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

Michael Becker (IU) & Maria Gouskova (NYU)

November 5, 2012

abstract

Speakers generalize paradigmatic relations from their lexicon, e.g., English speakers extend the existing *fling* ∼ *flung* to the nonce *spling* ∼ *splung*. Dominant theories in linguistics model such effects using product-oriented generalizations, replacing the older source-oriented rule-based theories. We report an experiment on the generalization of mid vowel deletion ("yer" deletion) in Russian, showing that speakers extend a source-oriented generalization in addition to product-oriented generalizations. We model the speakers' behavior using multiple product-oriented grammars, with grammar separation triggered by inconsistent paradigmatic behavior. The speaker infers which product-oriented grammar to use based on the wellformedness of the source given each grammar.

1 Introduction

A central goal of all linguistic theories is to characterize the knowledge that speakers generalize from their language experience. There is a growing consensus that such generalization is based on symbolic rule-based processes (Pinker & Prince 1988, 1994; Prasada & Pinker 1993; Pinker 1999). Pinker's dual route model divides generalizations into rule-based (e.g., "English past tense is formed by the addition of −ed") vs. analogical (e.g., "English past tense is formed by internal stem change, as in *sit* ∼ *sat*") . Albright & Hayes (2002, 2003, 2006) demonstrate that all generalizations are rule-based—even those that are only partially productive. Yet this line of research is fairly neutral with respect to how these generalizations are stated. For example, English verb stems can be turned into past tense forms via the addition of a suffix that ends in [t] or [d]; relatedly, past tense forms are expected to end in a [t] or a [d] sound. Which of the two formulations is the better characterization of speaker knowledge? Bybee and colleagues (Bybee & Slobin 1982; Bybee & Moder 1983; Bybee 2001) noted that generalizations can be stated in these two different ways: *product-oriented* generalizations only describe the output, while *source-oriented* generalizations also make reference to the input. The question, then, is whether

*Thanks to Adam Albright, Karthik Durvasula, Gillian Gallagher, Bruce Hayes, Vsevolod Kapatsinski, John McCarthy, Colin Wilson, and to the phonology reading groups at NYU and UMass Amherst. For their help with participant recruitment, we thank Anna Aristova, Yevgenia Gouskova, Stephanie Harves, Sofya Kasyanenko, Barbara Partee, and Kevin Roon. Lastly, we thank our anonymous participants for their help.

speaker knowledge is source-oriented or product-oriented.

Generative linguistics started with a source-oriented, or rule-based approach (Chomsky & Halle 1968), where morphophonological relations were expressed with rules that transform one form into another. In English, for example, the rule would say something like "add /d/ to a verb stem to make a past tense form", capturing the structure of a paradigm such as *need* ∼ *needed*. But past tense forms can be made without any change, e.g., *cut* ∼ *cut* or *shed* ∼ *shed*. This option is only available to verb stems that already end in [t] or [d] (Albright & Hayes 2003). A rule-based approach can make two separate rules, one that adds [t] or [d] to most verbs, and one that adds nothing to verbs that end in [t] or [d]. The fact that both rules are related, i.e. that they both generate past tense forms that end in [t] or [d], would be left unexpressed in this theory. Cases like these, and similar cases in other areas of phonology (Kisseberth 1970), have motivated the use of product-oriented approaches in theoretical linguistics, both in generative linguistics (starting with Optimality Theory, Prince & Smolensky 1993/2004) and in usage-based linguistics (starting with Hooper [Bybee] 1976). Recent work argues specifically that speakers prefer product-oriented generalizations in artificial language tasks (Becker & Fainleib 2009; Kapatsinski 2009, 2011).

From the perspective of Optimality Theory, 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 that is impossible to express using the theory's standard product-oriented markedness constraints.

In this paper, we show that the Russian lexicon has a source-oriented generalization, in particular, one that bans a marked structure in the input, and that native speakers productively extend this generalization to nonce words. 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. The evidence we present 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). In the model we propose, grammars can only capture product-oriented generalizations; the source-oriented generalization is captured by a grammar inference mechanism that assigns 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. Section 3 presents the results of a nonce-word study that tests these generalizations. 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. Section 5 concludes.

2 The yer sublexicon

This paper examines the willingness of Russian speakers to accept the deletion of the mid vowels [e] or [o] (traditionally called “yers”) from the final syllable of novel nouns. Deletion 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, e.g., [nos] ‘nose NOM’ ∼ [nosa] ‘nose GEN’, while in other nouns the mid vowel of the base is deleted, e.g., [rot] ‘mouth NOM’ ∼ [rta] ‘mouth GEN’, not *[rota]. The lexically specific, or unpredictable nature of yer deletion in individual nouns has long been recognized (Farina 1991; Gouskova 2012; Gouskova & Becker to appear; Halle 1973a; Lightner 1965; Melvold 1989; Yearley 1995). 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 CCC cluster (a cluster of three cosonants), the middle consonant is almost always an obstruent, as in [kast^{j}or ∼ kas$\underline{t}$ra] ‘fire’. Examples with a middle sonorant exist, but they are much rarer, e.g., [agn^{j}ets ∼ ag$\underline{n}$tsa] ‘lamb’; usually a middle sonorant blocks deletion, e.g., [mudr^{j}ets ∼ mudr^{j}etsa] ‘wise’, *[mud$\underline{r}$tsa] (Gouskova 2012; Gouskova & Becker to appear; Yearley 1995). This restriction means that yer words (i.e. nominatives that lose a vowel in the genitive) and non-yer words are different: a sonorant before a final vowel is rare among the nominative forms of yer words, but common among non-yer word.

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 (to appear). 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 paradigms often lack a yerless 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.

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. The unattested genitive form [pamsta] is a possible word of Russian, and it is a possible product of yer deletion (cf.

[kast^{j}or ∼ ka<u>str</u>a] 'fire'), but it is not a possible product of yer deletion from a nominative form that ends in a complex coda. Therefore, this 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.

Like complex codas, monosyllables are also common in the non-yer lexicon, making up 8% of it. In the yer lexicon, monosyllables exist (e.g., [son ∼ sna] 'dream'), but they are rare, making 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 suffixed form have just one syllable.

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 1992; 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, Russian speakers prefer clusters 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.

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. There are two reasons, however, to think that speakers do not track sonority profiles of consonants that are separated by a vowel. First, there is no known mechanism that would track the relevant sonority profiles in the nominative, where they are separated by a vowel, and also find 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.

The second reason is typological: 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 1992, 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 not [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. 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. Whether such long-distance restrictions on consonant sonority are harder for learners to discover or are ruled out by universal markedness principles, they are not typical of human language phonologies.

Yet another limitation on yer deletion is that 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 to appear). Gouskova & Becker (to appear) show that speakers extend the limitation to novel words in a rating study, finding the deletion of a mid vowel in, e.g., [riʂon ∼ riʂn-a] more acceptable than the deletion of the high vowel in, e.g., [karut ∼ kart-a]. This limitation is source-oriented, since the quality of the vowel is defined only in the base. It is impossible to restate the mid vowel generalization as a product-oriented one (e.g., "the stem-final syllable cannot have a mid vowel"): there are mid vowels in stem-final syllables of some yer words even after deletion has applied, as in [vʲetʲer ∼ vʲetr-a] 'wind NOM/GEN'.

