Lexical and phonological variation in Russian prepositions*

Tal Linzen, Sofya Kasyanenko, and Maria Gouskova

July 26, 2013

Abstract

Phonological rules can be variable in two ways: they can apply to a subset of the lexicon (lexical variation), or apply optionally, with a probability that depends on the phonological environment (stochastic variation). These two types of variation are occasionally seen as mutually exclusive. We show that the vowel-zero alternation in Russian prepositions ([s trudom] 'with difficulty' vs. [sə stinoj] 'with the wall') exhibits both types of variation at the same time. In two corpus studies and a nonce word experiment, we document novel stochastic factors that influence the alternation: similarity avoidance, stress position and sonority profile. These constraints interact additively, lending support to a weighted constraint analysis. In addition to the phonologically determined stochastic variation, we find significant lexical variation: phonologically similar nouns differ in the rate at which they condition the alternation in the prepositions. We analyze this pattern by augmenting the weighted constraints approach with lexical scaling factors.

1 Introduction

Much of the phonological literature is concerned with rules that apply categorically whenever their phonological context is present, but many phonological rules are variable: they sometimes fail to apply in the relevant phonological environments. This can happen in one of two ways. The variation can be lexical, such that the rule only applies to a subset of the lexicon (Inkelas et al. 1997, Zuraw 2000, Pater 2000, 2006). For example, Chomsky and Halle (1968) observe that derived forms in English conserve the stress pattern of the base, as in *impórt* ∼ *impòrtátion*. Yet there are many words that this rule does not apply to, even when their phonological form is very similar to the cases in which it does apply; for example, *consúlt*, which ends in a sonorant-stop sequence just like *impórt*, shifts its stress position in the derived word *cònsultátion* (Pater 2000). Another way in which rules can be variable is stochastic variation, where a rule applies optionally, with a probability that may depend on the phonological environment. One of the best-studied cases of stochastic variation is English *t/d* deletion: e.g., *west* can either be pronounced [wɛst] or [wɛs] (Guy 1980, Labov 1989, Guy 1991, Raymond et al. 2006, Coetzee and Kawahara in press, inter alia). The probability of *t/d* deletion is affected by a series of phonological factors; for example, it is more likely to apply when the syllable containing *t/d* is unstressed than when it is stressed (Labov 1989).

Lexical and stochastic variation have typically been studied as two separate phenomena. Indeed, some models of phonology consider them to be mutually exclusive: in Lexical Phonology, for example, stochastic

*For feedback that led to many improvements to this work, we would like to thank Lisa Davidson, Gillian Gallagher, Vera Gribanova, Gregory Guy, Alec Marantz, and audiences at NYU, Tel Aviv University, SYNC, and OCP 10 in Berlin, as well as three anonymous reviewers and the associate editor. We would also like to thank our many anonymous participants on the Internet, as well as Anna Aristova, Stephanie Harves, Yevgenia Gouskova, Barbara Partee, and Kevin Roon for help with recruiting the participants.

rules apply at a phonetic level of representation, which is blind to the lexical content of the utterance (Kiparsky 1985). Documented effects of lexical frequency on English *t/d* deletion call this assumption into question (Bybee 2000, Coetzee and Pater 2011). Recently, Coetzee and Kawahara (in press) have analyzed lexical frequency effects on stochastic variation using a weighted constraints approach (Legendre et al. 1990, Smolensky and Legendre 2006, Pater 2009, Coetzee and Pater 2011) augmented with "scaling factors" determined by the word's frequency. What is unclear is how this architecture would account for cases in which stochastic variation is influenced by the *identity* of the word. Indeed, examples of lexical idiosyncrasies in stochastic rules have been reported only sporadically and are typically limited to very common function words, such as *and* in the case of *t/d* deletion (Neu 1980, Guy 2007).

We explore the interaction between stochastic and lexical variation in the case of the vowel-zero alternation in the Russian prepositions [s], [v] and [k] (Matushansky 2002, Timberlake 2004, Katz 2005, Steriopolo 2007, Gribanova 2009a, Blumenfeld 2011). These prepositions either surface as a single consonant (C) or as a consonant-vowel (CV) sequence ([sa/sə], [va/və] and [ka/kə]). The vowel is [a] when the following syllable is stressed and [ə] otherwise (see Crosswhite 1999, Padgett and Tabain 2005, Bethin 2006, and others).[1] Other prepositions, such as [za] 'behind', [po] 'along' and [na] 'on', do not alternate.

Before words beginning with a vowel or a single consonant, only the C forms are found. When the words begin with a consonant cluster, both CV and C forms are attested. The choice of form depends to some extent on the phonology of the cluster. Clusters with rising sonority, such as [pr], normally condition the C forms, as shown in (1). The prepositions [s] and [v] typically surface as CV when the following cluster starts with a consonant identical to the preposition, as shown in (2).

(1) C form conditioned by clusters with rising sonority

a.	s prikázəm	*sə prikázəm	'with the order'
b.	v bróf^{j}	*va bróf^{j}	'in the eyebrow'
c.	k slanú	*kə slanú	'towards the elephant'

(2) CV form used if following cluster starts with the same consonant as the preposition

a.	*s stərikóm	sə stərikóm	'with the old man'
b.	*v vrém^{j}ə	va vrém^{j}ə	'in time'

The generalizations in (1) and (2) have been described as categorical (Matushansky 2002, Timberlake 2004, Steriopolo 2007). As we will see in detail, this is generally not the case: these patterns are strong tendencies rather than inviolable constraints. Moreover, there are cases of relatively free variation between C and CV forms before certain cluster-initial words. This is particularly common when the first consonant of the word is a sonorant (Matushansky 2002, Steriopolo 2007), as shown in (3).

[1] All transcriptions are in IPA, except where noted. We do not transcribe palatalization before [e] and [i]; to indicate palatalization contrasts before unrounded high vowels, we use [i] and [ɨ]. Our transcriptions are fairly broad but reflect obstruent voicing alternations and vowel reduction. When we refer to prepositions in isolation, we use the string [so], following the Russian orthography. The pretonic allophone of /o/ and /a/ has been transcribed in the literature as [ʌ] (Avanesov 1968), [ɐ] (Crosswhite 1999) or [a] (Barnes 2004). Since the precise phonetic quality of the vowel is not relevant to our analysis, we transcribe it as [a] for readability. Likewise, we transcribe unstressed front vowels (reduced /i/ and /e/) as [i] rather than the phonetically more accurate [ɪ]. We use the following abbreviations in glosses: "acc" for "accusative", "nom" for "nominative", "dat" for "dative", "gen" for "genitive", "inst" for "instrumental", "prep" for "prepositional", "sg" for "singular", "pl" for "plural".

(3) Variation between the C and CV prepositions before the same word

s mnóz̧əstvəm ∼ sa mnóz̧əstvəm 'with a great number/set'

Variation is not always possible, however. Timberlake (2004, p. 177) observes that the CV form of the preposition is categorically selected before certain common function words, in a way that is not predictable from the phonology of the word-initial cluster (see (4)). This suggests that the choice between the C and CV forms is not determined entirely by phonological factors.

(4) Some function words categorically condition CV prepositions

*s mnój sa mnój 'with me'
*s fsémi sa fsémi 'with all'

These properties of the alternation make it an ideal test case for studying the interaction between lexical and phonological variation. Moreover, the alternation is represented orthographically, which makes it possible to study its properties in a large orthographic corpus. In the two corpus studies described in the paper, we uncovered a set of new stochastic phonological constraints, and found that the lexical variation is much more extensive than previously known and exhibits several surprising morphological properties (see Table 1 for a summary of the results of the two corpus studies).

We document three groups of new phonological constraints. First, we show that the alternation is sensitive to the position of stress in the following word: the vowel is more likely to appear when the following syllable is stressed. Second, the dissimilation pressure which favors CV forms turns out to be more widespread than previously thought: it applies across intervening consonants (e.g., in [s̱+fs̱]) and between labials with different manners of articulation ([v+m]). This suggests that the alternation is conditioned not just by geminate avoidance but by a more general version of the Obligatory Contour Principle (OCP, Leben 1973, McCarthy 1986 et seq.). Finally, we confirm and elaborate the effect of sonority sequencing: we show that the alternation is affected not just by the sonority of the first consonant (Steriopolo 2007) but by the sonority profile of the entire word-initial cluster (e.g., [sa+vḏ] is more likely than [sa+vṟ]).

We report on an experiment that confirmed that Russian speakers productively apply the phonological constraints to nonce words ("wugs", Berko 1958). We find that the interaction between the constraints is additive: the more constraints are violated in the cluster, the more likely is the vowel to surface. We model the results in a Maximum Entropy grammar with weighted constraints and probabilistic outputs (Goldwater and Johnson 2003, Hayes and Wilson 2008; Coetzee and Pater 2008, Pater 2009, Potts et al. 2010).

To capture the different CV rates conditioned by different morphemes with similar phonological properties, we propose that individual morphemes can adjust the language-wide constraint weights (cf. Pater 2006). Our implementation takes the scaling factors approach proposed by Coetzee and Kawahara (in press) as its starting point but modifies it in several ways. The scaling factors analysis allows us to straightforwardly analyze a register effect, whereby ecclesiastical words as a group (such as [krést] 'cross') show a higher CV rate than other words in the language.

