosen because both [ʂx] and [xl] were in the list of eligible pairs. This yielded 301 consonant triplets. A total of 3698 nonce words were tested, for an average of 14 ratings per consonant combination. Some of the consonant combinations used in the experiment are shown in Table 5.

There were 45 filler paradigms, each with one faithful genitive and one genitive with an unattested alternation in the manner or place of articulation of the last consonant, e.g., [ʧʲipan ∼ ʧʲipan-a, ʧʲipa<u>m</u>-a].
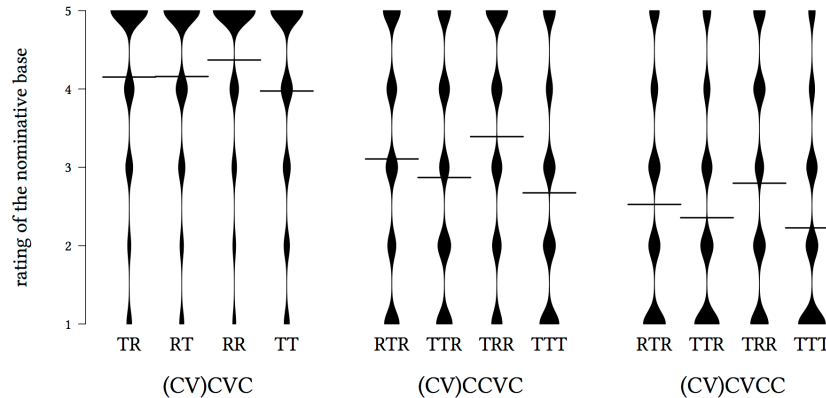
Figure 2: Rating of the nominative base (115 participants), by cluster position and sonority profile. Cluster-free bases are best, complex codas are worst. Bases whose last two consonants are sonorants (RR and TRR) are preferred.

3.4 Results

This section starts with a survey of the participants' treatment of the nominative base in §3.4.1, followed by their treatment of the faithful genitive in §3.4.2, then their treatment of the unfaithful, vowel-deleted genitive in §3.4.3. A regression analysis is given in §3.4.4. The raw experimental results are available at `http://becker.phonologist.org/projects/yers/`.

In each subsection, we discuss four generalization: the effect of the complex coda in the base, the preference for a TR sonority profile, monosyllabicity, and the choice of mid vowel ([e] vs. [o]).

3.4.1 Nominative bases

Participants gave the highest ratings to the cluster-free nominative forms (CV)CVC (4.16 on the 1–5 scale), followed by forms with a non-final cluster (CV)CCVC (3.01), and forms with complex codas (CV)CVCC (2.48), as seen in the bean plot in Figure 2. Within each of these three categories, words whose last two consonants are sonorants, i.e., (T)RR, were rated higher than words with one or no sonorants (3.52 vs. 3.12).

Monosyllabicity had a strong but inconsistent effect across categories: monosyllables were rated higher in the cluster-free forms (4.36 vs. 3.97) and in the complex coda forms (2.71 vs. 2.24), but not in the non-final cluster forms (2.80 vs. 3.22). The mid vowel used ([e] vs. [o], e.g., [ʂel] vs. [ʂol]) had a minuscule effect (3.20 vs. 3.23), and the effect was very small across phonological sizes and sonority profiles.
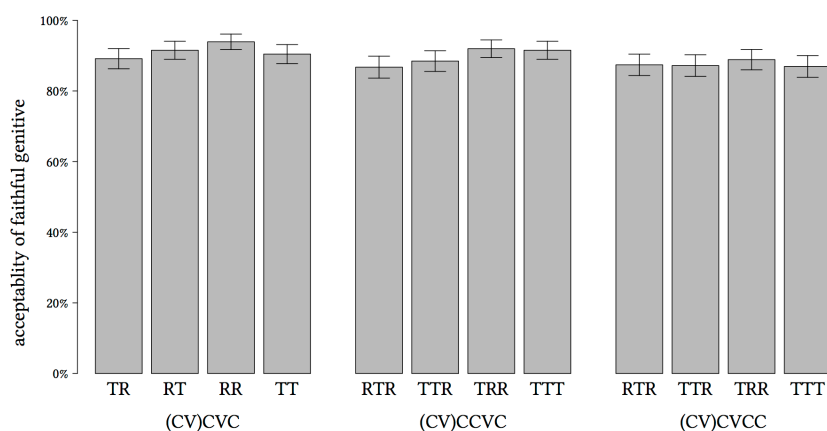
11

Figure 3: Acceptability of the faithful genitive (115 participants). The faithful genitives are highly acceptable, especially (T)RR.

3.4.2 Faithful genitives

The faithful genitive was found acceptable in 90% of all trials, as seen in Figure 3. In this figure and throughout the paper, error bars represent 95% confidence intervals. This high acceptability changed little across categories, but nevertheless mirrored the pattern of the nominatives, with 91% for cluster-free forms, 90% for forms with a cluster before the base's final vowel, and 88% for forms with a cluster after the base's final vowel. Again mirroring the rating of the nominative forms, (T)RR forms are slightly more acceptable than others (92% vs. 89%). The monosyllabicity effect mirrored that of the nominatives as well.

3.4.3 Vowel-deleted genitives

The overall acceptability of the vowel-deleted genitives was 25%, as seen in Figure 4. Vowel-deleted genitives were most acceptable for cluster-free bases (40%), less acceptable for bases with a non-final cluster (21%) and least acceptable for bases with a complex coda (15%). It is rather striking that vowel-deleted genitives were judged differently based on the syllabic profile of their base; recall that bases with both non-final and final clusters give rise to the same tri-consonantal clusters in the vowel-deleted genitive, e.g., [pişxla] from either [pişxol] or [pişoxl].

The effect of sonority profile does not mirror the ratings of the nominative forms: (T)RR vowel-deleted genitives are less acceptable than others (22% vs. 26%), not more acceptable. The vowel-deleted forms are most acceptable when the resulting cluster ends in an obstruent followed by a sonorant, i.e. (C)TR, with acceptability at 27%, vs. 24% for other sonority profiles.

The monosyllabicity effect did not mirror the ratings of the nominative forms either: monosyllabic vowel-deleted genitives were uniformly less acceptable than disyllabic ones
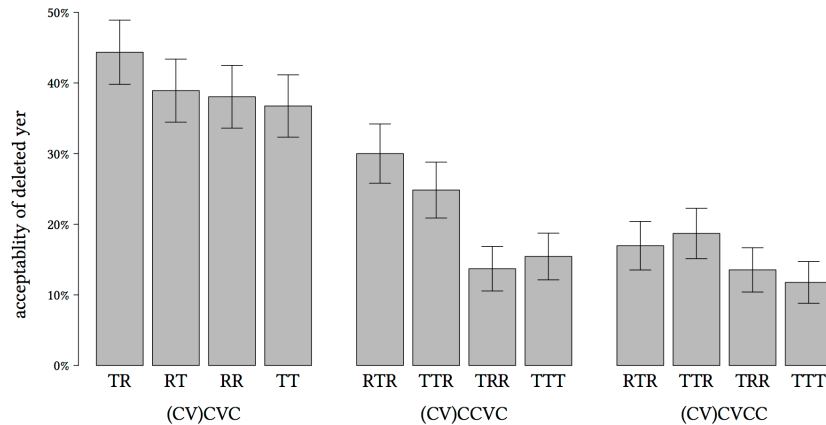
Figure 4: Acceptability of the unfaithful, vowel-deleted genitive (115 participants). The vowel-deleted genitives are more acceptable in TR-final forms (TR, RTR, TTR) and when the base has a simple coda.

(22% vs. 28%), e.g. [pișox ∼ pișxa] ≻ [șox ∼ șxa]. This was true for cluster-free bases (43% vs. 36%), non-final-cluster bases (23% vs. 19%), and complex coda bases (18% vs. 13%).

The vowel-deleted genitives also show a difference between [e] and [o], with the deletion of [o] being more acceptable (27% vs. 24%), e.g., [xol ∼ xla] ≻ [xʲel ∼ xla].

3.4.4 Inferential statistics

The statistical analysis was performed with a mixed-effects logistic regression model, using the *lme4* package (Bates & Maechler 2009) in R (R Development Core Team 2012). The dependent variable was the acceptance of the vowel-deleted genitive. The following predictors were used:

- *nominative.response:* the rating of the nominative base on the 1–5 scale.
- *faithful.genitive:* a binary predictor that was true if the faithful genitive was accepted.
- *order:* a binary predictor that was true if the faithful genitive was presented to the participant before the vowel-deleted genitive.
- *tri.consonantal:* a binary predictor that was true if the cluster created by yer deletion was tri-consonantal, contrasting (CV)CCVC and (CV)CVCC bases with (CV)CVC bases.
- *complex.coda:* a binary predictor that was true if the nominative base had a complex coda, contrasting (CV)CVCC bases with (CV)CVC and (CV)CCVC bases.
- *disyllabicity:* a binary predictor that was true if the nominative or vowel-deleted genitive were disyllabic (both have the same number of syllables).
- *vowel:* a binary predictor that was true for the base vowel [o], and false for [e]

- *TR:* a binary predictor that was true for TR, RTR, and TTR consonant combinations.