The distribution of vowel deletion, then, is not completely unpredictable in Russian. Only mid vowels can delete, but not in the presence of a complex coda. Deletion is preferred when it impacts polysyllables, and when it creates a cluster that ends in an obstruent followed by a sonorant.

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 tested the source-oriented generalization about the syllabic profile of the nominative base: deletion is blocked from a base like [poʂm] due to its final complex coda, while deletion from a base like [pʂom] will be allowed, even though both give rise to the same genitive form [pʂma]. The tested clusters included those that end in an obstruent followed by a sonorant (TR), e.g., [ʂl], [mʂl], [pʂl], in addition to clusters than end in two sonorants (RR), e.g., [xmn], and two obstruents (TT), e.g., [xʂt]. To test both monosyllables and disyllables, an initial CV sequence was added in half of the trials, making, e.g., [kipoʂm] and [kipʂom] based on [poʂm] and [pʂom]. The choice between the two mid vowels [e] and [o] was 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 (`http://odnoklassniki.ru`, `http://vkontakte.com`, `http://livejournal.com`) and through word of mouth.

The participants volunteered 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).

The use of online experiments has been gaining much traction in recent years, e.g., Becker et al. (to appear); Daland et al. (2011); Zuraw (2007), including comparisons of in-lab and on-line participants (Sprouse 2011).

3.2 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., [ʂox], and two genitive forms made by adding the suffix [a]. One genitive form was fully faithful to the base, e.g., [ʂoxa], while the other was 'yerless', or missing the mid vowel of the base, e.g., [ʂxa]. These nonce paradigms covered a wide range of consonant clusters in a variety of phonological environments, as schematized in Table 1.

Consonant pairs were used to make CVC bases, e.g., the pair [ʂx] was combined with the mid vowels [e] and [o] to make the monosyllables [ʂex] and [ʂox] in the nominative. The genitives were the two faithful forms [ʂexa] and [ʂoxa] and the yerless form [ʂxa]. 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 yerless genitive [piʂxa].[1]

[1]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.

			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 1: Nonce word types; each participant saw one each of the 48 word types. C = consonant, V = vowel, T = obstruent, R = sonorant.

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]. These four bases each had a faithful genitive with the suffix [a], and all four had the same yerless genitive [ʂxla]. Randomly chosen initial CV sequences were attached to create CVCCVC and CVCVCC bases, e.g., [muʂxel], [muʂxol], [muʂexl], [muʂoxl], all four with the yerless genitive [muʂxla].

The consonants were chosen from among the obstruents [p b t d g ʦ ʧʲ f s z ʂ ʐ x], marked T, and the sonorants [m n r l v], marked R, as seen in Table 2. This covered all the consonants of Russian (ignoring palatalization) except for [j], [ʃʃʲ], and [k kʲ]. We excluded [j] and [ʃʃʲ] because they do not combine freely with other consonants, and thus would not allow us to balance their appearance in different positions within clusters; [j] is limited to vowel-adjacent position, while [ʃʃʲ] is limited to pre-sonorant position in clusters. See below for [k kʲ].

The consonants were to be spelled without palatalization or non-palatalization (velarization) symbols. This spelling convention means that consonants were to be read palatalized before [e] and velarized elsewhere, except for four consonants: [ʦ ʂ ʐ] are always non-palatalized, and [ʧʲ] is always palatalized, regardless of following vowel letter. See further discussion of spelling in §3.5 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. Some of the possible combinations were not covered to make hypothesis testing more tractable. Since the distribution of sonorants in Russian CCC clusters is rather limited, we used only one of the three combinations that puts a

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

sonorant before another sonorant (namely TR̲R) and one of the four combinations that puts a sonorant before an obstruent (namely R̲TR). One additional combination (namely TR̲T) was found to be quite bad in Gouskova & Becker (to appear).

Some consonant combinations were excluded for phonotactic reasons. Pairs or triplets that contained two adjacent identical consonants or adjacent affricates were eliminated.[2] Any [t] or [ʦ] followed by [s] was eliminated, because the sequence would not be sufficiently distinct from plain [ʦ]. Combinations of obstruents were excluded if they did not agree in voicing, to make sure that the spelling agrees with the intended pronunciation without any voicing alternations.

Some consonant combinations were eliminated for morphological reasons: those that begin with [v], [s], or [k], because these consonants can be interpreted as prepositions, and those that end with [k] and [ʦ], because they can be interpreted as yer suffixes [-(o)k] and [-(e)ʦ]. This effectively eliminated [k kʲ] from the study.

Some combinations were eliminated for lexical reasons: consonant pairs were excluded if they appeared in real words with mid vowels (e.g., [dl] was excluded because it appears in the word [dol] 'valley'). Consonant triplets were excluded if they appeared in real words regardless of vowel quality (e.g., [psx] was excluded because it appears in [psʲix] 'psycho'). A stricter criterion could be applied to consonant triplets, because this stricter criterion still left hundreds of eligible triplets.

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 3.

The fillers were also three-word paradigms. These consisted of of an unaffixed

[2]In a study of Yefremova's (2000) dictionary, we found that adjacent affricates were severely underattested, with an observed/expected value of .06, yet adjacent stridents were not underattested (O/E .97).

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, tʂn, 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 3: examples of consonant combinations used in target items

nominative base, e.g., [ʧʲipan], a faithful genitive, e.g., [ʧʲipan-a], and a genitive with an unattested alternation in the manner or place of articulation of the last consonant, e.g., [ʧʲipam̲-a]. There were 45 filler paradigms.

To help participants assign the intended morphological categories to the test items, 74 frame sentences were created that presented a context for 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 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 1, 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 the 74 frame sentences.

The participants were told to expect sentences containing made up words that looked like Russian words. The first sentence included a word that they were asked rate on a scale from 1 to 5, with 1 for words that were very strange and did not look Russian, and 5 for words that looked normal and could easily be words of Russian. Following were two sentences with declined words, and the participant had to decide whether the words could be declined forms of the first word they rated. Participants were asked not to think too long about the answer but rather provide their first impressions. Subsequently, a short training example was shown with the word [kudat], which was offered with the comment that most Russian participants rated it as a "5": it could easily be a Russian word that they simply hadn't heard before. It was further offered that other speakers considered the form [kudata] to be a possible declined variant of [kudat], but they did not consider the word [kudasa] to be a possible form of [kudat]. Word stress or other alternations were not mentioned in the instructions; Gouskova & Becker (to appear) 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.[3] The unaffixed base form was presented first, in a random frame sentence where the base form would be interpreted as a nominative singular noun. The participant was asked to rate it on a scale of 1 (worst) to 5 (best) as a possible word of Russian, by pressing a button.[4] Once one of the five buttons was pressed, 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 yerless (suffixed but missing the base's final mid vowel). The participant was asked to decide whether or not the genitive form was acceptable by pressing one of two buttons. Once one of the buttons was pressed, a second genitive form was shown, again asking for a yes-or-no judgment. Each set of buttons was disabled once a choice was made, so participants could not go back and change their minds. The order of presentation of the faithful and yerless genitives was randomized.