Lexical variation interacts with the phonological constraints we identified. For example, the word [dvor] 'yard' shows a strong tendency to appear with the CV forms of all three prepositions; yet this tendency is amplified for [v], presumably due to similarity avoidance: the mean CV rate for [s] and [k] before [dvor] is 62%, whereas for [v] before [dvor] the rate is a near-categorical 96%. We show that the joint effect of lexical and phonological constraints arises naturally in our additive weighted constraints framework without any additional stipulation.

The structure of the paper is as follows. Section 2 examines the alternation in the Russian National Corpus. Section 3 reports an additional corpus study, performed using a Web search engine. Section 4 reports the results of the nonce word experiment designed to test the productivity of the phonological patterns we found in the corpus studies. Section 5 proposes a set of constraints that account for the phonological patterns we found, and models the results of the nonce word experiment using Maximum Entropy Grammar (Goldwater and Johnson 2003, Hayes and Wilson 2008). Section 6 takes up lexical variation, and Section 7 concludes.

2 Corpus study 1: Russian National Corpus

This section presents the results of a study of the prepositional vowel-zero alternation in the Russian National Corpus (RNC). We compiled a list of all cluster-initial words in Russian and counted how many times each variant of the prepositions appeared before each of those words in the RNC. The methodology of this study is described in Section 2.1. The comprehensive database obtained in this way allows us to explore the phonological factors that condition the alternation, both known and new ones (Section 2.2). We then show that these phonological patterns do not fully explain the distribution of the alternation, and that lexical variation must be admitted (Section 2.3). Finally, Section 2.4 addresses the connection between prepositional vowel-zero alternations and Russian *yer* deletion (e.g., [l^jón] ∼ [l^jn-á] 'linen (nom/gen sg)'). Traditional accounts treat the vowel-zero alternation in theses prepositions as a case of yer alternation, therefore predicting that *yer* words should categorically select the CV form. Contrary to this traditional view, we propose that monosyllabic words of this type largely condition prepositional alternations based on their phonotactics, with some lexical variation.

2.1 Methods

2.1.1 Corpus searches

Words that start with a simple onset invariably condition C forms. We therefore focused on cluster-initial words. Even though the prepositions can appear before words of several different syntactic categories, including adjectives and participles, we chose to limit our corpus searches to nouns, in order to keep the data set to a manageable size. We constructed a list of nouns with cluster-initial inflected forms based on a digital version of Zaliznjak's (1977) dictionary (Usachev 2004). To simplify the automatic selection of inflected forms, we only considered paradigms that had exactly 12 forms—6 for each case in both the singular and the plural, without paradigm gaps or two alternative forms for the same combination of case and number. We further excluded from our sample those lemmas for which only some of the inflected forms started with a cluster; for example, we excluded the root [mn-] 'me', which can surface either as [mnój] 'me (inst)' or [min^já] 'me (gen)'. Finally, we excluded lemmas that had a frequency of fewer than 5 tokens per million, since they were unlikely to have a significant number of occurrences with the prepositions of interest. To reduce the number of online queries sent to the Russian National Corpus site, we performed this preliminary elimination offline, using a frequency database (Sharoff 2005). All of the data discussed in the article are posted at http://tallinzen.net/projects/russian_preps/.

There are five ways in which the prepositions [k], [s] and [v] can combine with a word, as shown in (5).

We generated up to 20 search queries for each lemma: singular and plural forms, in each of the five case forms listed in (5), preceded by either form of the preposition. The total number of queries was 5502. The queries were performed using an automated computer program in April 2011, when the corpus consisted of 150 million words. Words that had fewer than 10 tokens in the RNC with either form of a preposition were excluded from further analysis with that preposition. For example, a dative form that appeared 6 times with [k] and 7 times with [ko] was included in the sample for [k], but one that had 6 [k] tokens and 3 [ko] tokens was not. In total, 601 inflected forms met the threshold for [k]/[ko], 1133 for [v]/[vo] and 1017 for [s]/[so].

Phonological factors	
OCP: Adjacent sibilants (s + {s, z, ş, z̧}) and adjacent labial fricatives (v + {f, v}) tend to be avoided in the beginning of clusters (see 2.2)	[sə stərikóm] > [s stərikóm]
Non-adjacent OCP: multiple sibilants and labials (of any manner of articulation) tend to be avoided in the same cluster, even when they are not adjacent (see 2.2, 3.2.3)	[sa vzórəm] > [z vzórəm]
Stress: CV forms are more likely before stressed syllables than before unstressed ones (CV+σ́ > CV+σ̆) (see 3.2.2)	[sa vdóxə] > [sə vdaf͡tsá]
Sonority slope: CV forms are uncommon in clusters with steeply rising sonority, and become more likely before level or falling sonority clusters (see 3.2.4)	[s prikázəm] > *[sə prikázəm]; [sə vdaf͡tsá] > [s vridámi]
Lexical factors	
Lexical idiosyncrasies: Lexical items with similar phonological properties differ widely in their tendency to appear with CV forms (see 2.3)	[sə dvará], [və dvaré] vs. [z dvajniká], [v dvajniké]
Inflected form idiosyncrasies: Two inflected forms of the same lexical item, occasionally associated with different senses of the word, may differ in their tendency to appear with CV forms (see 2.3.2, 2.3.3)	[z dn^{j}óm] ‘with the (holi)day’ vs. [sa dn^{j}á] ‘from the day’
Monosyllabic yer words: Words assumed to contain an underlying “yer” vowel do not pattern uniformly with respect to the prepositional C/CV alternation, suggesting that the prepositional alternation is not an instance of the more general yer alternation (see 2.4)	[ka sn-ú] ‘towards sleep’ vs. [k psú] ‘towards the dog’
Conservative contexts: ecclesiastical and archaic contexts, as well as certain frozen expressions, favor CV forms (see 2.3.1, 2.3.4)	[liz̧át^j va práxi] ‘be dead (lit. lie in ashes)’ vs. [práx k práxu] ‘ashes to ashes’

Table 1: Overview of the results of the two corpus studies

(5) Case forms of nouns in preposition phrases with [s], [v] and [k]: sample queries for [dvor] ‘yard’

		Singular	Plural	
a.	[k/ko] + dative (‘towards’):	g/kə dvar-ú	g/kə dvar-ám	‘towards the yard(s)’
b.	[s/so] + genitive (‘from’):	z/sə dvar-á	z/sə dvar-óf	‘from the yard(s)’
c.	[s/so] + instrumental (‘with’):	z/sə dvar-óm	z/sə dvar-ámi	‘with the yard(s)’
d.	[v/vo] + accusative (‘into’):	v/va dvór	v/və dvar-ɨ́	‘into the yard(s)’
e.	[v/vo] + prepositional (‘in’):	v/və dvar-é	v/və dvar-áx	‘in the yard(s)’

2.1.2 Identifying CV preference

As noted above, the C forms are the default in the language, and they are the only grammatical option before simple onsets (non-clusters). Most of this section is therefore concerned with the conditions under which a word may be preceded by the CV form. Due to the noisy nature of corpus results, a single occurrence of a word with a CV preposition does not necessarily indicate that the word is grammatical with the CV form. To identify inflected forms that robustly take a CV preposition, we used the following simple procedure. For a given word form and preposition, the number of CV forms can be seen as a draw from a binomial distribution. We estimated a 95% confidence interval for the proportion of CV prepositions using the Clopper and Pearson (1934) method, implemented as *binom.test* in R (R Core Team 2012). We then selected for further investigation those words for which the lower bound of the confidence interval was higher than 5%—that is, words for which we could be fairly confident that the CV rate was higher than 5%. We also excluded inflected forms that appeared fewer than 20 times in the corpus with either variant of a given preposition (except as noted). In total, 120 forms met these criteria and are discussed in the rest of this section.

2.1.3 Consonant inventory and sonority

Since much of the following discussion concerns consonant natural classes, we give the consonant inventory of Russian in (6).

(6) Consonant inventory of Russian

	Labial	Dental	Palatoalveolar/Palatal	Retroflex	Dorsal
Stops	p pʲ b bʲ	t tʲ d dʲ			k kʲ g gʲ
Fricatives	f fʲ	s sʲ z zʲ	ʃʃʲ	ʂ ʐ	x xʲ
Affricates		t͡s	t͡ʃʲ		
Nasals	m mʲ	n nʲ			
Laterals		l lʲ			
Trills/flaps		r rʲ			
Glides	v vʲ		j		

Surface [f] can be derived from /v/ by devoicing, but there are also non-alternating examples, usually from a loan source (e.g., [ftór] ‘fluorine’, from Greek *phthoros*, or [fól] ‘a foul’, from English). This latter distinction is reflected in the orthography, a fact which is relevant for our study. Finally, all of the consonants in the inventory can appear in word-initial CC clusters (see Davidson and Roon 2008 for review) except for [j], which must be adjacent to a vowel (Padgett 2008).

We use the scale in (7) in our discussion of sonority slope effects. To calculate the sonority slope between the first and second consonant in a cluster, we subtract the sonority rank of the first consonant in the cluster from the sonority rank of the second. Thus, [r_2t_6] has a sonority slope of −4, and [d_6n_4] has a sonority slope of +2.

(7) Sonority scale:

$v_1 > r_2 > l_3 > n_4 > s_5 > t_6$

Most of this scale is uncontroversial, but a note is in order on the labial continuant [v]. This sound is famous in the literature on Russian for its ambiguous status (Hayes 1984, Kiparsky 1985, Padgett 2002): in voicing assimilation, it patterns with sonorants in supporting a voicing contrast in the preceding obstruent (e.g., [svét] 'light' vs. [zvón] 'ringing'), but with obstruents in undergoing the rule itself (e.g., /v+bar/ → [v bár] 'into the bar' vs. /v+par/ → [f pár] 'into the steam'). We therefore assume the sonority scale given in (7), while acknowledging that [v] may occasionally show ambiguous patterning.