All the variables were centered using R's *scale* command, to reduce correlations between them. Correlations were further reduced by residualization: *faithful.genitive* was residualized on *nominative.response*, *tri.consonantal* was residualized on *nominative.response* and *faithful.genitive*, *complex.coda* was residualized on *nominative.response*, *faithful.genitive*, and *tri.consonantal*, and *disyllabicity* was residualized on *nominative.response* and *faithful.genitive*. This brought all correlations near zero. It also meant that the grammatical effects of interest, viz. *tri.consonantal*, *complex.coda*, *disyllabicity*, *vowel*, and *TR*, are ones that go above and beyond the rating of the nominative base and the acceptability of the faithful genitive.[6]

To start, a model was fitted that had the following fixed effects: *tri.consonantal*, *complex.coda*, *disyllabicity*, *vowel*, *TR*, and all of their interactions except the interaction of *tri-consonantal* and *complex.coda*, and in addition to those, *nominative.response*, *faithful.genitive* and *order*. The model had random intercepts for *participant* and *consonant combination*. The model was then pared down, starting with the four-way interactions, each time making sure that the model is not significantly changed with an ANOVA model comparison test. All the interactions were removed this way, except for the two-way interaction of *TR* and *tri.consonantal*.

Next, random slopes were added to the model, starting with a fully crossed model that had all predictors as random slopes for *participant*, and all the between-item predictors as random slopes for *consonant combination*. This model did not converge. To reach a model with maximal random slopes that nonetheless converges, we followed Barr et al. (2013), removing the random slope term with the smallest variance and refitting the model. This was repeated until a model that converges was reached. This model, reported in Table 6, has *nominative.response*, *faithful.genitive*, *tri.consonantal*, and *disyllabicity* as random slopes for *participant* and no random slopes for *consonant combination*. This final model enjoys low collinearity measures ($\kappa =$ 1.91, VIF$\leq$1.42, calculated using *mer-utils*, by Austin Frank, available at `github.com/aufrank/R-hacks`).

The model in Table 6 shows that yer deletion is significantly more acceptable if the nominative base is rated high, and significantly less acceptable if the faithful genitive was accepted, and if the faithful genitive was shown first. More importantly, the model also confirms that yer deletion is significantly less acceptable when it creates tri-consonantal clusters, and when the cluster originates from a complex coda. Deletion is significantly more acceptable in disyllables, when the stem vowel is [o], and when the cluster ends

---

[6]The minimization of collinearity via predictor centering and residualization has been criticized by Wurm & Fisicaro (2014). When we ran the regression without centering and residualization, the result was very similar to the model we report in Table 6, except that *vowel* did not reach significance. In particular, the source-oriented reduced acceptance of deletion in the presence of a complex coda and the product-oriented preference for deletion when it creates a (C)TR cluster were both highly significant.

	β	SE(β)	z	$p\,(>|z|)$
(Intercept)	−1.71	.15	−11.10	
nominative.response	.49	.09	5.62	<.0001
faithful.genitive	−1.01	.11	−9.12	<.0001
order	−.30	.04	−7.30	<.0001
tri.consonantal	−.65	.06	−10.16	<.0001
complex.coda	−.24	.05	−4.90	<.0001
disyllabicity	.23	.05	4.79	<.0001
vowel	.14	.04	3.45	<.001
TR	.35	.05	6.70	<.0001
TR:tri.consonantal	.12	.08	2.45	<.05

Table 6: mixed effects logistic regression model for the experiment (115 participants)

in an obstruent followed by a sonorant. The interaction term shows that the preference for (C)TR clusters is stronger in tri-consonantal clusters. Since the correlations in the model were reduced to a minimum (including centering and residualizing as explained above), each effect can be interpreted individually, e.g., the presence of complex codas significantly reduces the acceptability of yer deletion above and beyond the dispreference for tri-consonantal clusters and independently of the effect that these complex codas have on the rating of the nominative base.

3.5   Norming study: phonotactics & spelling

Our main study left two potential questions for the interpretation of the results. Firstly, one wonders whether the judgments of the nominative bases are influenced by the presence of the genitives, e.g., perhaps participants could be adjusting their rating of the nominative in reference to their expectation about the goodness of the vowel-deleted genitive. Secondly, the materials were presented to the participants in Russian spelling, using Cyrillic letters, which gives participants a little leeway in the phonological interpretation. In particular, the letter ⟨e⟩ normally but not obligatorily causes the previous consonant to be read palatalized, and the vowel letter itself can also be read as [ʲo]. Additionally, the disyllabic items were not marked for stress, thus requiring participants to make a decision about stress location. To address these ambiguities in interpretation, a norming study was conducted where a group of participants was asked to read aloud and rate the nominative bases.

Participants: 15 native speakers of Russian, 11 women and 4 men, ranging in age from 20 to 61 (mean 33); all living in Moscow. None reported any speech or hearing problems. All had some familiarity with a major Indo-European language such as English, German, French, or Spanish. The participants were paid the equivalent of $15 for their time.
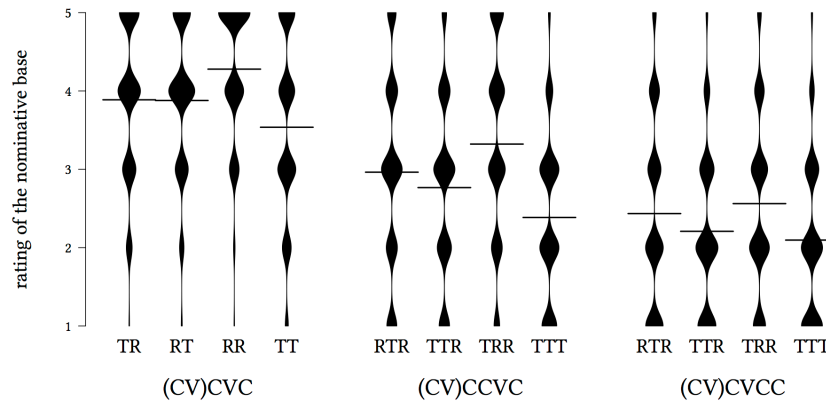
Figure 5: Norming study (15 participants): ratings of bases in a word likelihood task

Materials and procedure: The list of nominative bases from the main study was used. The participant sat in a quiet room with a Russian-speaking experimenter, who asked them to read aloud each word that appeared on a computer screen, and to rate the word on a scale of 1 (worst) to 5 (best), based on how plausible it was as a word of Russian. The word appeared on the computer screen in Cyrillic orthography, in isolation (no frame or context). The participant read the word and then gave it a rating, which the experimenter typed in. Each word was recorded as an audio file for further annotation. Each person rated at least 150 words, chosen at random from the list of words used in the main study. The participants rated and pronounced a total of 1783 different words, which constitutes 48% of the items in the main study.

For further analysis, a native speaker of Russian annotated each audio file in Praat (Boersma & Weenink 2009) for three properties: the vowel pronounced in the last syllable ([e] or [o]), the presence of palatalization on the consonant preceding the vowel, and the location of stress (initial or final). These materials are available for download at `http://becker.phonologist.org/projects/yers/`.

Results: Figure 5 shows the ratings by prosodic shape and sonority profile. These results are nearly identical to those shown in Figure 2, with highest ratings for (T)RR items. The ratings also followed those of the main study in terms of monosyllabicity and vowel quality. We conclude that the ratings of the nominative bases in the experiment were not influenced by the judgments of the genitive forms.

The disyllabic items were pronounced with final stress in 85% of all tokens. Initial stress appeared most often on CVCVC items (23%), less often on CVCCVC items (15%), and least often on CVCVCC items (8%).

The vowel letter ⟨o⟩ was pronounced as [o] in 99.7% of all tokens. The preceding consonant was pronounced non-palatalized (velarized) in 99.9% of all tokens, except before the inherently palatalized consonant [tɕʲ].

As for palatalization and vowel quality of ⟨e⟩, the expectation is for [ʧʲ] to always be palatalized and for [ts ʂ ʐ] to never be palatalized. The other consonants are expected to be palatalized before [e]. The vowel letter ⟨e⟩ was pronounced [e] in 98.7% of all tokens and [ʲo] in the remaining 1.3%. With the pronunciation [e], the preceding consonant was pronounced as expected before [ʧʲ ʂ ʐ ts]. The remaining consonants were palatalized in 91% of all tokens.

We infer from this that our internet participants read the nonce words with the pronunciation we were hoping for: the letter ⟨e⟩ is pronounced [e] with a preceding palatalized consonant over 91% of the time. The letter ⟨o⟩ is pronounced [o] with a preceding non-palatalized consonant 99% of the time. Stress is assigned to the final syllable in 85% of all disyllabic tokens.

3.6   Discussion

Russian speakers accept yer deletion in nonce words, relying on both source-oriented and product-oriented generalizations. The source-oriented effect of a complex coda in the base significantly reduced the acceptability of deletion. The product-oriented preference for (C)TR clusters made deletion significantly more acceptable, while (C)TR consonant combinations were not rated higher than others in the nominative source.