Once the participant responded to all the items, they were presented with a form that invited them to provide demographic information, as listed in §3.1.

3.4 Results

This section starts with a survey of the participants' treatment of the nominative base, followed by their treatment of the faithful (yerful) genitive, followed by their treatment of the unfaithful yerless genitive. The raw data is available at `http://becker.phonologist.org/projects/yers/`. Participants gave the highest ratings to the cluster-free nominative forms (CV)CVC (4.16 on the 1–5 scale), followed by forms with two consonants before the final vowel (CV)CCVC (3.01), and forms with complex codas (CV)CVCC (2.48). 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). This can be seen in Figure 2. In this figure and throughout the paper, error bars represent 95% confidence intervals. 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 the disyllabic CVCCVC was preferred over the monosyllabic CCVC (3.22 vs. 2.80). This is shown in Figure 3.

[3]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.

[4]We used a scale of 1–5 rather than 1–7 because Russians are familiar with the 1–5 scale 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.

In this river, there lives a long **ʂer**

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

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

Ivan caught a long **ʂra**

Can this word be a declined variant of the word **ʂer**?

[no] [yes]

Ivan caught a long **ʂera**

Can this word be a declined variant of the word **ʂer**?

[no] [yes]

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

The mid vowel used, [e] vs. [o], had a miniscule effect (3.20 vs. 3.23), and the effect was very small across phonological sizes and sonority profiles.

The faithful genitive was found acceptable in 90% of all trials. This high acceptability changed little across categories, but nevertheless mirrored the pattern of the nominatives, with 91% for cluster-free forms, 90% for pre-yer clusters, and 88% for forms with complex codas. Again mirroring the rating of the nominative forms, (T)RR forms are slightly more acceptable than others (92% vs. 89%). These effects can be seen in Figure 4. The monosyllable effect mirrored that of the nominatives: CVC genitives were more acceptable than CVCVC (93% vs. 90%), and CVCC genitives were more acceptable than CVCVCC (89% vs. 86%), but CVCCVC were more acceptable than CCVC (90% vs. 89%).

Of most interest here is the acceptability of genitives with vowel deletion ("yerless genitives"). The overall acceptability of the yerless words was 25%. Yerless genitives were most acceptable for (CV)CVC bases (40%), less acceptable for (CV)CCVC bases (21%) and least acceptable for (CV)CVCC bases (15%). Recall that (CV)CVC bases give rise to bi-consonantal clusters, e.g., the cluster [ʂx] in [(pi)ʂxa], while both (CV)CCVC and (CV)CVCC bases give rise to the same tri-consonantal clusters, e.g., [ʂxl] in [(mu)ʂxla]. This difference between (CV)CCVC and (CV)CVCC mirrors the ratings of the bases, but it is rather surprising given that both have (CV)CCCa genitives. The effect of sonority profile does not mirror the ratings of the nominative forms: (T)RR yerless genitives are less acceptable than others (22% vs. 26%), not more acceptable. The yerless forms are most acceptable when the resulting cluster ends in an obstruent followed by a sonorant, i.e. (C)TR, as seen for TR, RTR, and TTR in Figure 5, with acceptability at 27%, vs. 24%

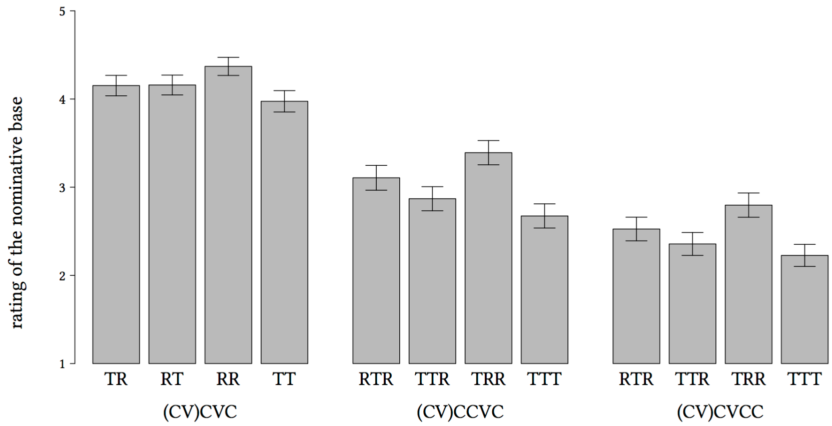


Figure 2: Rating of the nominative base, by cluster position and sonority profile (115 participants). Cluster-free (CV)CVC is best, (CV)CVCC is worst.

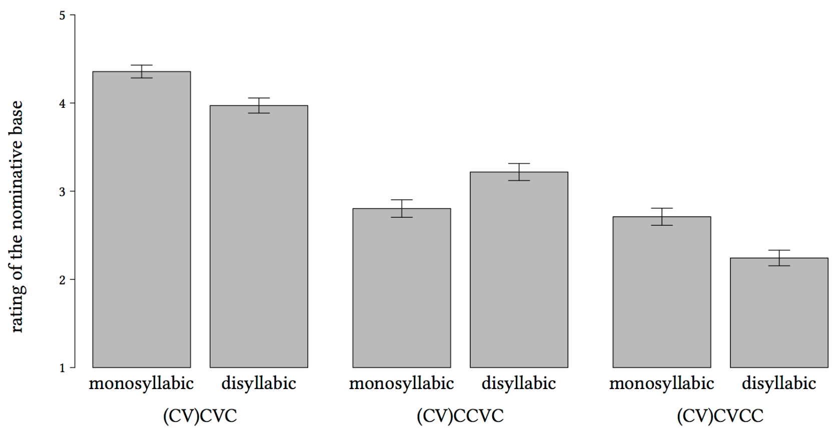


Figure 3: Rating of the nominative base, by cluster position and phonological size. Monosyllabic CVC and CVCC are preferred, as is disyllabic CVCCVC.

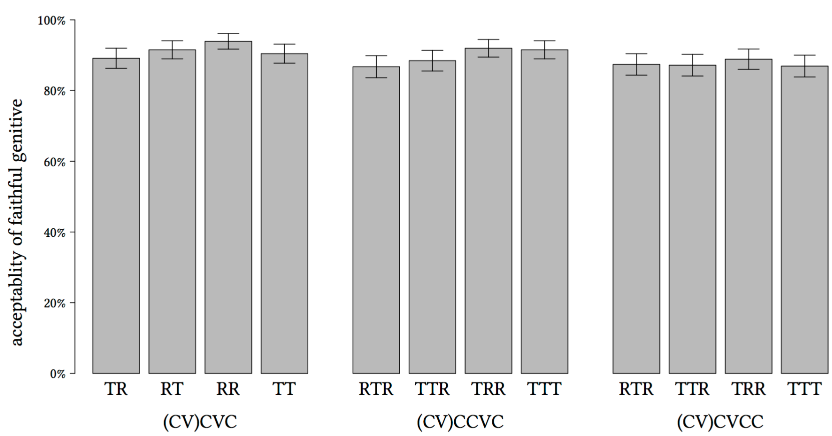


Figure 4: Acceptability of the faithful genitive, by cluster position and sonority profile. The faithful genitives are highly acceptable, especially (T)RR.