2.2 Phonological factors

The corpus results confirmed the generalization that rising sonority clusters typically condition C forms of the prepositions. Most words beginning with steeply rising clusters such as [pr] showed CV rates of less than 1%, although there was considerable lexical variation in this group as well (see Section 2.3 below). We also confirmed the generalization that CV forms of the prepositions [s] and [v] are used to avoid clusters of identical or similar consonants—[so] is used before sibilant-initial clusters (e.g., [sə stərikóm 'with an old man'), and [vo] is used before the labial continuants [v] and [f] (e.g., [va fránt͡sii] 'in France'). We did not find evidence that [ko] is used more than [k] before the dorsal segments [k], [g] and [x], in line with Matushansky (2002) and Steriopolo (2007) (but contra Timberlake 2004).

The proportion of the CV forms of [s] before sibilants was high, but not 100%. While all sibilants show a strong preference for the CV preposition [so], this preference is stronger when the word starts with [s] (mean proportion of the CV form: 94%) than when it starts with [ʂ] (86%) or [z] (88%). The difference between CV rate in [s] and in the other sibilants is significant ($p < 0.05$ in both cases, Wilcoxon test). This suggests that the strength of the OCP effect depends on the degree of similarity between the preposition and the following consonant (see Section 3.2.1 below for similar results and discussion). The strong but noncategorical preference for the CV form before sibilants is also corroborated in our nonce word experiment (Section 4). Thus, the alternation is gradient even before sibilants: while there is a strong preference for [so] before [svint͡sá], occasionally a speaker will produce [s svint͡sá] 'from lead'.

We additionally found that the OCP effect is more extensive than previously thought: it applies to more consonants than has been reported, and to non-adjacent consonants within the same cluster. Figure 1a shows that among the words that robustly occur with the CV form [so], this CV form is more strongly favored when the *second* consonant of the word-initial cluster is a sibilant, as in the lemma [vzór] 'look': the median proportion of CV forms is 90% when the second consonant is a sibilant, compared to 58% when it is not. The preposition [v/vo] behaves in a similar way (Figure 1b): when the second consonant of the following word is a labial continuant, there is a higher proportion of CV forms.[2]

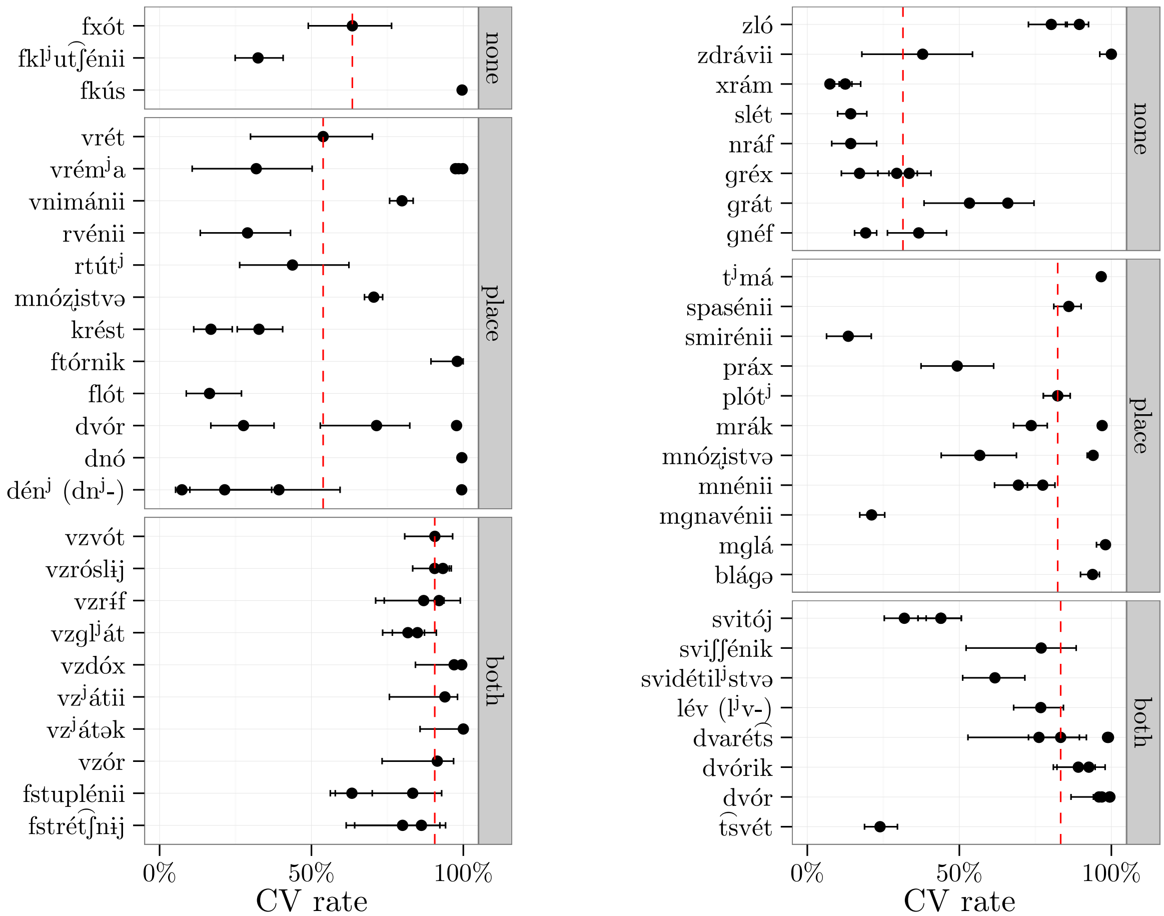
[2] Anticipating the discussion in Section 2.4, it should be noted that some of the examples in Figure 1a have prefixes that have yer alternations, although they are quite limited in productivity (Pesetsky 1979, Svenonius 2004, Gribanova 2009b, 2010,

Moreover, we found that the OCP holds not only between adjacent consonants that share place and manner, but also between homorganic consonants that differ in manner. Words with a labial consonant anywhere in their initial cluster, regardless of its manner of articulation, favor the CV form [vo] more than words that do not have a labial consonant (Figure 1b). For example, a word like [mrák] 'darkness', which starts with the labial [m], will be more likely than a word like [nráf] 'character' to appear with [vo], all other things being equal.

By contrast, when the preposition [s] is followed by a coronal consonant, it only dissimilates from it when that consonant is a sibilant: we did not find evidence that it dissimilates from [t], [d] or [n]. This sensitivity to manner of articulation in coronals but not in labials mirrors the situation in Arabic and Muna: co-occurrence constraints in these languages distinguish subclasses of coronals, but not labials or dorsals, possibly because coronals are perceived as more heterogeneous than consonants in other places of articulation (Coetzee and Pater 2008). Since Russian does not have a non-fricative monoconsonantal coronal preposition (e.g. [t] or [n]), there is no way to determine whether the dissimilation pressure applies to coronals with identical manner of articulation, as in Arabic or Muna, or specifically to sibilants, as in Chol (Gallagher and Coon 2009).

We did not find similarity avoidance effects after the dorsal preposition [k], not even when the following word began with a [k]. Russian differs in this respect from languages with co-occurrence restrictions in roots, which tend to apply them regardless of place of articulation (Coetzee and Pater 2008, Frisch et al. 2004). Thus, the dissimilation pressure is not a general property of monoconsonantal prepositions but is specific to the prepositions [v] and [s]; whether this is for phonological reasons or due to lexical variation among the prepositions is hard to tell.

Gouskova 2012). Examples with prefixes include /v+zor/ [vzór] 'glance' and /v+xod/ [fxót] 'entrance'. The initial clusters in [vrét] 'harm', [ftórnik] 'Tuesday', and [vrémʲə] 'time' are not morphologically analyzable. As for /vz/- and /fs/-initial clusters, these all contain the /v-/ prefix. The probability of the [so] preposition is high even for words without shared place ([fxót], [fkús] and [fklʲut͡ʃénii]), but it is difficult to tell whether this is because they have a devoiced /v/ as their first segment and have a marked sonority profile or because they have a prefix that historically had a yer in it (a la the Lower analysis described in Section 2.4).



(a) s/so

(b) v/vo

Figure 1: Non-adjacent OCP between the preposition and the word-initial cluster. Error bars indicate 95% binomial confidence intervals. The category "both" includes cases in which one of the consonants in the word-initial cluster has the same manner and place of articulation as the preposition; in "place", only the place of articulation is shared; and in "none" neither place nor manner are shared. The dashed line represents the median CV rate within each category. Note that for [v/vo] "both" and "place" both have a higher median than "none", whereas for [s/so] "place" patterns with "none", and only "both" words show a higher CV rate.[3]

To ascertain that the differential OCP effects were statistically significant, we fitted mixed-effects logistic regression models to the set of words that robustly take the CV forms, using the R package *lme4* (Bates et al. 2012). We fitted two separate models, one for [v] and one for [s], with a three-level predictor for similarity: between the preposition and the following cluster no features in common (baseline level), shared major place, or shared place and manner.[4] We included a by-lemma random intercept in the models. The standard errors of the estimates were calculated using bootstrapping as implemented in the R library *boot* (Davison and

[3] Glosses: (a) 'entry', 'turning on', 'taste', 'harm', 'time', 'attention', 'yearning', 'mercury', 'a large amount/math set', 'cross', 'Tuesday', 'fleet', 'yard', 'bottom', 'day', 'platoon', 'adult', 'explosion', 'glance/look', 'sigh', 'capture', 'bribe', 'gaze', 'coming from the opposite side'; (b) 'evil', 'health', 'temple', 'footprint', 'character', 'sin', 'city', 'wrath', 'darkness', 'salvation', 'humility', 'ash', 'flesh', 'darkness', 'a large amount', 'opinion', 'moment', 'mist', 'good/blessing', 'saint', 'evidence', 'priest', 'lion', 'palace', 'yard-diminutive', 'yard', 'color'.