Two additional generalizations had a significant effect on yer deletion: monosyllabicity and the quality of the yer. Deletion was significantly less acceptable in monosyllabic products, an effect that was consistent across consonantal profiles in the vowel-deleted genitive, but not in the nominative source. Speakers found [o] to be significantly more deletable than [e], which we suggest is due to the loss of palatalization with [e], e.g., [xʲel ∼ xla] vs. the more faithful [xol ∼ xla]. We return to both of these generalizations in §4.3.

We replicated some of the findings of Gouskova & Becker (2013) on a larger scale, namely the source-oriented complex coda effect, and the monosyllabicity effect. The preference for (C)TR clusters and for the deletion of [o] are new findings. The current experiment further answers some questions that were unresolved in Gouskova & Becker's (2013) studies. Those experiments asked participants to rate a suffixed/genitive word as a form of the nominative on a scale of 1–7, but the status of the nominative base was not assessed independently, so it was not clear whether people rated a paradigm poorly because they didn't like deletion in that context or because they didn't like the base in the first place. In the current study, the rating of the base is separate from the judgment of deletion, which clarifies that the complex coda effect is source-oriented. Furthermore, the present study establishes that complex codas affect both monosyllables and disyllables, and that this effect cuts across all types of clusters. Gouskova & Becker (2013) only tested the complex coda effect in monosyllables, with a small number of clusters that differed between CVCC and CCVC forms. Our experiment tested the same consonant triplets

in (CV)CVCC and (CV)CCVC words, showing that the effect was robust in a variety of contexts. Finally, the current study asked participants to judge faithful genitives, further demonstrating the participants' ability to judge each part of the paradigm separately.

4   Analysis

The results from §3 show that Russian speakers extend generalizations about the distribution of yers from the lexicon onto nonce words. We claimed that at least one of these generalizations is source-oriented (deletion dispreferred if its source ends in a complex coda), and that at least one of these generalizations is product-oriented (deletion preferred if the produced cluster creates a (C)TR cluster).

We start with §4.1, which presents a theoretical model that uses one product-oriented grammar for the yer words, another product-oriented grammar for the non-yer words, and a grammar inference mechanism that expresses the source-oriented generalization. Section §4.2 reports the results of a pair of learning simulations, on the nominative source and the genitive product, that allow the diagnosis of generalizations as either source-oriented or product-oriented. Section 4.3 discusses the role of faithfulness in the model. Section 4.4 integrates the results from §4.2 and §4.3 in a single implementation of the theory from §4.1. The results are compared to the source-oriented Minimal Generalization Learner in §4.5, and §4.6 concludes.

4.1   A model of grammar inference

We propose a model of morphophonological learning that identifies paradigmatic operations, and then generalizes over the subset of the lexicon that shares an operation. In the case of Russian, "formation of the genitive by addition of [a]" would be the operation that identifies non-yer masculines. A different operation, viz. "formation of the genitive by addition of [a] and deletion of the stem-final vowel", identifies the yer masculines. Feminines and neuters will form other sublexicons. The identification of an operation triggers the formation of a sublexicon, which in turn triggers learning of two grammars for that sublexicon: a *gatekeeper grammar*, which is a purely phonotactic grammar for the sources to the sublexicon, and a *grammar proper*, which is a fuller grammar for the products of the sublexicon. The gatekeeper assigns a high probability to the words that are likely to be part of the sublexicon. The effect of the gatekeeper grammar is driven entirely by markedness constraints; faithfulness constraints are always vacuously satisfied in it. This vacuous satisfaction follows from assuming the base as the underlying representation (Hayes 2004; see also Albright 2008; Becker 2009; Hayes 1995, 1999). The grammar proper assigns a probability to the products of the operation that defines the sublexicon. It uses both markedness constraints, for example those that prefer TR clusters, and faithfulness

constraints, such as those that penalize vowel deletion.

Under this view, a Russian speaker holds two sublexicons and four grammars for their masculine nouns. For the yer words, this would include the yer sublexicon and two grammars: the yer gatekeeper grammar GK_{yer}, and the yer grammar proper GP_{yer}. For the non-yer words, there is the the non-yer sublexicon, the gatekeeper non-yer grammar $GK_{non\text{-}yer}$, and the non-yer grammar proper $GP_{non\text{-}yer}$. All of these grammars are taken to be constraint-based. A grammar (GP or GK) can also be seen as a function from words (or, more precisely, input-output mappings) to probabilities.

When Russian speakers are given a nonce nominative base "NOM" and a nonce genitive "GEN" to go with it, and are asked whether "GEN" is acceptable, they calculate the probability of the paradigm [NOM $\sim$ GEN], which is the sum of the probabilities assigned to the paradigm by the yer grammars and the non-yer grammars, as shown in (1).[7] The probability assigned to the genitive by the yer grammar proper ($p(GEN|GP_{yer})$) is modulated by the willingness of the yer gatekeeper grammar to allow the base in ($p(GK_{yer}|NOM)$), and similarly for the non-yer grammar.

(1) $p(NOM \sim GEN) = p(GK_{yer}|NOM) \cdot p(GEN|GP_{yer}) + p(GK_{non\text{-}yer}|NOM) \cdot p(GEN|GP_{non\text{-}yer})$

Focusing on the vowel-deleted genitive, the one generated by the yer grammar proper, we can simplify the expression in (1) to (2); see Appendix A for details. The advantage of (2) is that we can estimate each of its components. The probabilities in the numerator of (2) are those assigned to the nominative by the yer gatekeeper and to the vowel-deleted genitive by the yer grammar proper (as estimated by the models in §4.2 and §4.4). The denominator is the overall acceptability of the nominative, which was given by the participants on a 1–5 scale (easily transformed to probabilities by subtracting them from 6 and taking the inverse). Combined, this predicted probability is expected to be proportional to the participants' yes/no judgements of the vowel-deleted genitives.

(2) $p(NOM \sim GEN_{yer}) \mathrel{\overset{\propto}{\approx}} \dfrac{p(NOM|GK_{yer}) \cdot p(GEN_{yer}|GP_{yer})}{p(NOM)}$

The model we present here, illustrated with an example of Russian vowel deletion, is meant to be general: the model computes the probability of a paradigm in a language as a function

[7]In the experiment reported in §3, speakers were explicitly asked about the goodness of the nominative as well as the genitive, but the theory advanced here is not specific to this task. Thus, Gouskova & Becker's (2013) experiment asked participants only about the relationship between nominatives and genitives, but the results were similar, with dispreference for nominatives with a complex coda. We assume that people use the gatekeeper grammar whenever they assess paradigmatic relationships, not just when they are asked to judge the nominative. A similar result is predicted with a task involving the feminine, where the yer appears in the genitive plural. Given a genitive plural that ends in a complex coda, e.g., [poʂm], it is expected that speakers will accept the nominative singular [poʂma], but not the nominative singular [pʂma]. Indeed, the facts of Russian align with this prediction: historic yer feminines such as [mestʲ] 'revenge' and [lestʲ] 'flattery' no longer alternate in inflectional paradigms (Gouskova 2012), retaining the vowel throughout.

of its probability in each of the language's sublexicons. Each sublexicon has a gatekeeper grammar, based on the sources to the sublexicon; the speaker uses an inference mechanism to assess the ability of a novel source to pass through a given sublexicon. Each sublexicon also has a grammar proper, computed over the products of the sublexicon. Faithfulness constraints are vacuously satisfied in the gatekeeper grammar, and are thus limited in their influence to the grammar proper.

4.2   Diagnosing generalizations

To test whether the generalizations that we identified about the distribution of yers can be plausibly found in the nominative sources or in the genitive products, or both, we used the *UCLA phonotactic learner* (henceforth UCLAPL, Hayes & Wilson 2008), which we trained on sources and products separately. This learner accepts a list of words as its training data and another list of words for testing. It uses the training data to generate markedness constraints, and once the constraints are induced, they are used to predict the probability of each word in the testing data.

The UCLAPL was trained on the 1,902 masculine yer words (as described in §2), and since this learner expects at least 3,000 words for training, the list was duplicated. The model also takes a list of the sounds of the language with their feature specifications. The sounds from Table 4 were used with the features of Hayes (2009). The model was tested on the 5,516 target nonce words that were presented to the participants in the experiment. The model's default parameters were used (30 constraints, no projected tiers). The training materials and the results for all of the simulations described here are available at `http://becker.phonologist.org/projects/yers/`.

The UCLAPL produces a list of the generated markedness constraints with the weights that were assigned to them. In addition, it produces the result of applying these constraints to the testing data with the assigned violations and the overall harmony generated. These harmonies are the natural logarithms of the probabilities assigned to the items, with the highest harmony being zero (=100%). The UCLAPL's detailed output enables the assessment of the model both quantitatively and qualitatively: the probabilities can be used in a regression model to predict the participants' responses (yes/no acceptance of the unfaithful genitive), while the violation marks can identify the markedness constraints that account for the generalizations that the learner projected from the training data.