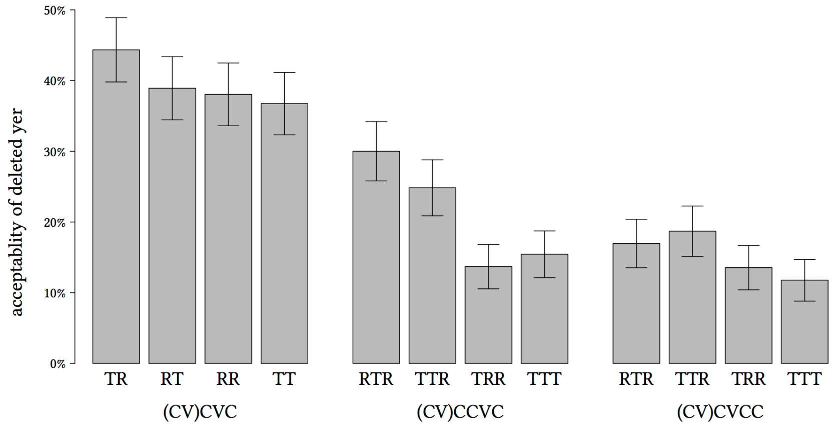


Figure 5: Acceptability of the unfaithful, yerless genitive, by cluster position and sonority profile. The yerless genitives are more acceptable in TR-final forms (TR, RTR, TTR).

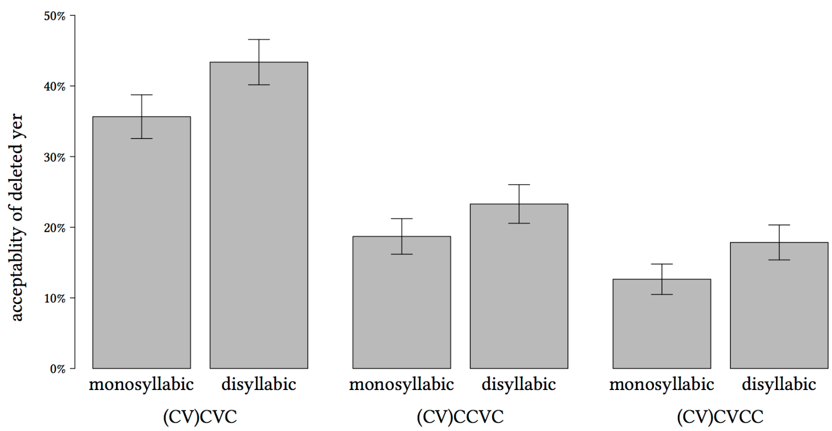


Figure 6: Acceptability of the unfaithful, yerless genitive, by cluster position and phonological size. Monosyllables are uniformly dispreferred.

for other sonority profiles. The monosyllabicity effect did not mirror the ratings of the nominative forms either: monosyllabic yerless genitives were uniformly less acceptable than disyllabic ones (22% vs. 28%), as seen in Figure 6. This was true in (CV)CVC bases (43% vs. 36%), (CV)CCVC bases (23% vs. 19%), and (CV)CVCC bases (18% vs. 13%). The yerless genitives also show a difference between [e] and [o], with the deletion of [o] being more acceptable (27% vs. 24%). The effect is stronger for (CV)CVC bases (42% vs. 37%) and (CV)CCVC bases (23% vs. 19%) than it is for (CV)CVCC bases (16% vs. 15%).

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 independent variable was the acceptance of the yerless 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 yerless 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 yerless 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.

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*. All the variables were then 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.

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. (2012), 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 4, 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 < 1.42, calculated using *mer-utils*, by

	β	SE(β)	z	$p\,(>\lvert z\rvert)$
(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 4: mixed effects logistic regression model for the experiment (115 participants)

Austin Frank, available at `https://github.com/aufrank/R-hacks`).

The model in Table 4 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 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 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 yerless 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 [^{j}o]. Additionally, the disyllabic items were not marked for stress, thus requiring participants to make a decision about stress location. To answer these questions, a norming study was conducted where

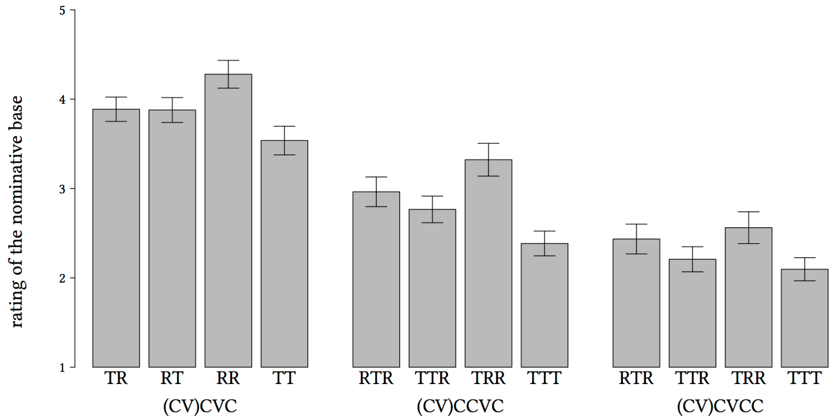


Figure 7: Norming study (15 participants): ratings of bases in a word likelihood task

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 age 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.

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 7 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 number of syllables 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 was heard 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 [ʧʲ].

As for palatalization and vowel quality of ⟨e⟩, recall that the expectation is for [ʧʲ] to always be palatalized and for [ʦ 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 ʦ]. The remaining consonants were palatalized in 91% of all tokens.

We conclude that Russian speakers overwhelmingly pronounce 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 extend yer deletion to nonce words, using both source-oriented and product-oriented generalizations. Tri-consonantal clusters are significantly less acceptable if they originate from bases with complex codas, confirming the productivity of this source-oriented generalization. The nominative base is more acceptable when its last two consonants are sonorants (RVR, TRVR, TVRR), but deletion is significantly more acceptable when the last two consonants are an obstruent followed by a sonorant (TVR, RTVR, TTVR). The goodness of (C)TR clusters, then, is product-oriented, since it holds true of the genitive product and not of the nominative source (though see §4 below about this classification).

In addition to these two generalizations, we found that deletion is significantly more acceptable in disyllabic forms than in monosyllabic forms. This contrasts with the ratings of nominative bases, where disyllables were dispreferred in (CV)CVC and (CV)CVCC bases. This shows that speakers successfully separate their judgments of the bases from their judgments of the yerless forms, but it does not necessarily inform the question of whether the monosyllabicity effect is source-oriented or product-oriented, since both monosyllabic nominative and monosyllabic genitives are equally dispreferred in the yer lexicon.

Finally, Russian speakers found [o] to be significantly more deletable than [e] in our experiment. While these vowels are missing from the genitive product, it would be rash to say that this effect is source-oriented. First, there is no significant difference in their frequency in the yer lexicon, nor is there a difference in their acceptability in the base here. Second, the deletion of [e] normally triggers a loss of palatalization on the consonant that is prevocalic in the base, such as the [x^j] in [x^jel $\sim$ xla] (although palatalization loss depends on the morphological and segmental context). Thus, the difference between [e] and [o] might be one of faithfulness within the paradigm, and does not inform broader asymmetries either in the lexicon at large or in the yer lexicon. We return to this point in §4.3.