[4] Recall from Section 2.1 that there are no sounds that share manner but not place with [v] and can form a cluster in Russian. For [s], the class of sounds that share manner but not major place would be the fricatives [x] and [f] (orthographically voiceless); there were not enough clusters of this kind in the corpus that appeared with [so] to allow us to model the variation.

Hinkley 1997, Canty and Ripley 2012). In each iteration of the bootstrap procedure, a random subset of the words was selected with replacement (i.e., the same word was occasionally included more than once). The logistic regression model was then fitted to this subset of the words. This process was repeated 10000 times, allowing us to estimate the distribution of the regression coefficients. Visual examination confirmed that the bootstrapped regression coefficients were normally distributed, justifying the use of the t distribution to calculate the significance levels of the regression coefficients. The statistical models confirm the graphical impression (Table 2). In the case of [s], similarity in both manner and place of articulation increased the likelihood of the CV form, but similarity in place only did not. In the case of [v], similarity in place of articulation was enough to increase the the likelihood of the CV form. Manner similarity did not give an additional boost to the CV form for [v].

Predictor	Coefficient	SE	t	p-value
Place similarity only	−0.38	0.26	-1.46	0.14
Place and manner	0.86	0.09	9.5	<0.001

(a) s/so

Predictor	Coefficient	SE	t	p-value
Place similarity only	1.95	0.32	6	<0.001
Place and manner	2.09	0.43	4.85	<0.001

(b) v/vo

Table 2: Similarity avoidance in words from the Russian National Corpus that are compatible with CV prepositions: logistic regression results. We used treatment coding of the categorical predictor, with the reference level being the absence of any shared feature between the preposition and the word-initial cluster.

2.3 Lexical factors

Although a number of phonological properties can be used to predict the form of the preposition, the alternation cannot be reduced to phonological factors alone. For example, (8) shows two words that start with the same cluster. The two CV rates are unlikely to be derived from the same underlying distribution for [mn]-initial words: the difference between the rows is highly significant ($p < 10^{-15}$, Fisher's exact test).

(8) Lexical asymmetries between [mn]-initial words

a.	s mnénijəm	98%	sa mnénijəm	2%	'with the opinion'	($n = 443$)
b.	s mnóz̧əstvəm	29%	sa mnóz̧əstvəm	71%	'with the set'	($n = 926$)

We discuss four properties of the lexical variation in this section. First, we identify what appears to be a systematic register effect (Section 2.3.1). Section 2.3.2 discusses lexical variation of more arbitrary character, including cases in which the exact same phonological form behaves differently depending on its meaning (homonyms). In Section 2.3.3, we discuss the role of morphological structure: some morphemes show different CV rates depending on morphosyntactic context and morphological structure. Finally, Section 2.3.4 describes cases of idiosyncratic CV-preference which is limited to the context of a specific idiom.

2.3.1 Register effect: ecclesiastical words

One context in which CV prepositions occur more often than would be expected from their phonological shape is in ecclesiastical use. The word [krést] 'cross' is one such example. The cluster [kr] has a steeply rising sonority profile, and accordingly [kr]-initial words such as [kréslə] 'armchair', [kravát^{j}] 'bed' and [krɨló] 'wing' always appear with C prepositions. At the same time, [krést] occurs a substantial number of times with CV forms, across cases:

(9) CV-preference in [krést] 'cross'

a.	k kristú	57%	kə kristú	43%	'towards the cross'	($n = 227$)
b.	s kristá	68%	sə kristá	32%	'from the cross'	($n = 162$)

The word [krést] 'cross' is not an isolated case. Other examples are given in (10):

(10) CV-preference in other ecclesiastical words

a.	k spaséniju	65%	kə spaséniju	35%	'towards salvation'	($n = 426$)
b.	g blágu	48%	ka blágu	52%	'towards the blessing'	($n = 179$)
c.	g grixú	47%	kə grixú	53%	'towards sin'	($n = 100$)

The unusually high CV rate in ecclesiastical words is likely related to the fact that the CV forms were historically the only forms of the prepositions (Vlasto 1986).[5] Words related to religion are more conservative and tend to maintain features that are no longer productive in the language. For example, the Modern Russian consonant [ɡ] was once pronounced as a continuant, [ɣ]. While this pronunciation variant is almost entirely obsolete in Standard Russian, it is still possible in some lexical items with sacral connotation, such as the interjections [ɣóspədi] 'Lord' and [jej bóɣu] 'I swear to god' (Timberlake 2004, p. 23, Gouskova 2012). The phoneme [l^{ʕ}] in Arabic is another example: it only appears in the word [ʔal^{ʕ}ˈl^{ʕ}aːh] 'god' (Ferguson 1956).

2.3.2 Lexical idiosyncrasy in homonyms

Homonyms present a clear case of lexical variation which is purely arbitrary from the phonological point of view. For example, [léf] can mean either 'a lion' or 'Lev (name)', as in the author Lev Tolstoy. When it means 'lion', the CV form [so] is used almost exclusively; when it means 'Lev' the form of the preposition shows considerably more variation:

(11) The two types of [léf]

s l^{j}vóm	4%	sa l^{j}vóm	96%	'with the lion'	($n = 85$)
s l^{j}vóm	55%	sa l^{j}vóm	45%	'with Lev'	($n = 73$)

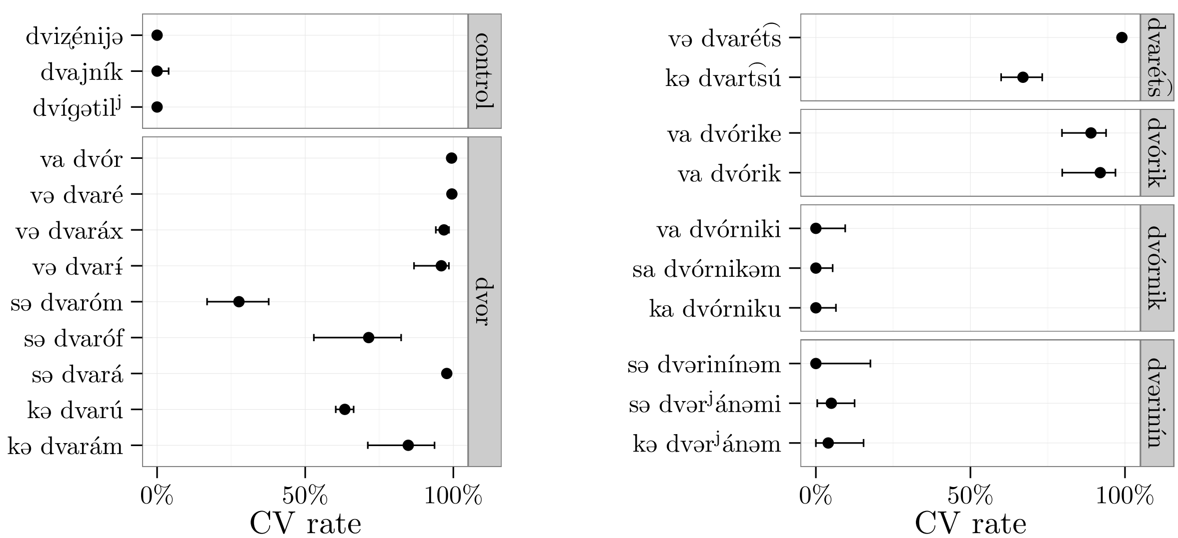
The difference in CV rate between two homophones proves conclusively that the alternation can be conditioned by lexical items, not just by phonological strings: "late" or postlexical phonology has no way to distinguish the two homophones (see Section 6.6 and Blumenfeld 2011).

[5] A reviewer points out that ecclesiastical words are more likely to occur in literal quotes from the Bible and older liturgical texts. This raises the possibility that at least some of the overall preference of these words for the CV forms reflects older language usage which is no longer productive, rather than a synchronic register effect, as we suggest. Informal corpus searches suggest that quotes from old texts do not account for all of the CV forms that occur with ecclesiastical words. In future work, the productivity of the CV-preference of ecclesiastical words can be assessed by collecting grammaticality judgments.

Another case of homonym-related variation is [mnóʐəstvə]. Gribanova (2009a) notes that [s mnóʐəstvəm], with the C form of the preposition, means ‘with a mathematical set’, while [sa mnóʐəstvəm] means ‘with a great number (of something)’. We found some support for this characterization, though it was not clear-cut. Out of the 340 occurrences of [s mnóʐəstvəm], 6 were used to mean “with a set”.[6] Out of the 561 occurrences of [sa mnóʐəstvəm], only one was used to mean “with a set”. This asymmetry is statistically significant ($p = 0.01$, Fisher’s exact test), but it is not the case that [s mnóʐəstvəm] has to refer to mathematical sets. Alternatively, this may be a difference in register: formal registers favor CV forms more than informal ones (recall the ecclesiastic register effect in Section 2.3.1).