The learner was first trained on yer-word nominatives (e.g., [rot] 'mouth', [kastʲor] 'fire', [dʲenʲ] 'day' etc.) and tested on nonce nominatives (e.g., [xʲel], [ʂom], etc). The results of this simulation are shown in Figure 6a, grouped by prosodic shape and sonority profile. The highest probabilities were assigned to the cluster-free nonce bases, followed closely by bases with medial clusters, with complex coda bases ranking lowest, matching the participants' preferences. The low probability of the words with complex codas is due
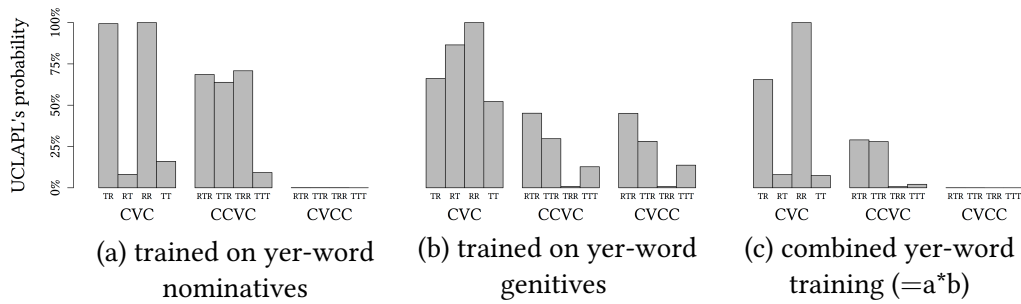
Figure 6: UCLAPL's predictions for nonce words after training on yer lexical words, e.g., [rot ∼ rta] 'mouth'.
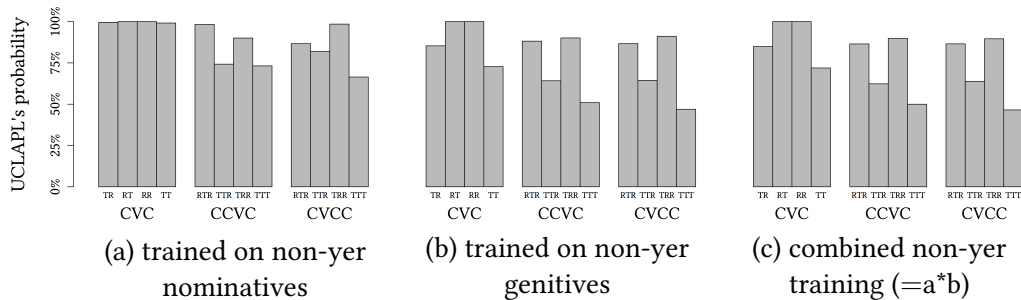

Figure 7: UCLAPL's predictions for nonce words after training on non-yer lexical words, e.g., [nos ∼ nosa] 'nose'.

to a pair of constraints that were induced by the learner: *CX# (=no consonant followed by another segment at the end of the word) and *C# (=no consonant at the end of the word), i.e. the learner identified that yer nominatives do not end in two consonants, and applied this generalization to the nonce words. A second simulation used yer-word genitives (i.e., unfaithful, vowel-deleted genitives, e.g., [rta] 'mouth', [kastra] 'fire', [dnʲa] 'day' etc.) and tested on nonce unfaithful genitives (e.g., [xla], [ʂma], etc), with results shown in Figure 6b. Higher probabilities were assigned to the nonce (CV)CC-V genitives, and lower probabilities to triconsonantal clusters. Naturally, without access to the base, the predicted probabilities of CCVC and CVCC bases are identical, and the complex coda effect could not be found. To summarize, then, the phonotactic learner notices the lack of complex codas in the nominative source, and projects this as a source-oriented generalization onto the training data.

As for the sonority profile, recall that participants found yer deletion to be most acceptable in CTR bases (e.g., [tunzʲem], [tupʂor]), and least acceptable in TRR and TTT bases (e.g., [lʲipron], [lufxotʃʲ]). This preference for CTR clusters is missing from the source oriented learner in Figure 6a, which correctly assigns low probabilities to TTT bases, but overestimates the goodness of TRR bases. The product-oriented learner in

21

Figure 6b correctly assigns high probabilities to CTR, and low probabilities to TRR and TTT, matching the participants' responses. The UCLAPL matched the preference for CTR clusters only under product-oriented training, confirming the expectations developed in §2. The learner's performance on CC clusters is a little disappointing in both cases: the participants accepted deletion in TR bases most often, and found RT, RR, and TT clusters all equally good. The phonotactic learner overestimates the goodness of RR items both in the sources and in the products.

The two training regimens, then, allowed the learner to discover two generalizations: one about the badness of complex coda sources, and one about the goodness of CTR products. But each training set only discovered one generalization. The two grammars that the UCLAPL learned can be combined by multiplication (or equivalently, addition of the harmony scores followed by exponentiation), e.g., by multiplying the probability that the nominative-trained grammar assigned to [ʂom] by the probability the genitive-trained grammar assigned to [ʂma], and similarly for each nonce item. The product of the predictions of the two grammars, shown in Figure 6c, is also an estimate for the numerator of (2), and thus approximates the predicted goodness of a vowel-deletion paradigm such as [ʂom ∼ ʂma]. The combined grammar models both the badness of complex codas and the goodness of CTR clusters, expressing the generalizations of its components.

The generalization about the goodness of cluster-free items was discovered by both grammars, and is thus compatible with either source-oriented or with product-oriented explanations, or both. Similarly, the preference for polysyllables over monosyllables was discovered by both grammars, and is thus compatible with either explanation. The preference for deletion of [o] over [e] wasn't discovered by either grammar; we return to this point in §4.3 below.

To assess the differences between the yer sublexicon and the non-yer sublexicon, and to assess the importance of the training data, the phonotactic learner was run with the sources and products of the the non-yer words. In these two simulations, the learning data consisted of the list of 18,661 non-yer words (again duplicated, to match the procedure with the yer words). The results are shown in Figure 7, again with the nominatives, the genitives, and the combined non-yer grammar that is the product of the first two. Figure 7a shows the results of training on non-yer nominatives (e.g., [nos] 'nose', [most] 'bridge', [gusʲ] 'goose', etc.) and testing on the nonce nominatives as above (e.g., [xʲel], [ʂom], etc). Figure 7b shows the results of training on non-yer genitives (e.g., [nosa] 'nose', [mosta] 'bridge', [gusʲa] 'goose', etc.) and testing on the nonce unfaithful genitives as above (e.g., [xla], [ʂma], etc). Figure 7c shows the result of multiplying the predictions of the two grammars. Training on non-yer nominatives led to a much worse match with the participants' responses, showing neither the lower acceptability of complex codas nor the preference for CTR clusters. The preference for CTR clusters was not discovered by

training	χ^2	df	p-value
yer words, nominative sources	32.7	1	$<.0001$
yer words, genitive products	126.4	1	$<.0001$
non-yer words, nominative sources	13.6	1	$<.001$
non-yer words, genitive products	2.6	1	$>.1$

Table 7: The contribution of each training regimen to a superset model with the participants' responses as a dependent variable

training on the genitives, either. Following a reviewer's suggestion, we also trained the model on all masculines (both non-yer and yer); the results are virtually identical to those seen in Figure 7. This demonstrates that generalizations about words that undergo yer deletion are distinct from generalizations about Russian words in general, and cannot be discovered using a phonotactic grammar for the language as a whole (cf. Yearley 1995; Szpyra 1992).

To quantify how well the results of the four different simulations matched the participants' acceptance of vowel deletion, nested model comparison was used. We used the *glm* function in R to fit a superset logistic model that had the participants' acceptance of the vowel-deleted genitive as the dependent variable. The four predictors were the probabilities that each of the four runs of the UCLAPL assigned to the same items (nominative and genitive yer words, nominative and genitive non-yer words). Then, four subset models were fitted, each one having three of the four predictors. Four ANOVA model comparisons were used to assess the contribution of each missing predictor to the superset model. The use of model comparison alleviates any concerns about the large degree of collinearity between the four predictors. The result, in Table 7, shows that the contribution from training on yer words, as measured by the χ^2 statistic, outweighs the contribution of training on the non-yer words. The two yer word trainings both make sizable contributions, which is to be expected: each one captures one generalization that the other does not.

To summarize, we were able to model the participants' judgments of yer deletion in nonce words by using a combination of two phonotactic grammars, one trained on lexical yer words in the nominative source, and one trained on yer words in the genitive product. The resulting phonotactic grammars correctly discovered the source-oriented dispreference for complex coda and the product-oriented preference for the (C)TR sonority profile.

We used the UCLAPL to induce markedness constraints for each sublexicon separately. These constraints were used to show that the generalizations we identified can be found by an inductive process either in the source or in the product, or both. Yet classical constraint-

based theory assumes one set of constraints that is possibly universal or innate (Prince & Smolensky 1993/2004). The grammar inference model we developed in §4.1 is compatible with this view; in §4.4 we show that the model can be trivially adjusted to have a single set of markedness constraints per language that is ranked or weighted differently in each sublexicon. This constraint set may also include universal and/or innate constraints.

4.3 The role of faithfulness

The discussion up until now focused on two generalizations: the source-oriented dispreference for complex codas in yer words, and the product-oriented preference for (C)TR clusters. Yet there were two more significant generalizations in the experimental results: the preference for deletion in disyllabic words and the preference for deletion of [o] over [e].