We replicated some of the findings of Gouskova & Becker (to appear) on a larger scale: the monosyllabic effect and the source-oriented generalization about triconsonantal clusters derived from coda clusters. Our experiment answers some questions that were unresolved in Gouskova & Becker's (to appear) study. 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 in any way, 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 coda cluster effect is source-oriented. Furthermore, the present study establishes that the coda cluster effect affects both monosyllables and disyllables, and that it cuts across all types of clusters. Gouskova & Becker (to appear) only tested CVCC and CCVC monosyllables, not longer words, and there were enough differences between the items in the CVCC and CCVC categories that there might have been doubts about the source of the coda cluster effect. Our experiment tested the same exact consonant triplets in (CV)CVCC and (CV)CCVC words, showing that the effect was robust in a variety of contexts.

4 Analysis

The results from §3 show that Russian speakers extend generalizations about the distribution of yers from the lexicon onto nonce words. It was suggested 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).

This section starts in §4.1 with 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. Then, §4.2 presents an integrated 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 4.3 discusses the role of faithfulness in the model. Section

4.4 compares the result to the source-oriented Minimal Generalization Learner. Section 4.5 concludes.

4.1 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, the UCLA phonotactic learner (Hayes & Wilson 2008) was used to examine the sources and the 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 evaluate the testing data and generate a wellformedness score for each word.

The model 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 2 were used with the features found in 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 Hayes & Wilson (2008) model 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 wellformedness scores generated. These scores are the exponents of the probabilities assigned to the items, with the highest wellformedness score being zero (=100%). To better match the figures from §3, we negated and normalized these scores. The model's detailed output enables the assessment of the model both quantitatively and qualitatively: the wellformedness scores can be used in a regression model to predict the participants' responses (yes/no acceptance of the yerless 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 and tested on nominatives, and then it was trained and tested on yerless genitives. The results of these two runs are shown in Figure 8, grouped by prosodic shape and sonority profile. Training on nominatives (shown in Figure 8a) led to assigning the highest scores to the nonce (CV)CVC bases, followed closely by (CV)CCVC bases, with (CV)CVCC bases ranking lowest, matching the participants' preferences. The low scores of the words with complex codas is due 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. Training on yer genitives (shown in Figure 8b) led to assigning the highest scores to the nonce (CV)CVC words, but without a difference between (CV)CCVC and (CV)CVCC. This was entirely expected, given that (CV)CCVC bases and (CV)CVCC bases are indistinguishable in their yerless genitives. Nevertheless, this result offers a poorer match to the participants' preferences. 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.

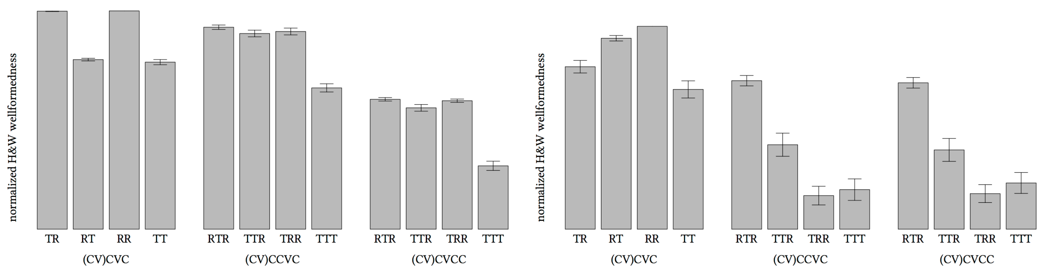
Recall that participants found yer deletion to be most acceptable in RTR and TTR bases, and least acceptable in TRR and TTT bases. This preference for (C)TR clusters is missing from the source oriented learner in Figure 8a, which correctly assigns low scores to TTT bases, but overestimates the goodness of TRR bases. The product-oriented learner in Figure 8b correctly assigns high scores to RTR and TTR, and low scores to TRR and TTT, matching the participants' responses. The phonotactic learner 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 codas, and one about the goodness of RTR and TTR items. But each training set only discovered one generalization. Training on the nominative source led to the discovery of the source-oriented generalization that (CV)CVCC bases make bad yer words. Training on the genitive products led to the discovery of the product-oriented generalization that RTR and TTR items make good yer words.

The generalization about the goodness of (CV)CVC items over either (CV)CCVC or (CV)CVCC 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.

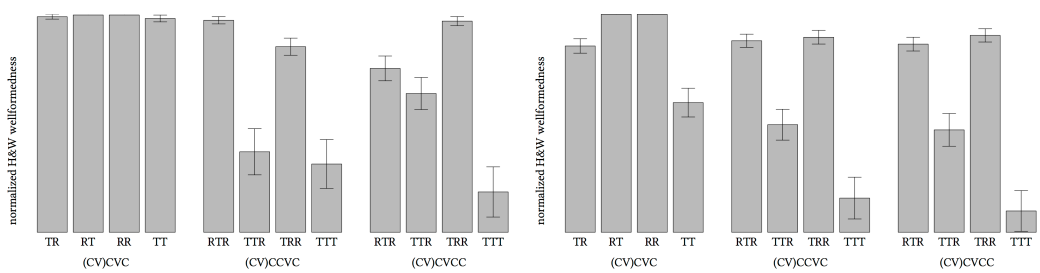
The preference for deletion of [o] over [e] wasn't discovered by either grammar; we turn this point in §4.3 below.

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 runs, 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 9. Training on non-yer nominatives led to a much worse match with the participants' responses. In the source-oriented training, (CV)CVCC items were wrongly assigned scores that are higher than the (CV)CCVC items.



(a) trained on yer nominative sources (b) trained on yer genitive products

Figure 8: The wellformedness scores assigned by the UCLA phonotactic learner to the experimental nonce words after it was trained on lexical yer words.



(a) trained on non-yer nominatives (b) trained on non-yer genitives

Figure 9: The wellformedness scores assigned by the UCLA phonotactic learner to the experimental nonce words after it was trained on lexical non-yer words.

In both trainings, TTR items were wrongly assigned scores that were lower than the TRR items. Unsurprisingly, then, generalizations about yer words cannot be found by looking at non-yer words, be the generalizations source-oriented or product-oriented.

To quantify how well the results of these four runs match the participants' behavior, model comparison was used. We used the *glm* function in R to fit a superset model that had the participants' acceptance of the yerless genitive as the dependent variable. The four predictors were the scores that each of the four runs assigned to the same items. Then, four subset models were fitted, each one having three of the four predictors, with one missing. Four ANOVA model comparisons were used to assess the contribution of each run 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 5, shows that the contribution from training on yer words, as measured by the χ^2 score, far outweighs the contribution of training on the non-yer training. The two yer word trainings make fairly comparable contributions, which is to be expected: each one captures one generalization that the other does not.