2.3.3 The role of morphology

Many words maintain their idiosyncratic preference for the CV form across prepositions and throughout the morphological paradigm. For example, the lemma [dvór] strongly favors CV prepositions, throughout the paradigm (Figure 2a; see also Matushansky 2002, fn 40). This is despite the fact that it starts with the rising sonority cluster [dv], which does not normally condition CV prepositions; indeed, the words [dvígətilʲ] ‘engine’ and [dviʐénijə] ‘movement’ invariably appear with C prepositions.



(a) (b)

Figure 2: (a) Idiosyncratic CV-preference throughout the paradigm in [dvór] ‘yard’: Other lemmas beginning with [dv] virtually never occur with CV prepositions (the control items are collapsed across prepositions and cases), but [dvór] takes the CV prepositions fairly consistently (7 out of the 8 forms found in the corpus). The tendency towards the CV form is enhanced when the preposition is [v], in line with the non-adjacent OCP effect discussed in Section 2.2. (b) Behavior of words that are morphologically derived from [dvór]. The words [dvarét͡s] ‘palace’ and [dvórik] ‘small yard’ behave like the base form, taking the CV form exclusively after [v] and frequently after [k]. Conversely, the words [dvərinín] ‘nobleman’ and [dvórnik] ‘groundskeeper’, also derived from [dvor], almost always take C forms. This contrast shows that variation can be affected by suffixes.[7]

[6] Three of the hits came from the same text, Yaschenko (2002), which uses [s mnóʐəstvəm] twice and [sa mnóʐəstvəm] once. Unfortunately, the structure of the RNC does not allow for easy searches within each text.

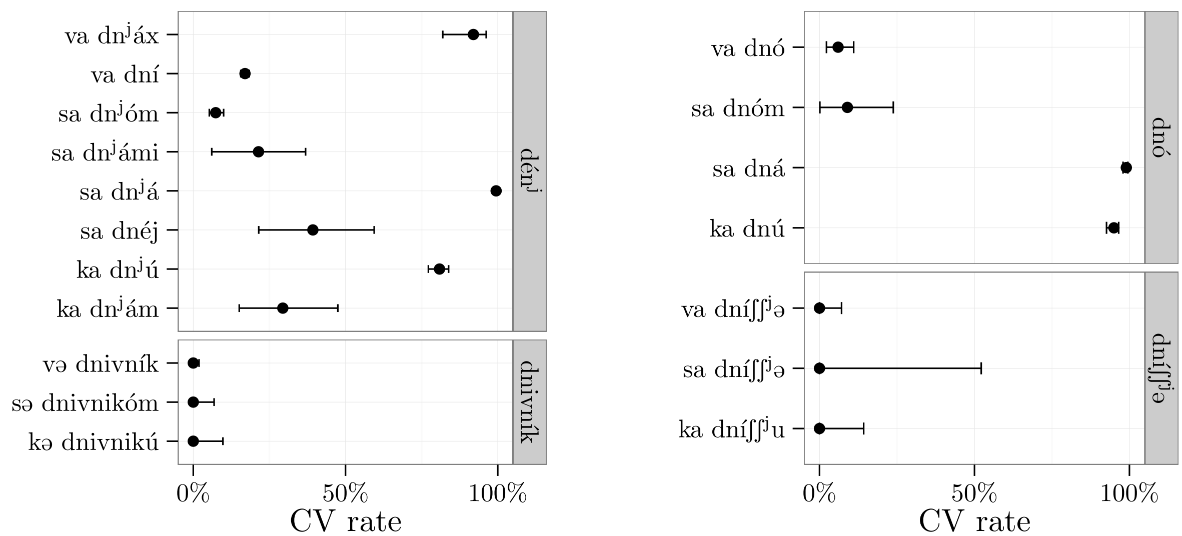
Most lexical items behave like [dvór] 'yard', in that they show fairly consistent CV rates across inflected forms. There are, however, cases of variable behavior within the inflectional paradigm of a word, which cannot be explained by the phonological properties of preposition-word junctures. This is particularly evident in the case of the two homophonous prepositions [s] in Russian, meaning 'with' and 'from', which take different case forms of nouns. Some nouns only show a CV-preference with one of the homophones of the preposition: [sa dnʲá] 'from the day' but [z dnʲóm] 'with the day' (Figure 3a). The word [dnó] 'bottom' behaves similarly (Figure 3b).

Upon closer examination, it turns out that the difference between [sa dnʲá] 'from the day' and [z dnʲóm] 'with the day' lies not in the homophone itself ('from' vs. 'day'), but in the sense of the word 'day' which is typically used with each homophone. The phrase [z dnʲóm] 'with the day' occurs primarily in salutatory expressions referring to holidays (as in, [pəzdravlʲáju z dnʲóm raʐdénijə] 'I wish you a happy birthday' or [z dnʲóm pabédɨ] 'happy Victory Day'). On the other hand, [sa dnʲá] 'from the day' usually refers to days in the general sense of the word (e.g., [sa dnʲa prinʲátijə kanstitút͡sii] 'from the day of the enactment of the constitution'). The holiday use of 'day' is not limited to 'with': of the 19 hits of [z dnʲá] 'from the day' in the RNC, 7 are used in the birthday/holiday context; there are non-holiday uses, but they come from older texts. Of the 50 uses of [sa dnʲóm] in the corpus, one is a quote from an Old Church Slavonic text, and many refer to an ecclesiastical holiday ([sa dnʲóm ángilə], 'with one's Name Day'). There are also uses in the birthday context. Thus, the holiday use of 'day' tends to favor the C preposition, whereas the non-holiday use mostly conditions CV prepositions.

There is a similar asymmetry between [z dnóm] 'with the bottom' and [sa dná] 'from the bottom'. Yet while this asymmetry is similar to the one in the 'day' case, there is no evidence that the two meanings of 'bottom' are different. Likewise, there is no reason to assume that the prepositions 'with' and 'from' are responsible for this distinction; we did not find overall any difference in CV rates between the 'with' and 'from' senses of [s] in the Yandex corpus study (see Section 3.2.5). It is possible that [dnó] 'bottom' is analogizing to [dénʲ] 'day', which is an order of magnitude more frequent in the corpus.

The effect of derivational morphology is equally heterogeneous, with derived words varying in the extent to which they inherit the CV preference of the base (Figure 2b). For instance, [dvór-ik] 'yard-diminutive' has a strong CV preference, whereas [dvór-nik] 'groundskeeper' does not. Certain suffixes (e.g. [-ik] 'diminutive') have a systematic effect on the CV preference of derived words, either maintaining it, reducing it or eliminating it. We do not discuss derivational context effects any further here for reasons of space; see Gouskova and Linzen (in preparation).

[7]Glosses: (a) 'movement', 'double', 'engine', 'into the yard', 'in the yard', 'in the yards', 'into the yards', 'with the yard', 'from the yards', 'from the yard', 'towards the yard', 'towards the yards'; (b) 'with the nobleman', 'with the noblemen', 'towards the noblemen', 'into the yard-diminutive', 'in the yard-diminutive', 'in the groundskeepers', 'with the groundskeeper', 'towards the groundskeeper', 'into the palace', 'towards the palace.'



(a) (b)

Figure 3: Variable CV rates within the paradigm: the two homophones [s] ‘with’ and [s] ‘from’ behave differently before both [dén^j] (dn^j-) ‘day’ and [dnó] ‘bottom’. The morphologically derived [dnivník] ‘diary’ (from [dén^j]) and [dníʃʃ^{j}e] ‘bottom of a ship’ (from [dnó]) always take C forms, as expected from their phonological form.[8]

2.3.4 Fixed expressions

Certain words show a CV-preference only in the context of a specific idiom (Blumenfeld 2011). This phenomenon appears to be particularly common in the ecclesiastical register. The word [práx] ‘ash’, for example, begins with a rising sonority cluster that almost always favors the C form of prepositions. For example, the word [pridlaẓénijəm] ‘proposal (inst)’, which starts with the same cluster, appears 956 times in the corpus with the C preposition [s], and never with the CV preposition [sə]. With most prepositions, [práx] too behaves as expected from its initial cluster, taking the C form. With the preposition [v] ‘in’, however, almost 50% of the tokens in the Russian National Corpus are the CV form [va]. A closer inspection reveals that most of these tokens are part of the fixed expression [liẓát^j va práxi] ‘be dead (lit. lie in ashes)’. The fixed expression [práx k práxu] ‘ashes to ashes’, with the preposition [k], does not behave in the same way (a search engine query turned up around 100 hits for [ka práxu], and a hundred thousand hits for [k práxu]).

A similar case is the word [plót^j] ‘flesh’, which appears with a CV preposition in the idiom [və platí] ‘in the flesh (figurative)’. One indication that this is an idiom with special phonology is the final stress on the suffix, which is not normal for the word (Zaliznjak 1977). The same lemma does not take CV prepositions in non-idiomatic phrases, such as [s plót^{j}ju] ‘with the flesh’, not *[sa plót^{j}ju]. Thus, unlike the two uses of [dn^j-] ‘day’ discussed in Section 2.3.3, the two allomorphs of the root [plót^j] ‘flesh’ are not even fully homophonous: they have different properties with respect to stress and to the preposition phonology.