The preference for yer deletion in disyllabic words seems to be compatible with a range of explanations, either source-oriented or product-oriented. The UCLAPL predicts the disyllabicity preference in both the source and product grammars using markedness constraints that penalize word-initial clusters (e.g., s̱xol ∼ s̱xla) more strongly than intervocalic clusters (e.g., pis̱xol ∼ pis̱xla). The effect does not have to be phonotactic, however. It could be due to the faithfulness component of the grammar proper. The deletion in a monosyllable, as in [s̱xol ∼ s̱xla], affects the initial (and only) syllable of the base, violating the special faithfulness that protects initial syllables (Beckman 1997, 1998; Becker et al. 2011, 2012; Gouskova & Becker 2013).

As for the vowel effect, where participants find [o] more deletable than [e], it cannot be a product-oriented effect, since neither vowel is present in the vowel-deleted product. Yet it is not likely to be a source-oriented phonotactic restriction, either. Among the yer words in the lexicon, [e] is somewhat more frequent than [o] (55% vs. 45%), and thus the nominative-trained UCLAPL finds [e] more deletable than [o], unlike the participants. The markedness-based UCLAPL cannot find the correct generalization, we argue, because the generalization is based on faithfulness, and has to do with the preservation of palatalization.

Before the Cyrillic vowel letter ⟨e⟩, consonants are usually palatalized. Thus, a paradigm such as ⟨хел ∼ хла⟩ is interpreted as [xʲel ∼ xla],[8] with a change from a palatalized consonant in the nominative to a non-palatalized (velarized) consonant in the genitive. With the vowel letter ⟨o⟩, no palatalization alternations are possible, and consonants are non-palatalized (velarized) in both members of the paradigm. In a full constraint-based theory with markedness and faithfulness, markedness enforces the alternation, demanding that [xla] be preferred to *[xʲla], while the faithfulness constraint

---

[8]As established in §3.5, Russian speakers choose the pronunciation [e] with preceding palatalization well over 90% of the time. Another possible reading is [xʲol ∼ xla], with an identical palatalization alternation.

Ident(back) penalizes the discrepancy between [x] and [xʲ]. Since Ident(back) only assigns violations when the palatalization-inducing ⟨e⟩ is deleted, it can account for the observed trend in the participants' responses.

Support for this view comes from the treatment of the consonant [lʲ], which retains its palatalization in real yer words (e.g., [lʲev ~ lʲva], *[lva] 'lion'), but not in the experiment (e.g., [lʲeş ~ lşa]), making the nonce paradigms less Russian-like in this case.[9] Indeed, participants accepted the deletion of ⟨e⟩ before [lʲ] much less frequently than before ⟨o⟩ (14% vs. 25% compared to 24% vs. 28% elsewhere), lending support to the faithfulness-based view.

The UCLAPL simulations presented in §4.2 had no access to faithfulness, and thus could not express this aspect of the results. In §4.4 below, we present a model that builds on the UCLAPL's results with the addition of faithfulness constraints. This model allows us to show that the incorporation of Ident(back) makes a significant improvement to the predictive ability of the model.

In the literature on lexical subpatterns in phonology, it has been claimed that subgrammars may differ only in their ranking/weighting of faithfulness (Benua 1997; Fukazawa 1999; Ito & Mester 1999). More recent work shows that both faithfulness and markedness need to differ in ranking/weighting between subgrammars (Pater 2000, 2006; Gouskova 2007, 2012; Flack 2007; Jurgec 2010, and others), as we propose here.

4.4 Implementation

In §4.2 above, the UCLAPL was used to induce and weight markedness constraints that simulate gatekeeper grammars and also the markedness components of grammars proper; the division of the lexicon into sublexicons was done manually. In this section, we present the results from a complementary learning model that automates the division of the lexicon into sublexicons, creates gatekeeper grammars and full grammars proper that include faithfulness constraints, and generates genitive products from nonce nominative sources.

This learner does not induce its own constraints; rather, a set of constraints is given by the analyst and these constraints are used in all of the grammars. A full description of this implementation goes beyond the scope of this paper; this task is taken up in Allen & Becker (2014). Additionally, an implementation similar to ours is offered in Moore-Cantwell & Staubs (2014).

As its training data, the learner was given the 1902 yers paradigms (nominative-

---

[9]The loss of palatalization in lexical yer deletion depends on the morphological context and the features of the palatalized consonant (Farina 1991; Kochetov 2006; Padgett to appear). Our nonce words were constructed to look morphologically simple, and in morphologically simple cases, only [lʲ] consistently retains its palatalization (cf. [kazʲol ~ kazl-a] 'goat', [dʲenʲ] ~ [dnʲa] 'day' but [ulʲej ~ ulʲj-a] 'beehive', [lʲon ~ lʲna] 'linen').
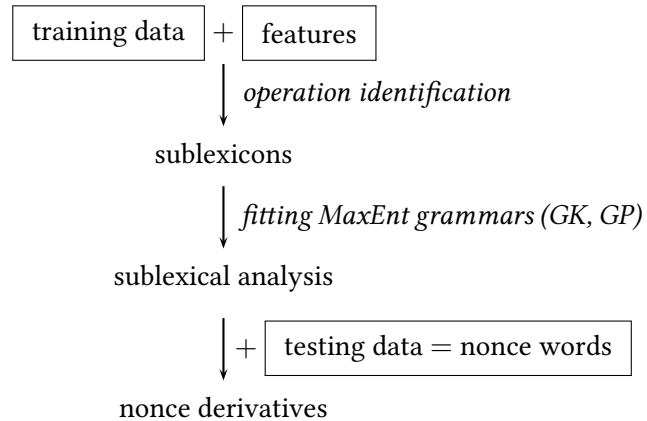
Figure 8: Summary of the Sublexical Learner. The training data, testing data, feature chart, and constraint set are supplied by the analyst.

genitive pairs) described in §2 and a random selection of 600 non-yer paradigms. As testing data, the model received the same 5,516 nominatives that the participants were given, for which it then generated genitive forms with a probability distribution over them. The model goes through the steps summarized in Figure 8. The probabilities of the vowel-deleted genitives, aggregated by sonority profile, syllable structure, and monosyllabicity, are shown in Figure 9. The simulation is available at the learner's website, `http://sublexical.phonologist.org/`.

Figure 9 shows the results of training the model with the constraints that were induced by the UCLAPL, obtained by collecting all the constraints that the UCLA found in all of its runs. The impressive correlation in Figure 9 is surely due to the quality of the UCLAPL's constraints, which does a lot of the linguistic heavy lifting here. The simulation was helped by adding the faithfulness constraints IDENT(back) and IDENT(back)/lateral. An ANOVA model comparison confirms that adding faithfulness makes a small but significant improvement to the predictive power of the model compared to a simulation with markedness constraints only ($\chi^2(1)$ = 10.4, $p < .005$).

There are two rather minor differences between the theoretical model we describe in §4.1 and the implemented learner we describe here. Firstly, our implementation weights its constraints by comparing each sublexicon to the other sublexicons of the language, while the model in §4.1 relies on the UCLAPL, which weights its constraints by comparing each sublexicon to a simulated rich base (a "salad" in Hayes & Wilson 2008's terms). For discussion of this point, see Hayes (2014). Secondly, the implementation does not incorporate the language-wide probability of the base, i.e., it only uses the numerator of (2) from §4.1.
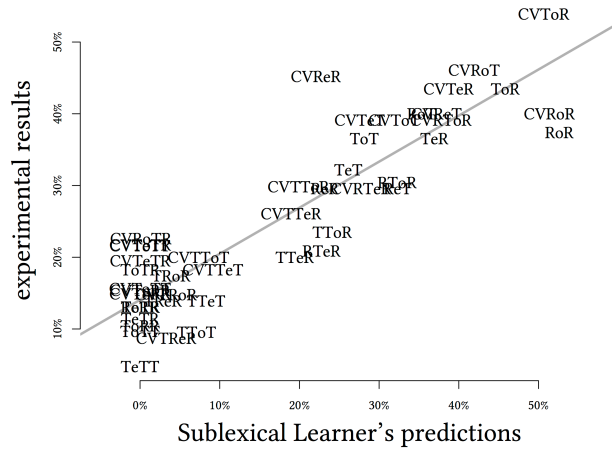
Figure 9: Sublexical Learner's predicted acceptability of yer deletion, by sonority profile, monosyllabicity, and base vowel, using the UCLAPL's constraints.

4.5   Comparison with a rule-based model

Another perspective on the question of source-oriented generalizations comes from the Minimal Generalization Learner (MGL, Albright & Hayes 2002, 2003, 2006). As Albright & Hayes note, the MGL uses SPE rules, which describe source-oriented generalizations. It starts by creating a word-specific rule for each lexical item, and then aggregates rules that share the same change (e.g., deletion of [o]), creating more general rules by pairwise comparison. Each rule is assigned a confidence score that is based on the number of words it derives correctly and the number of words it derives incorrectly (e.g., incorrectly deleting a vowel from a non-yer word).

When we trained the MGL on the Russian lexicon,[10] it generated vowel deletion rules that always specified a mid vowel followed by a single consonant, followed by a word boundary. Thus, it successfully captured the source-oriented generalization that yers are deleted when followed by exactly one consonant. When the MGL was tested on the experimental nonce forms, it applied deletion rules to bases that end in a single consonant, but not to bases with complex codas, mimicking the participants' reluctance to accept deletion in these cases. As for the product-oriented preference for (C)TR clusters, however, the MGL performed badly, wrongly assigning the highest confidence to deletion that created (T)TT clusters, with lower confidence for RT, and lower still for (C)TR clusters. The MGL also found [e] to be more deletable than [o], like the nominative-trained UCLAPL.