To summarize, we were able to closely replicate our experimental participants' judgments of yer deletion in nonce words by using two phonotactic grammars that were trained on lexical yer words in the nominative source and in the genitive product.

training	χ^2	df	p-value
yer words, nominative sources	70.9	1	$<.0001$
yer words, genitive products	74.6	1	$<.0001$
non-yer words, nominative sources	3.3	1	$<.1$
non-yer words, genitive products	15.6	1	$<.0001$

Table 5: The contribution of each training regimen to a superset model with the participants' responses as a dependent variable

The resulting phonotactic grammars correctly discovered the source-oriented constraint against coda clusters and the product-oriented sonority constraint. We now turn to a general model of how speakers decide whether words can undergo lexically restricted rules.

4.2 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 purely phonotactic grammar for the sources to the sublexicon, and a fuller grammar for the products of the sublexicon. The first grammar is a gate-keeper: it assigns a high probability to the words that are likely to be part of the sublexicon. The effect of the gate-keeper 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 second grammar 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 gatekeeper yer grammar GK_{yer}, and the yer grammar proper G_{yer}. For the non-yer words, it's the the non-yer sublexicon, the gatekeeper non-yer grammar $GK_{non\text{-}yer}$, and the non-yer grammar proper $G_{non\text{-}yer}$. All of these grammars are taken to be constraint-based. A grammar (G or GK) can also be seen as a function from words (or, more precisely, input-output mappings) to probabilities.

When a Russian speaker is given a nonce nominative base "NOM" and a nonce genitive "GEN" to go with it, and is asked whether "GEN" is acceptable, they calculate the probability of the paradigm [NOM ∼ 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).[5] The probability assigned by the yer grammar proper is modulated by the willingness of the gatekeeper grammar to allow the base in, and similarly for the non-yer grammar. In the rest of this subsection, we will demonstrate that we can estimate the probability in (1) using the wellformedness scores from the UCLA phonotactic learner and the participants' rating of the nominative base.

(1) $p(\textsc{nom} \sim \textsc{gen}) = p(GK_{yer}|\textsc{nom}) \cdot p(\textsc{gen}|G_{yer}) + p(GK_{non\text{-}yer}|\textsc{nom}) \cdot p(\textsc{gen}|G_{non\text{-}yer})$

To focus on the yerless genitives of interest, we continue the discussion with the example paradigm [ʂer ∼ ʂra], where the mid vowel of the nominative base is conspicuously missing in [ʂra]. Simply plugging in [ʂer] and [ʂra] into (1) yields (2).

(2) $p(\text{ʂer} \sim \text{ʂra}) = p(GK_{yer}|\text{ʂer}) \cdot p(\text{ʂra}|G_{yer}) + p(GK_{non\text{-}yer}|\text{ʂer}) \cdot p(\text{ʂra}|G_{non\text{-}yer})$

The second summand in (2) is easy to calculate: $G_{non\text{-}yer}$ does not allow deletion, so the faithfulness component in $G_{non\text{-}yer}$ will heavily penalize [ʂra], making $p(\text{ʂra}|G_{non\text{-}yer})$ very close to zero, thus making the product $p(GK_{non\text{-}yer}|\text{ʂer}) \cdot p(\text{ʂra}|G_{non\text{-}yer})$ very close to zero.

(3) $p(\text{ʂra}|G_{non\text{-}yer}) \approx 0$

(4) $p(GK_{non\text{-}yer}|\text{ʂer}) \cdot p(\text{ʂra}|G_{non\text{-}yer}) \approx 0$

This effectively eliminates the second summand, leaving (5).

(5) $p(\text{ʂer} \sim \text{ʂra}) \approx p(GK_{yer}|\text{ʂer}) \cdot p(\text{ʂra}|G_{yer})$

Since [ʂra] is quite likely as a yer genitive, $p(\text{ʂra}|G_{yer})$ is rather high. An estimate for $p(\text{ʂra}|G_{yer})$ is given by the genitive-trained UCLA phonotactic learner, which finds TR clusters such as [ʂr] quite acceptable as products (see Figure 8b).

We also need $p(GK_{yer}|\text{ʂer})$, the probability that [ʂer] is a word in the yer sublexicon, for which Bayes' theorem can be used, as shown in (6). The probability we need is equal to the probability that the yer gatekeeper grammar assigns to [ʂer], multiplied by the probability of this grammar, and divided by the overall probability of [ʂer].

[5]In this experiment, 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 (to appear) 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 [mest^j] 'revenge' and [lest^j] 'flattery' no longer alternate in inflectional paradigms (Gouskova 2012), retaining the vowel throughout.

(6) $p(\text{GK}_{yer}|\text{şer}) = \dfrac{p(\text{şer}|\text{GK}_{yer}) \cdot p(\text{GK}_{yer})}{p(\text{şer})}$

The denominator p(şer) is the overall probability of [şer] in the language, i.e. the sum of probabilities assigned to it by each of the grammars the speaker has (7). While we do not have estimates for the summands in (7), we do have an estimate for the sum: this is the participants' rating of the nominative base on the 1–5 scale.

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

We also have an estimate for $p(\text{şer}|\text{GK}_{yer})$: it is the probability that the nominative-trained UCLA phonotactic grammar assigns to the nominative sources. As for $p(\text{GK}_{yer})$, it is proportional to the size of the yer lexicon; it is just the proportion of yer words out of all of the words of the language (∼10%). It can thus be treated as a constant. Similarly, $p(\text{GK}_{non\text{-}yer})$ is the proportion of non-yer words in the language (∼90%). Together, $p(\text{GK}_{yer})$ and $p(\text{GK}_{non\text{-}yer})$ add up to one.[6]

The probability that [şer] is a yer word, then, is proportional to the probability that the yer grammar assigns to it, as in (8), divided by the acceptability of [şer].

(8) $p(\text{GK}_{yer}|\text{şer}) \propto \dfrac{p(\text{şer}|\text{GK}_{yer})}{p(\text{şer})}$

We now have an estimate of p(şer ∼ şra): it is approximately proportional to the probability that the nominative-trained UCLA phonotactic learner assigns to [şer], multiplied by the probability that the genitive-trained UCLA phonotactic learner assigns [şra], divided by the probability that the participants assigned to the nominative [şer]. Recall that we get from (9) to (10) because $p(\text{şra}|\text{GK}_{non\text{-}yer})$ is very close to zero. The reason $p(\text{şra}|\text{GK}_{non\text{-}yer})$ is close to zero is its faithfulness violation, which the phonotactic