[8] Glosses: (a) ‘in days’, ‘into days’, ‘with the day’, ‘with days’, ‘from the day’, ‘from days’, ‘towards the day’, ‘towards days’, ‘into the diary’, ‘with the diary’, ‘towards the diary’; (b) ‘into the bottom’, ‘with the bottom’, ‘from the bottom’, ‘towards the bottom’, ‘into the ship’s bottom’, ‘from the ship’s bottom’, ‘towards the ship’s bottom’.

2.4 Sonority and lexical idiosyncrasy in monosyllabic "yer" stems

Vowel-zero alternations in Russian prepositions are sometimes analyzed as a special case of a more general rule known as *yer* (jer) deletion (Lightner 1965, 1972, Halle 1973, Pesetsky 1979, Melvold 1989, Farina 1991, Yearley 1995, and many others). In the first part of the section (2.4.1), we review the aspects of the rule that are relevant to the prepositional alternations, and the predictions of the yer deletion account for monosyllabic words that contain a yer vowel. We contrast the yer deletion account with an alternative account that attributes the behavior of monosyllabic yer words to phonotactic constraints and lexical variation. We then present corpus data supporting the phonotactics-plus-lexical-variation account (2.4.2).

2.4.1 "Yers" and prepositions

In addition to the three prepositions discussed in this paper, Russian has lexically restricted vowel-zero alternations in many other morphemes. The vowel in [l^{j}ón] 'linen (nom)', for example, deletes when a vowel-initial affix is added in [l^{j}n-á] 'linen (gen)'. The alternation is analyzed as deletion because the backness of the vowel is not fully predictable; it is either [o] or [e] (Yearley 1995). The vast majority of Russian nouns do not have deletion (Gouskova and Becker to appear). In (12) we list a few examples of monosyllabic yer nouns, and in (13) we list similar nouns that do not exhibit such alternations.

(12) Examples of monosyllabic yer nouns

	Nom sg	Gen pl	Gloss		Nom sg	Gen pl	Gloss
a.	léf	l^{j}v-óf	'lion'	c.	ṣóf	ṣv-óf	'seam'
b.	róf	rv-óf	'ditch'	d.	l^{j}ón	l^{j}n-óf	'linen'

(13) Some similar monosyllabic nouns that do not alternate

	Nom sg	Gen pl	Gloss		Nom sg	Gen pl	Gloss
a.	lóf	lóv-əf	'catch'	c.	ṣéf	ṣéf-əf	'chief'
b.	r^{j}óf	r^{j}óv-əf	'roar'	d.	lés	lis-óf	'forest'

The dominant phonological analysis of yer deletion (Lightner 1972, Pesetsky 1979, Kenstowicz and Rubach 1987, Melvold 1989, Scheer 2006) marks alternating vowels underlyingly as different from regular vowels and posits a rule called Lower, which realizes a yer if it is followed by another yer in the underlying representation. A subsequent rule deletes all other yers. The nominative case suffix, null in masculine nouns, is a yer underlyingly (written as O in (14)); it triggers the realization of the stem yer and then deletes. Non-yer suffixes such as the genitive plural condition yet deletion.

(14) Traditional analysis: yers are only realized if followed by other yers (Lower rule)

a. /rOt + O/ $\rightarrow$ rót 'mouth (nom sg)'
b. /rOt + ov/ $\rightarrow$ rt-óf 'mouth (gen pl)'

If prepositional vowel-zero alternations are part of the same pattern, this analysis predicts that the prepositional vowel should delete when the following syllable has a non-alternating vowel, but it should be realized if the following vowel is also a yer. There are indeed some examples where this prediction is confirmed: words such as [rót] 'mouth' condition CV forms of alternating prepositions:[9]

(15) Yer realization in prepositions

a.	/sO + rOt + a/	→	sa rtá	'from the mouth'
b.	/sO + rabót + ɨ/	→	s rabótɨ	'from work'

As we have already seen, however, the prepositional vowels show up before nouns that do not have an underlying yer, either for phonotactic reasons or because of arbitrary lexical idiosyncrasy (Matushansky 2002). The influence of phonotactics is illustrated by the fact the putative prepositional yer is occasionally realized to prevent a geminate ([s] + [s]). Some examples of lexical variation are pointed out by Steriopolo (2007) and Blumenfeld (2011): there are yer monosyllables that condition C prepositions (e.g., [s ps-óm] 'with the dog', [k pn^j-ú] 'towards the stump'), and there are non-yer words that condition CV prepositions (Steriopolo cites [k rtúti] 'toward mercury' but notes that there is a [ka rtúti] variant). The phonotactic and lexical influences on the alternation in monosyllabic nouns suggest that the vowel-zero alternation in prepositions is distinct from yer alternations in nouns.

2.4.2 Corpus data

The two accounts outlined in the previous section make different predictions for monosyllabic yer words: the Lower analysis predicts that prepositions followed by yer words should always be CV, whereas the phonotactics-plus-lexical-variation analysis predicts that the prepositions' shape will depend on the following cluster. To test these predictions, we considered the preposition [ko] 'towards'; the results are shown in Figure 4. This figure plots the distribution of [ko] before ten monosyllabic yer words in the Russian National Corpus. We chose the preposition [k]/[ko] because it does not exhibit OCP effects (Matushansky 2002, Steriopolo 2007). The only phonological factor that should affect the choice of prepositional form is the shape of the word-initial cluster—specifically, its sonority profile (see Section 2.1.3 for the sonority scale used here, and Sections 2.2 and 3.2.4 for the role of sonority in the alternation).

Since the sample size is small, we lifted the restriction that a form must appear more than 20 times in the corpus. As shown in Figure 4, monosyllabic yer words indeed often appear with CV prepositions, though to different extents. For example, the form *[k rtú] 'towards the mouth' is categorically disallowed, the two forms [g zlú] and [ka zlú] 'towards evil' are equally frequent, and the C form [k pn^{j}ú] 'towards the stump' is much more common than [ka pn^{j}ú]. Words in our sample that started with a falling sonority cluster such as [rt] or [l^{j}d] were more likely to appear with the CV preposition [ko] than words starting with rising sonority cluster. High frequency words had a higher chance of being associated with CV prepositions; for example, the high frequency form [dn^{j}ú] 'day (dat sg)' had a higher CV rate than [zlú] 'evil (dat sg)' (medium frequency) and [psú] 'dog (dat sg)' (low frequency)—even though their sonority slopes are similar. For most of the words in the figure, the proportion of [ko] decreases as sonority slope increases—the outliers are [snú] 'sleep/dream' and [dn^{j}ú] 'day', which are also the highest-frequency words (frequency is shown via font size).

[9] In order to work for cases such as [rót], the analysis requires the assumption that the preposition and the following noun are phonologized in the same cycle. If the yer of the noun is deleted before the preposition is considered, the preposition will always lose its vowel. Matushansky (2002) proposes that the yer phonology of prepositions and nouns is resolved postcyclically. See Section 6.6 for discussion.

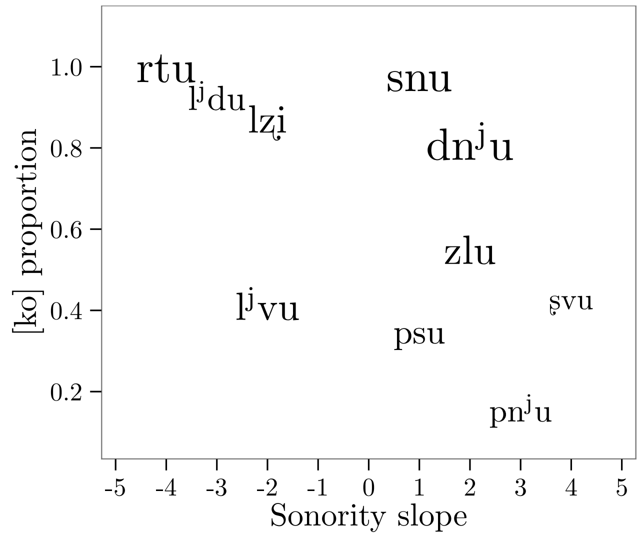


Figure 4: Sonority slope and frequency effects on the likelihood of the CV prepositions [ko] before monosyllabic yer words. Font size represents the number of tokens in the Russian National Corpus of the word after [k] or [ko] combined (on a logarithmic scale), ranging from 6 for [k/ko s̥vú] 'towards the seam' to 570 for [k/ko dnʲú] 'towards the day'.[10]

To assess the effects of sonority slope and frequency in monosyllabic yer words, we fitted a logistic regression model in R. As our frequency measure, we used the log-transformed number of tokens of the word with either form of the preposition, [k] or [ko]. Log-transformed frequency was somewhat correlated with sonority slope ($r = -0.37$). To reduce collinearity in the model, we regressed out sonority slope from frequency. Residualized frequency was highly correlated with frequency ($r = 0.92$), suggesting that it could be interpreted in a similar way to the original variable. The resulting logistic regression model indicated that both effects are highly significant. An inspection of the model residuals revealed three potential outliers: [lʲvú] 'lion (dat sg)', [lʲdú] 'ice (dat sg)' and [snú] 'sleep (dat sg)'. Removing these three items and refitting the model did not change the qualitative pattern: the CV form is more likely to be used when the sonority slope is smaller, and when the form is more frequent ($p < 0.001$ in both cases). The two factors interacted such that the sonority slope effect was larger for more frequent words ($p = 0.01$).