A measurement of the MGL's performance is given in Table 8, compared to the

---

[10]To accommodate the MGL's computational capacity, the training data had to be simplified in a few ways: (1) all 1902 yer words were included, but only a random sample of 3000 non-yer words was used; (2) the genitive suffix was removed, so that the environment of deletion rules would include word-final consonants; (3) all palatalization alternations were removed; (4) place features were removed.

training	AIC	BIC
UCLAPL, nominative + genitive	6019	6039
UCLAPL, genitive products	6055	6068
UCLAPL, nominative sources	6154	6168
Minimal Generalization Learner	6197	6210

Table 8: Comparison of the MGL and the UCLAPL (lower is better).

UCLAPL's. In each model, the dependent variable is the participants' acceptance of the vowel-deleted genitive. The models were compared on their AIC scores (Akaike Information Criterion) and BIC scores (Bayesian Information Criterion), where lower scores represent a better fit. The result shows that the predictive power of the UCLAPL outweighs that of the MGL.

4.6 Summary

The approach developed in this section is based on the partitioning of lexicons into sublexicons according to the morphophonological changes they undergo. Each sublexicon gives rise to two constraint-based grammars: a purely phonotactic gatekeeper grammar where only markedness constraints are active, and a grammar proper that includes active markedness and faithfulness. When given a novel word, the speaker has to infer which grammar to send it to, and this inference mechanism allows the speaker to extend source-oriented generalizations from the real words of the language to novel words.

Our sublexical analysis for the masculine second declension nouns is fully compatible with the existence of other (phonotactic) grammars in the language. In fact, we assume that language-wide phonotactic learning is the first step in phonological acquisition, an assumption that is argued for extensively in, e.g., Mehler et al. (1988); Jusczyk et al. (1994, 1999), among many others. In addition to the sublexicons and the language-wide grammar, there very well may be additional grammars, e.g., for all nouns, for all nominative singulars, for all masculines, etc. These wider-scope grammars may account for speakers' use of changes in unexpected phonological environments, as noted by Bybee & Slobin (1982); Albright & Hayes (2003); Kapatsinski (2013), among others.

The UCLA phonotactic learner was doubly useful in the development of our approach: firstly, it helped us diagnose generalizations as either source-oriented or product-oriented by training it on sources and products separately. Secondly, we used the constraints it induced in our own implementation of the sublexical approach.

Another aspect of Russian yers that made the UCLAPL useful is their relative rigidity: the vast majority of words are either clearly yer words or clearly non-yer words, while very few words allow variation. This property of the data allowed a rather straightforward

partitioning of the lexicon. Other cases of lexically-specific behavior, however, often involve a large group of words that show variability, as in the case of the Hungarian dative (Hayes & Londe 2006; Hayes et al. 2009, see below). The implementation we presented in §4.4 allows the membership of lexical items in sublexicons to be gradient, and together with the UCLAPL's constraints, can cover a wider range of cases.

While the mechanism was developed for the Russian case at hand, we extend it to a variety of other cases of stem alternations and allomorph selection in Allen & Becker (2014). In the English past tense, for example, verbs of the *drink* ∼ *drank* type either end in a velar (*sneak* ∼ *snuck, dig* ∼ *dug*) or in a nasal (*swim* ∼ *swum, win* ∼ *won*), but ideally both — a velar nasal (Albright & Hayes 2003; Bybee & Slobin 1982; Bybee & Moder 1983; Bybee 2001). These preferences are captured with a phonotactic grammar that covers just these verbs. Extending the approach to allomorph selection also offers a solution to otherwise puzzling results, such as the tendency in the Hungarian dative to prefer the [nɛk] allomorph with bases that end in a complex coda, or a sibilant, or a coronal sonorant (Hayes et al. 2009; Hayes 2014). We discover these generalizations with a phonotactic grammar computed over the [nɛk]-taking sources. There are many known cases where allomorph selection is sensitive to the phonological shape of the base, yet there is no clear connection between the shape of the base and the shape of the allomorph (Bobaljik 2000; Paster 2006; Bye 2007; Embick 2010; Nevins 2011). One famous example comes from the Italian definite article: this prefixal article shows up as [i] before single consonants, while [ʎi] appears before clusters and vowels (Lepschy & Lepschy 1998). Even more surprising is the definite article allomorphy in Haitian Creole (Klein 2003; Bye 2007; Embick 2010), where vowel-final nouns take [-a], and consonant-final nouns take [-la], e.g., [liv-la] 'the book' vs. [bɔkɔ-a] 'the sorcerer'. While these distributions do not seem to optimize anything in the resulting products, and in particular not anything about syllable structure, the distribution is readily available to a phonotactic grammar computed over the sources. Gouskova & Newlin-Łukowicz (2014) demonstrate, however, that suppletive allomorphy is sensisitive both to the phonotactic generalizations over bases and to the generalizations over derived words: in their elicitation study of Russian diminutive suppletive allomorphy, sublexical generalizations over both types of words predicted people's choice of allomorphs. Using the sublexical approach, these distributions are learned in terms of a phonological organization of the lexicon that is based on phonotactic product-oriented grammars, even if the selection of allomorphs does not phonologically optimize the products.

5   Conclusions

This paper offered evidence for a productive source-oriented markedness-based generalization: yer deletion is blocked if the source of the derivation ends in a complex coda. The lexicon survey in §2 showed this to be true of the real words of Russian, and the nonce word task in §3 showed that speakers productively extend the generalization, accepting genitives with deleted vowels significantly more often when the nominatives end in a single consonant.

Dominant theories in linguistics use product-oriented generalizations, both in generative linguistics and in usage-based linguistics, and there is some evidence that speakers prefer product-oriented generalizations in artificial grammar tasks (Becker & Fainleib 2009; Kapatsinski 2009, 2011, see also Kapatsinski 2013). The model we propose uses product-oriented constraint-based grammars, together with a mechanism for partitioning the lexicon and learning multiple grammars. Source-oriented generalizations are encoded in an additional layer of grammar inference, expressed as a phonotactic gatekeeper for each lexical partition.

Some source-oriented generalizations can be captured in Optimality Theory with faithulness constraints (Kapatsinski 2011), but not all. The restriction on complex codas in the nominative is markedness-based, and cannot be expressed in terms of faithfulness. In fact, this generalization is not opaque, in the sense of Kiparsky (1968) et seq. (see McCarthy 2007 for a recent overview). Models of opacity within Optimality Theory, such as Optimality Theory with candidate chains (McCarthy 2007), are unable to express the complex coda restriction, because these models' approach to opacity hinges on faithfulness.

The idea of partitioning the lexicon and learning multiple grammars per language is well-established in linguistics, in particular the theory of co-phonologies (Anttila 2002; Inkelas et al. 1996; Inkelas & Zoll 2007, inter alia). In this theory, the phonological grammar consists of a single constraint hierarchy in which certain rankings are crucially under-determined. Individual morphemes are associated with fully specified constraint hierarchies, which may conflict with other morphemes' rankings. Thus, each morpheme or morpheme class has its own sub-grammar, which is partially consistent with the language as a whole. Multiple grammars are also used in Stratal Optimality Theory (Kiparsky 2000 et seq.), where each grammar corresponds to a level of affixation, and each successive grammar effects changes from input to output. In our approach, changes from input to output are done in the grammar proper only, while the gatekeeper only vacuously maps forms to themselves. An alternative to co-phonologies is constraint indexation, in which there is only one constraint hierarchy for the language, but it contains duplicate constraints that apply only to individual morphemes or morpheme classes (Ito & Mester 1999; Fukazawa 1999; Pater 2000, 2006, 2008; Kawahara et al. 2002; Flack 2007; Gouskova

2007; Becker 2009; Becker et al. 2011, and others). Compared to these frameworks, our approach differs in using two grammars in each sublexicon, with one grammar serving as a gatekeeper and another grammar serving to derive outputs.

Our approach is also related quite directly to Morpheme Structure Constraints (MSC, Chomsky & Halle 1968; see Booij 2011 for a recent overview). MSCs are phonotactic constraints on morphemes, usually thought to hold of underlying representations: for example, in English, monosyllabic lexical morphemes cannot end in lax vowels (e.g., *[bɛ] is not a possible root of English); in Japanese, Sino-Japanese roots cannot be bigger than a syllable (Ito & Mester 1995). MSCs share key features with our phonotactic gatekeeper grammar: they do not necessarily trigger alternations/unfaithful mappings, and they may hold of only subsets of morphemes, such as native vocabulary (excluding loanwords). The difference between MSCs and our proposal is that our constraints are constraints on surface forms, not on underlying representations (though see Sommerstein 1974 for an early proposal for surface MSCs). There is also a theoretical question about whether MCSs hold of roots only, or whether they cover morphologically complex forms such as those in our training data; in practice, this difference does not seem to matter.