[6] The reader may have noticed that in (1) and (2), a probability is given that is 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, as seen in the following proof.
Repeating (1):
$p(\textsc{nom} \sim \textsc{gen}) = p(\text{GK}_{yer}|\textsc{nom}) \cdot p(\textsc{gen}|\text{G}_{yer}) + p(\text{GK}_{non\text{-}yer}|\textsc{nom}) \cdot p(\textsc{gen}|\text{G}_{non\text{-}yer})$
Substituting given (6):
$= \dfrac{\text{GK}_{yer}(\textsc{nom}) \cdot p(\text{GK}_{yer}) \cdot p(\textsc{gen}|\text{G}_{yer})}{p(\textsc{nom})} + \dfrac{\text{GK}_{non\text{-}yer}(\textsc{nom}) \cdot p(\text{GK}_{non\text{-}yer}) \cdot p(\textsc{gen}|\text{G}_{non\text{-}yer})}{p(\textsc{nom})}$
Since $p(\text{GK}_{yer})$ and $p(\text{GK}_{non\text{-}yer})$ add up to 1:
$= \dfrac{\text{GK}_{yer}(\textsc{nom}) \cdot p(\text{GK}_{yer}) \cdot p(\textsc{gen}|\text{G}_{yer})}{p(\textsc{nom})} + \dfrac{\text{GK}_{non\text{-}yer}(\textsc{nom}) \cdot (1 - p(\text{GK}_{yer})) \cdot p(\textsc{gen}|\text{G}_{non\text{-}yer})}{p(\textsc{nom})}$
We know that $p(\textsc{gen}|\text{G}_{yer}) \leq 1$ (like any probability), and also that $\dfrac{\text{GK}_{yer}(\textsc{nom})}{p(\textsc{nom})} \leq 1$ (because of 7), therefore:
$\dfrac{\text{GK}_{yer}(\textsc{nom}) \cdot p(\text{GK}_{yer}) \cdot p(\textsc{gen}|\text{G}_{yer})}{p(\textsc{nom})} \leq p(\text{GK}_{yer})$, and similarly for the other summand.
Therefore, $p(\textsc{nom} \sim \textsc{gen}) \leq p(\text{GK}_{yer}) + 1 - p(\text{GK}_{yer}) \leq 1$.

learner does not represent; thus we cannot trust the UCLA phonotactic learner to assess this probability (see more about faithfulness in §4.3 below).

(9) $$\text{p(ṣer} \sim \text{ṣra)} \propto \frac{\text{p(ṣer}|\text{GK}_{yer}) \cdot \text{p(ṣra}|\text{G}_{yer})}{\text{p(ṣer)}} + \frac{\text{p(ṣer}|\text{GK}_{non\text{-}yer}) \cdot \text{p(ṣra}|\text{G}_{non\text{-}yer})}{\text{p(ṣer)}}$$

(10) $$\text{p(ṣer} \sim \text{ṣra)} \mathrel{\overset{\propto}{\approx}} \frac{\text{p(ṣer}|\text{GK}_{yer}) \cdot \text{p(ṣra}|\text{G}_{yer})}{\text{p(ṣer)}}$$

To quantify the goodness of (10), the wellformedness scores of the nominative-trained and genitive-trained phonotactic learner were transformed to probabilities by taking the inverse of their exponents. The participants' ratings of the nominative on the 1–5 scale were transformed to probabilities by subtracting them from 6 and taking the inverse. Then, a new predictor was created by multiplying the first two probabilities and dividing by the third. This new predictor is our estimate for the probability of the yerless genitive as expressed in (10). The participants' yes/no judgements of these yerless genitives is expected to be proportional to that.

To measure this new predictor's ability to match the participants' acceptance of the yerless genitive, ANOVA model comparison was used. The new predictor makes a significant improvement over a model that has just the phonotactic learner's probabilities of the nominatives ($\chi^2(1) = 81.02$, $p < .0001$), over a model that has just the phonotactic learner's probabilities of the genitives ($\chi^2(1) = 40.18$, $p < .0001$), and over a model that has both ($\chi^2(1) = 10.28$, $p < .005$).

To summarize, we offer a model that computes the probability of a paradigm in a language as a function 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 get into 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 thus limited in their influence to the grammar proper.

The UCLA phonotactic learner induces markedness constraints for each sublexicon separately. This property of the learner was used in §4.1 to show that the generalizations we claimed to hold can be found by an inductive process either in the source or in the product, or both. Yes classical constraint-based grammars assume one set of constraints that is possibly universal or innate (Prince & Smolensky 1993/2004). The grammar inference model in this section is compatible with this view; 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 disyllabic yer words and the preference for deletion of [o] over [e].

The preference for disyllabic yer words seems to be compatible with either source-oriented or product-oriented explanations. Overall, Russian speakers prefer monosyllabic nominatives, which make ∼10% of the lexicon. Among the yer words, however, monosyllables are rare — less than 1%. Thus, there is a source-oriented generalization present in the lexicon for speakers to use. The effect could also be product-oriented, however, and due to the presence of the preceding vowel, e.g., the cluster [ʂxl] is more acceptable when it is intervocalic in [muʂxla] than when it is word-initial in [ʂxla]; intervocalic clusters are cross-linguistically preferred to word-initial clusters. The UCLA phonotactic learner offers no help here: it correctly assigns higher wellformedness scores both to disyllabic yer nominative sources and to disyllabic yer genitive products. The responsible constraints are similar in both cases: with source-oriented training, the main constraint that prefers polysyllables is *#[+consonantal,+voice][−voice], which could only penalize CTVC bases, as all other bases either have a vowel or a sonorant as their second segment. With product-oriented training, most of the work is done by *#[−continuat,+coronal][−sonorant], which penalizes word-initial clusters with an obstruent as their second member, again leaving polysyllables unpenalized. Neither set of constraints is very satisfying, because it limits the difference between monosyllables and polysyllables to a narrow range of clusters, whereas participants preferred deletion in polysyllables across the board, including all CC and CCC clusters.

The preference for disyllabic yer words doesn't have to be a phonotactic effect, however—either source-oriented or product-oriented. It could be due to the faithfulness component of the grammar proper. The deletion in a monosyllable, as in [ʂxol ∼ ʂxla], affects the initial (and only) syllable of the base, violating the special faithfulness that protects initial syllables (Becker et al. 2011, 2012; Beckman 1997, 1998; Gouskova & Becker to appear). Initial syllable faithfulness would not be sensitive to sonority or cluster type, penalizing deletion in monosyllables across the board, as is observed. The preference for disyllabic yer words is therefore likely to be due to this faithflness effect.

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 yerless product. Yet it is not likely to be source-oriented, either. It is not a fact about the yer nominative sources, since [e] is somewhat more frequent than [o] in the final syllable the yer words (55% vs. 45%). If anything, then, [e] should be more representative of the yer lexicon, and thus more deletable. As expected, the nominative-trained UCLA phonotactic learner finds [e] more deletable than [o], unlike the participants, while the genitive-trained learner finds no difference. The markedness-based UCLA phonotactic learner 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.[7] Thus, a paradigm such as ⟨хел ∼ хла⟩ is interpreted as [x^{j}el ∼ xla], 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 markendness and faithfulness, markedness enforces the alternation, demanding that [xla] be preferred to *[x^{j}la], while the faithfulness constraint Ident(back) penalizes the discrepancy between [x] and [x^j]. Since Ident(back) only assigns violations when the palatalization-inducing ⟨e⟩ is deleted, it can account for the observed trend in the participants' responses.