Monosyllabic yer words do not uniformly behave in the way predicted by Lower; the prepositional vowel occasionally deletes along with the noun's vowel. There are idiosyncrasies among these nouns as well. For instance, the root [sn-] (e.g. [ka snu]) appears with CV more often than would be expected from its phonotactics. The categorical preference of [ka rtú] over *[k rtú] may have been a productive pattern in older stages of the language, but it is not anymore, as indicated by the behavior of roots such as [s̥v-] and [ps-]. There is no evidence suggesting that monosyllabic yer words condition prepositional vowel alternations in a way that is qualitative different from other nouns: they condition CV forms based on their phonotactics and on lexically idiosyncratic CV-preference.

In summary, this section contrasted two views. According to one view, preposition vowels pattern like other yers: they appear when followed by a yer vowel underlyingly. According to the other view, the prepositional alternation before yer stems is governed by the same phonological generalizations as prepositions

[10] Glosses: [rtú] 'mouth (dat)', [lʲdú] 'ice (dat)', [lzí̥] 'lie (dat)', [lʲvú] 'lion (dat)', [snú] 'sleep (dat)', [dnʲú] 'day (dat)', [zlú] 'evil (dat)', [psú] 'dog (dat)', [s̥vú] 'seam (dat)', [pnʲu] 'stump (dat)'.

throughout the language, with potential lexical variation. The three facts described in this section—the gradience of the alternation, the interaction with phonological properties of the word-initial cluster, and the effect of word frequency—all suggest that the selection of the CV preposition is not a case of yer realization, which is a categorical phenomenon that applies across the board, but rather a combination of stochastic phonotactic rules and idiosyncratic properties of individual nouns.

3 Corpus Study 2: The [s/so] Alternation in Yandex

The results from Corpus Study 1 enabled us to discover new stochastic phonotatic rules and map the landscape of lexically conditioned variation. Due to the limited size of the Russian National Corpus, however, we were only able to find a relatively small sample of words that appear with CV prepositional forms at a non-negligible rate. In particular, many possible word-initial clusters were not represented in our sample at all. To gain access to a larger corpus of Russian texts, we used the web search engine Yandex (http://www.yandex.ru).

3.1 Methods

We generated a set of search queries using the methods described in Section 2.1. In contrast with the queries sent to the Russian National Corpus, we did not impose a pre-determined frequency threshold—since the corpus is several orders of magnitude larger, even rare forms are expected to occur with some frequency. The Russian National Corpus study showed that the alternation pattern before stop-initial clusters such as [pr] is mostly categorical and well understood, so we did not perform any queries for words starting with those clusters. Specifically, we restricted our queries to words with clusters that started with [v], [ʐ], [z], [l], [m], [n], [s], [f], [x], [t͡s], [t͡ʃ] and [ʂ]. For reasons of space and time, we only collected data for the preposition [s]/[so], for which we performed 86,136 search engine queries in total. The results can be downloaded as a CSV file from http://tallinzen.net/projects/russian_preps/.

The number of matches reported by search engines is known to be a rough and sometimes misleading estimate of the actual frequency of the form (Nunberg 2009). However, when the number of matches is small—in Google and Yandex, fewer than 1000—it is possible to get a more accurate frequency estimate by going through the results page by page and counting the number of matches. The results we report in this Section are therefore limited to inflected forms that appeared with both variants of the preposition less than 1000 times—if either variant crossed the 1000 result limit, we excluded the inflected form from our sample. In addition, we included only inflected forms that had more than 10 matches in total with either variant of the preposition.

The mean lemma frequency of the words included in this study was lower than in Corpus Study 1: RNC searches for the lemmas yielded 1120 tokens on average for the lemmas used in Corpus Study 2, compared to 14760 tokens for the lemmas used in Corpus Study 1.

3.2 Results

We explore the effect of a series of factors: the first consonant of the cluster, the position of stress, the case of the following noun (genitive or instrumental), the number of the following noun (singular or plural), and the sonority slope within the cluster. Search engine results are noisy, and the alternations are sensitive to lexical exceptions, so we expect to find a great deal of variability in many phonological contexts. For this reason, we do not use a single summary statistic, such as the mean CV rate for each context, but rather present richer plots which approximate the distribution of C vs. CV prepositions for each context.

The box plots in Figure 5 show the proportion of the [so] form of the preposition [s], broken down by the cluster-initial consonant of the following word. The horizontal line in the middle of each box represents the median CV rate for that particular cluster. The box extends between the first and third quartile of the data. The whiskers extend to the most extreme data point that is no more than 1.5 times the length of the box away from the box, and the dots represent "outliers", or points that are farther than 1.5 times the box's length. As the figure shows, the CV rate is more variable before some consonants than others. Clusters starting with non-sibilant obstruents (e.g., [t͡ʃr] or or [xv]) almost always condition the C form. There is more variability before sibilant-initial clusters, though the tendency is to favor the CV form. Finally, sonorant-initial clusters show a non-uniform pattern: the nasals [n] and [m] essentially pattern with non-sibilant obstruents in favoring the C form, whereas clusters that start with [v] and [l] show a great deal of variability.

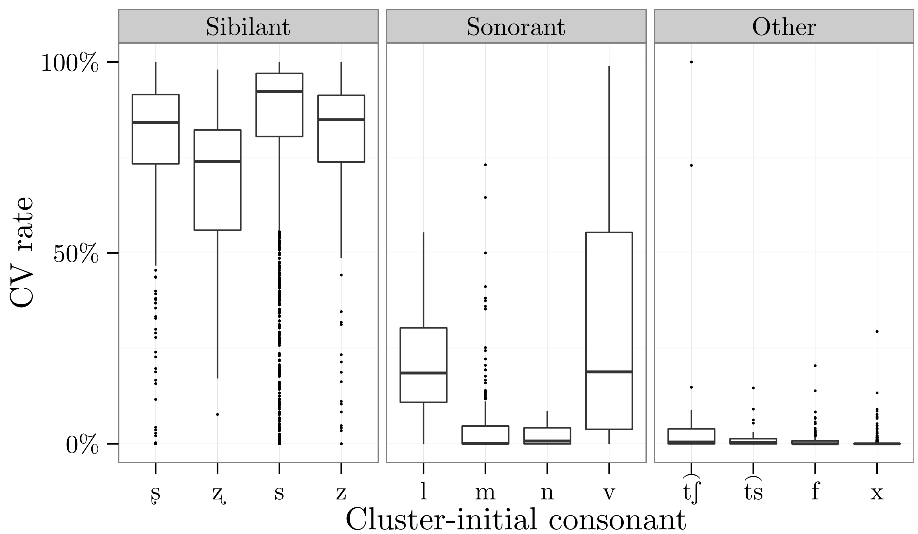


Figure 5: The probability of the CV form [so] in Yandex, by cluster-initial consonant.

3.2.1 Variation among sibilants

The next series of figures we present consists of kernel density plots, a rescaled and smoothed version of a histogram. Density plots approximate the probability density function of the population from which the data were sampled. Since density plots represent probabilities rather than raw counts, they facilitate comparisons across groups with different sample sizes. Figure 6 shows how CV rates vary among sibilant-initial clusters. In [s]-initial clusters, the most common probability of the CV form is over 95%: for most words that start with an [s]-initial cluster, the C form is quite rare. The CV rate decreases to around 90% for both [ʂ] and [z], which differ from [s] by one feature each (anteriority and voicing respectively). Finally, the sibilant [ʐ], which differs from [s] in both anteriority and voicing, shows an even lower typical CV rate of 80%. This suggests that the OCP, which encourages the presence of a vowel before sibilants, is stronger the more similar the preposition is to the consonant following it, in line with the results presented in Section 2.2.

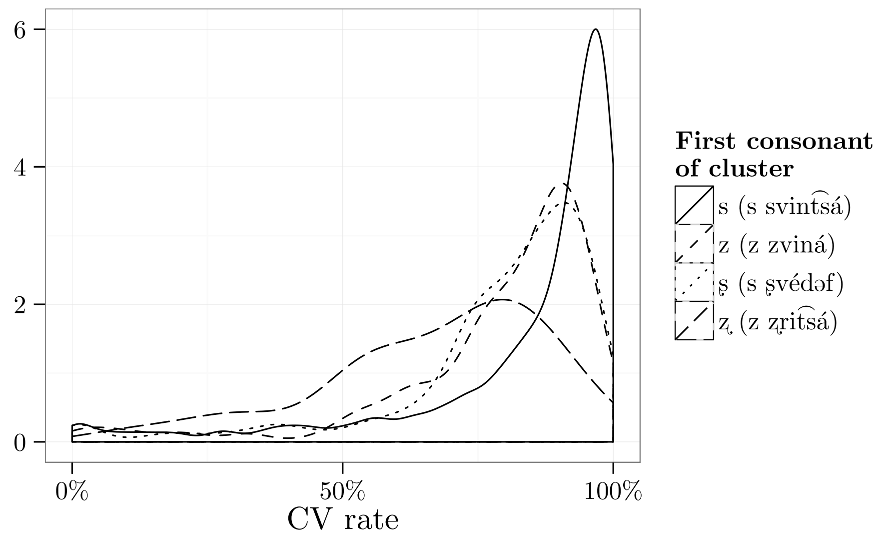


Figure 6: The probability of the CV form [so] before sibilant-initial clusters increases the more features are shared between that sibilant and [s]. (Glosses: ‘from lead’, ‘from a link’, ‘from Swedes’, ‘from a priest’)

This gradient similarity avoidance effect is not unique to Russian. Pająk and Bakovic (2010) find a similar effect of shared features for the Polish preposition [z], the etymological counterpart of the Russian [s]: when the preposition consonant and the following consonant share manner and sibilance, the CV form is more likely than when only sibilance is shared.