Shaw (2006) offers an analogical model that assigns nonce words to sublexicons, focusing on the distinction in Japanese between native words and Sino-Japanese words. The model calculates the similarity of a nonce word to all the words in each sublexicon, and uses this to estimate the more likely sublexicon. In contrast to Shaw's (2006) analogical model, our implementation is much more flexible and requires less supervision. Analogical models rely on a linguist to align bases and derivatives in a way that allows relevant generalizations to be discovered by the model, and thus analogical modeling is less promising as a path towards a completely unsupervised morphophonological learner.

The computational implementation of the model was discussed briefly. Its main interest is in its unsupervised mechanism for identifying groups of words that undergo identical morphophonological operations (e.g., "add [a] to make a genitive"), making the analysis of Russian demonstrably learnable. The implementation uses the markedness constraints induced by the UCLAPL, to which it adds built-in faithfulness constraints. While far from complete, the implementation brings us closer to a fully expressive learner for the morphophonological component of human language.

Appendix A   Predicting paradigm wellformedness

In §4.1, we stipulated that the predicted wellformedness of a paradigm is the sum of the probabilities it is assigned by each sublexicon, where the probability of each sublexicon is the probability assigned by the grammar proper to the derivative, modulated by the likelihood that the base belongs to the sublexicon. The formula in (3) is an illustration with two sublexicons, called *yer* and *non-yer*. To help focus on the vowel-deleted genitives of interest, we label the derivative GEN_{yer}.

(3) $p(\text{NOM} \sim \text{GEN}_{yer}) = p(\text{GK}_{yer}|\text{NOM}) \cdot p(\text{GEN}|\text{GP}_{yer})$
$+ p(\text{GK}_{non\text{-}yer}|\text{NOM}) \cdot p(\text{GEN}|\text{GP}_{non\text{-}yer})$

The second summand in (3) is easy to calculate. The non-yer grammar proper $\text{GP}_{non\text{-}yer}$ does not allow deletion, i.e., the faithfulness component in $\text{GP}_{non\text{-}yer}$ penalizes $[\text{GEN}_{yer}]$ heavily, making the probability $p(\text{GEN}_{yer}|\text{GP}_{non\text{-}yer})$ very close to zero. This in turn makes the product $p(\text{GK}_{non\text{-}yer}|\text{NOM}) \cdot p(\text{GEN}_{yer}|\text{GP}_{non\text{-}yer})$ very close to zero.

(4) $p(\text{GEN}_{yer}|\text{GP}_{non\text{-}yer}) \approx 0$

(5) $p(\text{GK}_{non\text{-}yer}|\text{NOM}) \cdot p(\text{GEN}_{yer}|\text{GP}_{non\text{-}yer}) \approx 0$

The second summand is effectively eliminated, leaving (6):

(6) $p(\text{NOM} \sim \text{GEN}_{yer}) \approx p(\text{GK}_{yer}|\text{NOM}) \cdot p(\text{GEN}_{yer}|\text{GP}_{yer})$

Since $[\text{GEN}_{yer}]$ is quite likely as a yer genitive, $p(\text{GEN}_{yer}|\text{GP}_{yer})$ is rather high. An estimate for $p(\text{GEN}_{yer}|\text{GP}_{yer})$ is given by the genitive-trained UCLAPL (see Figure 6b).

We also need $p(\text{GK}_{yer}|\text{NOM})$, the probability that $[\text{NOM}]$ is a word in the yer sublexicon, for which Bayes' theorem can be used, as shown in (7). The probability we need is equal to the probability that the yer gatekeeper grammar assigns to $[\text{NOM}]$, multiplied by the probability of this grammar, and divided by the overall probability of $[\text{NOM}]$.

(7) $p(\text{GK}_{yer}|\text{NOM}) = \dfrac{p(\text{NOM}|\text{GK}_{yer}) \cdot p(\text{GK}_{yer})}{p(\text{NOM})}$

The denominator $p(\text{NOM})$ is the overall probability of $[\text{NOM}]$ in the language, i.e. the sum of probabilities assigned to it by each of the gatekeeper grammars the speaker has (8). Alternatively, the probability of the base could come from a language-wide probabilistic phonotactic grammar. We do not have estimates for the summands in (8), but we do have an estimate for the wellformedness of $[\text{NOM}]$: this is the participants' rating of the nominative base on the 1–5 scale.

(8) $p(\text{NOM}) = p(\text{NOM}|\text{GK}_{yer}) + p(\text{NOM}|\text{GK}_{non\text{-}yer})$

We also have an estimate for $p(\text{NOM}|\text{GK}_{yer})$: it is the probability that the nominative-trained UCLAPL assigns to the nominative sources. The probability of the yer gatekeeper grammar $p(\text{GK}_{yer})$ is a constant; it is proportional to the size of the yer lexicon, i.e. the proportion of yer words out of all of the words of the language ($\sim$10%). Similarly, $p(\text{GK}_{non\text{-}yer})$ is the proportion of non-yer words in the language ($\sim$90%). Together, $p(\text{GK}_{yer})$ and $p(\text{GK}_{non\text{-}yer})$ add up to one.

The probability that [NOM] is a yer word, then, is proportional to the probability that the yer grammar assigns to it, divided by the acceptability of [NOM], as in (9).

(9) $p(\text{GK}_{yer}|\text{NOM}) \propto \dfrac{p(\text{NOM}|\text{GK}_{yer})}{p(\text{NOM})}$

Putting together (6) and (9) yields (10). We now have an estimate of $p(\text{NOM} \sim \text{GEN}_{yer})$: it is approximately proportional to the probability that the nominative-trained UCLAPL assigns to [NOM], multiplied by the probability that the genitive-trained UCLAPL assigns to [GEN$_{yer}$], divided by the probability that the participants assigned to the nominative [NOM].

(10) $p(\text{NOM} \sim \text{GEN}_{yer}) \mathrel{\overset{\propto}{\approx}} \dfrac{p(\text{NOM}|\text{GK}_{yer}) \cdot p(\text{GEN}_{yer}|\text{GP}_{yer})}{p(\text{NOM})}$

There is one more technical point to cover. The reader may have noticed in (3) that a probability is given as a sum of two other probabilities, and therefore there is a worry that the sum would be greater than one. Yet these sums are guaranteed to be smaller than or equal to one if we accept (8), as seen in the following proof.

All probabilities are less than or equal to one, therefore $p(\text{GEN}|\text{GP}_{yer}) \leq 1$ and $p(\text{GEN}|\text{GP}_{non\text{-}yer}) \leq 1$. The equality in (3) can therefore be turned into the following inequality:

(11) $p(\text{NOM} \sim \text{GEN}) \leq p(\text{GK}_{yer}|\text{NOM}) + p(\text{GK}_{non\text{-}yer}|\text{NOM})$

Substituting given (7):

(12) $p(\text{NOM} \sim \text{GEN}) \leq \dfrac{p(\text{NOM}|\text{GK}_{yer}) \cdot p(\text{GK}_{yer})}{p(\text{NOM})} + \dfrac{p(\text{NOM}|\text{GK}_{non\text{-}yer}) \cdot p(\text{GK}_{non\text{-}yer})}{p(\text{NOM})}$

Taking (8) and dividing both sides by $p(\text{NOM})$:

(13) $\dfrac{p(\text{NOM}|\text{GK}_{yer})}{p(\text{NOM})} + \dfrac{p(\text{NOM}|\text{GK}_{non\text{-}yer})}{p(\text{NOM})} = 1$

Each summand in (13) is $\leq$1. Substituting:

(14) $p(\text{NOM} \sim \text{GEN}) \leq p(\text{GK}_{yer}) + p(\text{GK}_{non\text{-}yer})$

Since by hypothesis $p(GK_{yer}) + p(GK_{non\text{-}yer}) = 1$:

(15) $p(\text{NOM} \sim \text{GEN}) \leq 1$.

References

Albright, Adam (2008). A Restricted Model of UR Discovery: Evidence from Lakhota. Ms. MIT.

Albright, Adam & Bruce Hayes (2002). Modeling English past tense intuitions with minimal generalization. In Michael Maxwell (ed.) *Proceedings of the sixth meeting of the ACL special interest group in computational phonology*. Philadelphia: ACL, 58–69.

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

Albright, Adam & Bruce Hayes (2006). Modeling productivity with the gradual learning algorithm: The problem of accidentally exceptionless generalizations. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.) *Gradience in Grammar*, Oxford University Press. 185–204.

Allen, Blake & Michael Becker (2014). Learning alternations from surface forms with sublexical phonology. Talk given at Phonlogy 2014.

Anttila, Arto (2002). Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* **20**. 1–42.

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

Bates, Douglas & Martin Maechler (2009). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42.

Becker, Michael (2009). *Phonological trends in the lexicon: The role of constraints*. Ph.D. dissertation, University of Massachusetts, Amherst, MA.

Becker, Michael & Lena Fainleib (2009). The naturalness of product-oriented generalizations. ROA 1036.

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

Becker, Michael & Jonathan Levine (2012). Experigen - an online experiment platform. Available at https://github.com/tlozoot/experigen.

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

Beckman, Jill (1997). Positional faithfulness, positional neutralization, and Shona vowel harmony. *Phonology* **14**. 1–46.

Beckman, Jill (1998). *Positional faithfulness*. New York: Routledge.