If the vowel effect is due to aversion to palatalization alternations, then two facts about Russian should interact with it. First, the consonants that don't have a palatalization contrast [ʧ^j ʂ ʐ ʦ] would have no reason to disprefer the deletion of ⟨e⟩, since they would not alternate in palatalization. This is indeed the case: participants accept deletion of [o] more frequently than deletion of [e] (28% vs. 24%) before all consonants other than [ʧ^j ʂ ʐ ʦ], but before [ʧ^j ʂ ʐ ʦ] the difference almost disappears at 26% vs. 25%. Second, the consonant [l^j] retains its palatalization in real yer words (e.g., [l^{j}ev ∼ l^{j}va], *[lva] 'lion'), but not in the experiment (e.g., [l^{j}eʂ ∼ lʂa]), making the nonce paradigms less Russian-like in this case.[8] Indeed, participants accepted deletion of ⟨e⟩ before [l^j] much less frequently than before ⟨o⟩ (14% vs. 25%). Both expectations, then, are borne out, lending support to the faithfulness-based approach to the vowel effect.

The grammatical model developed in §4.2 was built on the markedness-based UCLA phonotactic learner. Yet faithfulness can be readily incorporated into the grammar proper (the product-oriented grammar) in a systematic way: once the UCLA phonotactic learner is trained on the products, and markedness constraints are induced, a list of universal faithfulness constraints are added to the grammar, and the full constraint set is reweighted. This resulting grammar will thus incorporate the product-oriented markedness effects, such as the preference for TR clusters, and faithfulness-based effects,

[7] As established in §3.5, Russian speakers choose the pronunciation [e] with preceding palatalization well over 90% of the time.

[8] 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 [l^j] consistently retains its palatalization (cf. [kaz^{j}ol ∼ kazl-a] 'goat', [d^{j}en^j] ∼ [dn^{j}a] 'day' but [ul^{j}ej ∼ ul^{j}j-a] 'beehive', [l^{j}on ∼ l^{j}na] 'linen').

such as the dispreference for deletion of ⟨e⟩.

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

4.4 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, or 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 yer from a non-yer word).

When we trained the MGL on the Russian lexicon,[9] 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 forms, 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 UCLA phonotactic learner.

To measure the MGL's performance against the UCLA phonotactic learner, a series of models were fitted with the participants' acceptance of the yerless genitive as the dependent variable. One model had the confidence scores assigned by the MGL as its only predictor; two additional models used the results of the nominative-trained and the genitive-trained UCLA learner; the forth had two predictors, using both the nominative-trained and the genitive-trained results. The models were compared on their AIC scores (Akaike Information Criterion), where lower AIC represents a better fit. The result, in Table 6, shows that the predictive power of the UCLA phonotactic learner outweighs that

[9]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
UCLA phonotactic learner, nominative + genitive	6019.8
UCLA phonotactic learner, nominative sources	6078.8
UCLA phonotactic learner, genitive products	6092.3
Minimal Generalization Learner	6197.2

Table 6: Comparison of the MGL and the UCLA phonotactic learner

of the MGL.

4.5 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 can allow the speaker to extend source-oriented generalizations from the real words of the language to novel words.

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 were able to use the wellformedness scores it generated to estimate the probability of paradigms, as a part of our attempt to model the participants' responses.

Another aspect of Russian yers that made the UCLA phonotactic learner 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). To generalize our approach to the Hungarian case, the membership of lexical items in sublexicons would be made gradient. A future version of the UCLA phonotactic learner could be generalized in this way, allowing the contribution of data points to training to be weighted.

While the mechanism was developed for the Russian case at hand, it is easy to see how it extends to other cases of stem alternations and allomorph selection. 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 can be captured with a phonotactic grammar that covers just these verbs. Extending the approach to allomorph selection can also offer 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). A phonotactic grammar computed over the [nɛk]-taking sources would discover such generalizations. 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; Bye 2007; Embick 2010; Nevins 2011; Paster 2006). 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 (Bye 2007; Embick 2010; Klein 2003), 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 would be readily available to a phonotactic grammar computed over the sources. In other words, the distribution would be 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 optimize these products phonologically.

5 Conclusions

This paper adduced evidence for a source-oriented phonological generalization. Russian yer deletion can produce a tri-consonantal cluster, as in [kastʲor ∼ ka$\underline{\text{str}}$a] 'fire', but this product depends on its source: a tri-consonantal cluster can only be created if the source of the derivation ends in a single consonant. The lexicon survey in §2 showed this to be true of the real words of Russian. 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). The model we propose uses product-oriented constraint-based grammars, together with a mechanism for partitioning the lexicon and learning multiple grammars. Rather than being encoded in any of these grammars, source-oriented generalizations arise from a grammar inference mechanism.

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 (Becker 2009; Becker et al. 2011; Flack 2007; Fukazawa 1999; Gouskova 2007; Ito & Mester 1999; Kawahara et al. 2002; Pater 2000, 2006, 2008, 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). Morpheme Structure Constraints 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). Morpheme Structure Constraints 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. We prefer our gatekeeper to be constraint-based, since constraints are independently needed in the grammar proper to derive products from sources; a uniformly constraint-based approach is likely to be more tractable in the long run.

While the UCLA phonotactic learner allowed us to implement all the key pieces of our proposed model, a full implementation of the model is still in the works. There are two main missing ingredients: the incorporation of faithfulness constraints, and an unsupervised mechanism for identifying groups of words that undergo identical morphophonological operations (e.g., "add [a] to make a genitive"). A full model will also allow lexical items to be associated with sublexicons gradiently and/or stochastically, and thus express the behavior of words that exhibit variable behavior. This complete model will bring us closer to a fully expressive learner for the morphophonological component of human language.

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.

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 (2012). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Ms. University of Glasgow.

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.

Becker, Michael, Andrew Nevins & Jonathan Levine (to appear). Asymmetries in generalizing alternations to and from initial syllables. *Language* .

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. (1992). Phonological primes: Features or gestures? *Phonetica* **49**. 181–193.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis & Ingrid Norrmann (2011). Explaining sonority projection effects. *Phonology* **28**. 197–234.

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 (to appear). Russian yer alternations are governed by the grammar. *Natural Language and Linguistic Theory* Ms. NYU and UMass Amherst.

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 & 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**.

Hooper [Bybee], Joan B. (1976). *An Introduction to Natural Generative Phonology*. New York: Academic Press.

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.

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.
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.
Kisseberth, Charles (1970). On the functional unity of phonological rules. *Linguistic Inquiry* **1**. 291–306.
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.
Melvold, Janis (1989). *Structure and Stress in the Phonology of Russian*. Ph.D. dissertation, MIT, Cambridge, MA.
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`.

Pinker, Steven (1999). *Words and rules*. New York: Harper Perennial.

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

Pinker, Steven & Alan Prince (1994). Regular and irregular morphology and the psychological status of rules of grammar. In Suzi D. Lima, R. L Corrigan & Gregory Iverson (eds.) *The reality of linguistic rules*, Amsterdam: John Benjamins Publishing Company, vol. 26. 321.

Prasada, S. & Steven Pinker (1993). Generalisation of regular and irregular morphological patterns. *Language and cognitive processes* **8**. 1–56.

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.

Sprouse, Jon (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* **43**. 155–167.

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.

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.

Yefremova, T. F. (2000). *Novyj tolkovo-slovoobrazovateljnyj slovarj russkogo jazyka [New Dictionary of Russian]*. Moscow: Russkij Jazyk.

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

Zuraw, Kie (2007). The role of phonetic knowledge in phonological patterning corpus and survey evidence from tagalog infixation. *Language* **83**. 277–316.