Benua, Laura (1997). Affixes Classes are defined by Faithfulness. In Viola Miglio & Bruce Moren (eds.) *University of Maryland Working Papers in Linguistics 5. Selected Phonology*

Papers from Hopkins Optimality Theory Workshop 1997 / University of Maryland Mayfest 1997. 1–27.

Berent, Iris, Tracy Lennertz, Paul Smolensky & Vered Vaknin-Nusbaum (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology* **26**. 75–108.

Berent, Iris, Donca Steriade, Tracy Lennertz & Vered Vaknin (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* **104**. 591–630.

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

Blevins, Juliette (1995). The syllable in phonological theory. In John A. Goldsmith (ed.) *The Handbook of Phonological Theory*, Cambridge, MA, and Oxford, UK: Blackwell. 206–244.

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

Boersma, Paul & David Weenink (2009). Praat: Doing phonetics by computer, version 5.1.

Booij, Geert (2011). Morpheme structure constraints. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume & Keren Rice (eds.) *The Blackwell Companion to Phonology*, Wiley-Blackwell, vol. 4. xxx–xxx.

Bybee, Joan (2001). *Phonology and language use.* Cambridge University Press.

Bybee, Joan & Carol Lynn Moder (1983). Morphological classes as natural categories. *Language* . 251–270.

Bybee, Joan & Dan Slobin (1982). Rules and schemas in the development and use of the english past tense. *Language* **58**. 265–289.

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

Chomsky, Noam & Morris Halle (1968). *The Sound Pattern of English.* the MIT press.

Clements, George N. (1990). The role of the sonority cycle in core syllabification. In John Kingston & Mary Beckman (eds.) *Papers in Laboratory Phonology 1: Between the grammar and physics of speech*, New York: Cambridge University Press. 283–333.

Davidson, Lisa & Kevin Roon (2008). Durational correlates for differentiating consonant sequences in Russian. *Journal of the International Phonetic Association* **38**. 137–165.

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

Farina, Donna (1991). *Palatalization and Yers in Modern Russian Phonology: An Underspecification Approach.* Ph.D. dissertation, University of Illinois, Urbana, Urbana, IL.

Flack, Kathryn (2007). Templatic morphology and indexed markedness constraints. *Linguistic Inquiry* **38**. 749–758.

Fukazawa, Haruka (1999). *Theoretical implications of OCP effects on features in Optimality Theory.* Ph.D. dissertation, University of Maryland, College Park, MD.

Gouskova, Maria (2007). The reduplicative template in Tonkawa. *Phonology* **24**. 367–396.

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

Gouskova, Maria & Michael Becker (2013). Russian yer alternations are governed by the grammar. *Natural Language and Linguistic Theory* **31**. 735–765.

Gouskova, Maria & Luiza Newlin-Łukowicz (2014). Selectional restrictions as phonotactics over sublexicons. Ms. NYU.. (11 November 2014).

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.

Hayes, Bruce (1995). On what to teach the undergraduates: Some changing orthodoxies in phonological theory. In Ik-Hwan Lee (ed.) *Linguistics in the Morning Calm 3*, Seoul: Hanshin. 59–77.

Hayes, Bruce (1999). Phonological Restructuring in Yidiɲ and its Theoretical Consequences. In Ben Hermans & Marc van Oostendorp (eds.) *The derivational residue in phonology*, Amsterdam: Benjamins. 175–205.

Hayes, Bruce (2004). Phonological acquisition in optimality theory: the early stages. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in Phonological Acquisition*, Cambridge: Cambridge University Press. 158–203.

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

Hayes, Bruce (2014). Comparative phonotactics.

Hayes, Bruce & Zsuzsa Cziráky Londe (2006). Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* **23**. 59–104.

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

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

Hetzron, Robert (1975). Where the grammar fails. *Language* **51**.

Inkelas, Sharon, C. Orhan Orgun & Cheryl Zoll (1996). Exceptions and static phonological patterns: cophonologies vs. prespecification.

Inkelas, Sharon & Cheryl Zoll (2007). Is grammar dependence real? A comparison between cophonological and indexed constraint approaches to morphologically conditioned phonology. *Linguistics* **45**. 133–171.

Ito, Junko & Armin Mester (1995). Japanese phonology. In John Goldsmith (ed.) *Handbook of Phonological Theory*, Cambridge, MA: Blackwell. 817–838.

Ito, Junko & Armin Mester (1999). The phonological lexicon. In Natsuko Tsujimura (ed.) *The Handbook of Japanese Linguistics*, Oxford: Blackwell. 62–100.

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

Jusczyk, Peter, Paul Luce & Jan Charles-Luce (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* **33**. 630–645.

Jusczyk, Peter W, Derek M Houston & Mary Newsome (1999). The beginnings of word

segmentation in english-learning infants. *Cognitive psychology* **39**. 159–207.

Kapatsinski, Vsevolod (2009). *The architecture of grammar in artificial grammar learning: Formal biases in the acquisition of morphophonology and the nature of the learning task.* Ph.D. dissertation, Indiana University.

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

Kapatsinski, Vsevolod (2013). Morphological schema induction by means of conditional inference trees. In B. Cartoni, D. Bernhard & D. Tribout (eds.) *TACMO Workshop. Theoretical and Computational Morphology: New Trends and Synergies*.

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

Kiparsky, Paul (1968). Linguistic universals and linguistic change. In Emmon Bach, Robert Harms, Emmon Bach & Robert Harms (eds.) *Universals in Linguistic Theory*, New York: Holt, Rinehart and Winston. 170–202.

Kiparsky, Paul (2000). Opacity and cyclicity. *The Linguistic Review* **17**. 351–365.

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

Kochetov, Alexei (2006). Testing Licensing by Cue: A case of Russian palatalized coronals. *Phonetica* **63**. 113–148. ISSN 0031-8388.

Lepschy, Anna Laura & Giulio Lepschy (1998). *The Italian Language Today.* London: Hutchinson.

Lightner, Theodore (1965). *Segmental phonology of Modern Standard Russian.* Ph.D. dissertation, MIT, Cambridge, MA.

McCarthy, John J. (2007). *Hidden generalizations: Phonological opacity in Optimality Theory.* London: Equinox.

Mehler, Jacques, Peter Jusczyk, Ghislaine Lambertz, Nilofar Halsted, Josiane Bertoncini & Claudine Amiel-Tison (1988). A precursor of language acquisition in young infants. *Cognition* **29**. 143–178.

Melvold, Janis (1989). *Structure and Stress in the Phonology of Russian.* Ph.D. dissertation, MIT, Cambridge, MA.

Moore-Cantwell, Claire & Robert Staubs (2014). Modeling morphological subgeneralizations. In John Kingston, Claire Moore-Cantwell, Joe Pater & Robert Staubs (eds.) *Proceedings of Phonology 2013*, Linguistic Society of America.

Nevins, Andrew (2011). Phonologically-conditioned allomorph selection. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.) *The Blackwell Companion to Phonology*, Blackwell.

Padgett, Jaye (to appear). Russian consonant-vowel interactions and derivational opacity. In Wayles Browne, Adam Cooper, Alison Fisher, Esra Kesici, Nikola Predolac & Draga Zec (eds.) *Proceedings of FASL 18*, Ann Arbor, MI: Michigan Slavic Publications.

Paster, Mary (2006). *Phonological Conditions on Affixation*. Ph.D. dissertation, UC Berkeley, Berkeley, CA.

Pater, Joe (2000). Nonuniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* **17**. 237–274.

Pater, Joe (2006). The locus of exceptionality: Morpheme-specific phonology as constraint indexation. In Leah Bateman, Michael O'Keefe, Ehren Reilly & Adam Werle (eds.) *Papers in Optimality Theory III*, Amherst, MA: GLSA. 259–296.

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

Pertsova, Katya (2005). How lexical conservatism can lead to paradigm gaps. In Jeffrey Heinz, Andrew Martin & Katya Pertsova (eds.) *UCLA Working Papers in Linguistics*, Los Angeles, CA: UCLA, vol. 11. `http://www.linguistics.ucla.edu/faciliti/wpl/issues/wpl11/papers/Pertsova_Phono6.pdf`.

Prince, Alan & Paul Smolensky (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA, and Oxford, UK: Blackwell. Available as ROA-537 on the Rutgers Optimality Archive, http://roa.rutgers.edu.

R Development Core Team (2012). R: A language and environment for statistical computing. Vienna, Austria.

Shaw, Jason Anthony (2006). Learning a stratified lexicon. In Christopher Davis, Amy Rose Deal & Youri Zabbal (eds.) *Proceedings of NELS 36*, Amherst, MA: GLSA Publications, vol. 2. 519–530.

Sommerstein, Alan (1974). On phonotactically motivated rules. *Journal of Linguistics* **10**. 71–94.

Szpyra, Jolanta (1992). Ghost segments in nonlinear phonology: Polish yers. *Language* **68**. 277–312.

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

Weskott, Thomas & Gisbert Fanselow (2011). On the informativity of different measures of grammaticality. *Language* **87**. 249–273.

Wurm, Lee H & 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 Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory II (University of Massachusetts Occasional Papers in Linguistics)*, Amherst, MA: GLSA Publications. 533–571.

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