3.2.2 Stress

To investigate the remaining factors, we focused on words starting with [v]-initial clusters, since CV rates varied the most in this subset (Figure 5), and we had a relatively large number of such words in our sample (420 types).

We found that the position of stress in the noun interacts with the prepositional alternations. Figure 7 shows the probability of [so] before words stressed on the first syllable (solid line) and words stressed elsewhere (dashed line). The preference for the CV form is stronger in stress-initial words (i.e., [sa vdóxə] ‘from an inhale’) than in words with non-initial stress (e.g., [sə vdaft͡sá] ‘from a widower’). We were not able to find clear differences between words stressed on different non-initial syllables. This may be due to a theoretically significant difference between the first syllables and all the others, or due to the small number of items in many of the categories (for example, there were very few words that were stressed on the fourth syllable).

This effect of stress position is unexpected. At least for Moscow Russian, the preference could be restated as follows: [sa $\acute{\sigma}$] is more likely than [sə $\breve{\sigma}$]. Other dialects of Russian differ in the details of their vowel reduction patterns, but it is common for the immediately pretonic vowel to be more prominent—i.e., longer, less centralized, or bearing a high tone—than vowels in other positions (Crosswhite 1999, Padgett and Tabain 2005, Bethin 2006). Our analysis in Section 5 explains this pattern in terms of prominence: the preposition vowel is more likely to be retained when it can be parsed into an iambic foot.

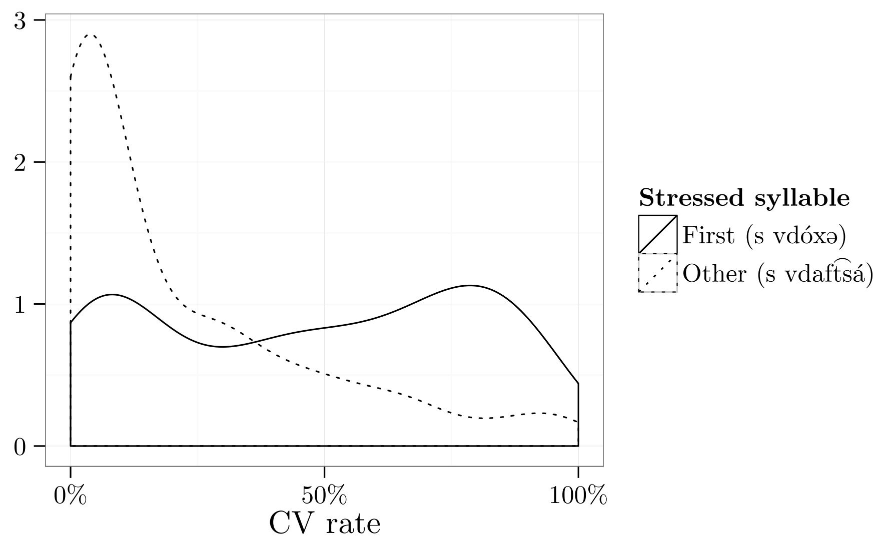


Figure 7: Words stressed on the first syllable are more likely to condition the CV form than words stressed on other syllables. (Glosses: 'from an inhalation', 'from a widower')

3.2.3 Non-adjacent OCP

We next inspect how the alternation is affected by the consonant that follows [v] in the word-initial cluster. In what follows, we abbreviate clusters of [v] with sibilants, sonorants, and the non-sibilant obstruents available in the sample as vS, vR and vK, respectively (see Table 3). Figure 8 shows that vS words are considerably more likely to condition the CV form than vK words. This is an OCP effect between non-adjacent consonants: selecting the C form of the preposition [s] before a word that starts with the cluster [vz] would create the cluster [svz], which has two sibilants, albeit in non-adjacent positions. This problem does not arise when the resulting cluster would be [svd].

vS	v + sibilant	s, z, ʂ, ʐ
vR	v + sonorant	r, l, m, n
vK	v + obstruent (non-sibilant)	b, d, k, p, t, x

Table 3: Notation used for vC clusters

A potential alternative explanation for the dispreference for [s+vS] clusters is that it is an instance of a more general pressure against multiple continuants in the same cluster, following the observation that clusters with similar elements are generally dispreferred (Ohala and Kawasaki-Fukumori 1997). There are three non-sibilant consonants in Russian that could be considered fricatives: [v], [f] and [x]. We set [v] aside because it is known to be ambiguous between a sonorant and an obstruent (recall section 2.1.3); [vf] is not a possible cluster, so that leaves [fx]. There is only one lemma in our data set that starts with this cluster: [fxaʐdénijə] 'entry'. Its CV rate is fairly low (12%), arguing against a dissimilation pressure for fricatives in general. Of course, a strong conclusion would be unwarranted based on a single example. At the very least, however, we have evidence that the dissimilation pressure is stronger between identical fricatives than between non-identical ones: the sequence [v+dv] is worse than [z+dv], at least in [dvor] and words derived from it (see Figure 2). The same example also indicates that the two sibilants need not be adjacent for the OCP effect to hold.

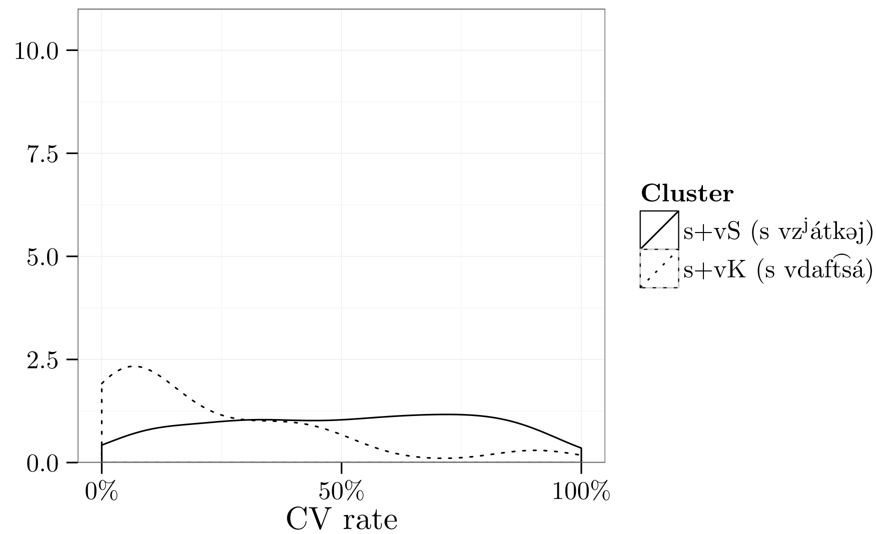


Figure 8: Non-adjacent OCP effect on the [s]/[so] alternation before [v]-initial clusters in Yandex: vS clusters condition the CV form more often than vK clusters (vS: v + sibilant; vK: v + non-sibilant obstruent). (Glosses: 'with a bribe', 'with a widower')

3.2.4 Sonority slope

The last phonological factor we consider in the Yandex corpus is the sonority of the consonant following [v]. As shown in Figure 9, [so] is more likely the less sonorous the second consonant in the cluster. In other words, [svC] clusters are more strongly dispreferred when there is a steeper drop in sonority between [v] and the following consonant. Words that start with a vR cluster, such as [vridámi] 'damages (inst)', almost always appear with [s]. On the other hand, there is more variation among words that start with a vK cluster, such as [vdaf͡tsá] 'widower (gen)', with some of them occasionally selecting the CV form. The difference between vK and vR clusters is smaller than the difference between vS and vK clusters, however. The differences among the sonorants [r, l, m, n] were not large (see the statistical analysis in Section 3.3 below).

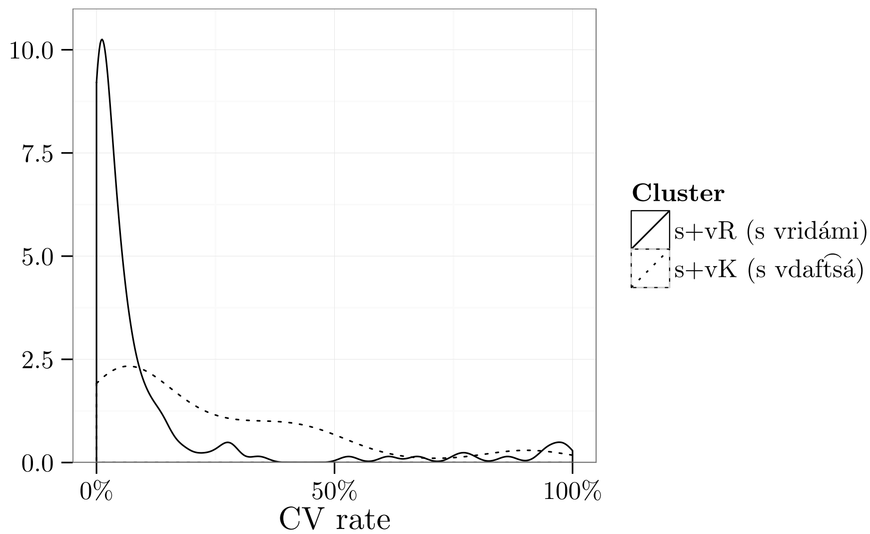
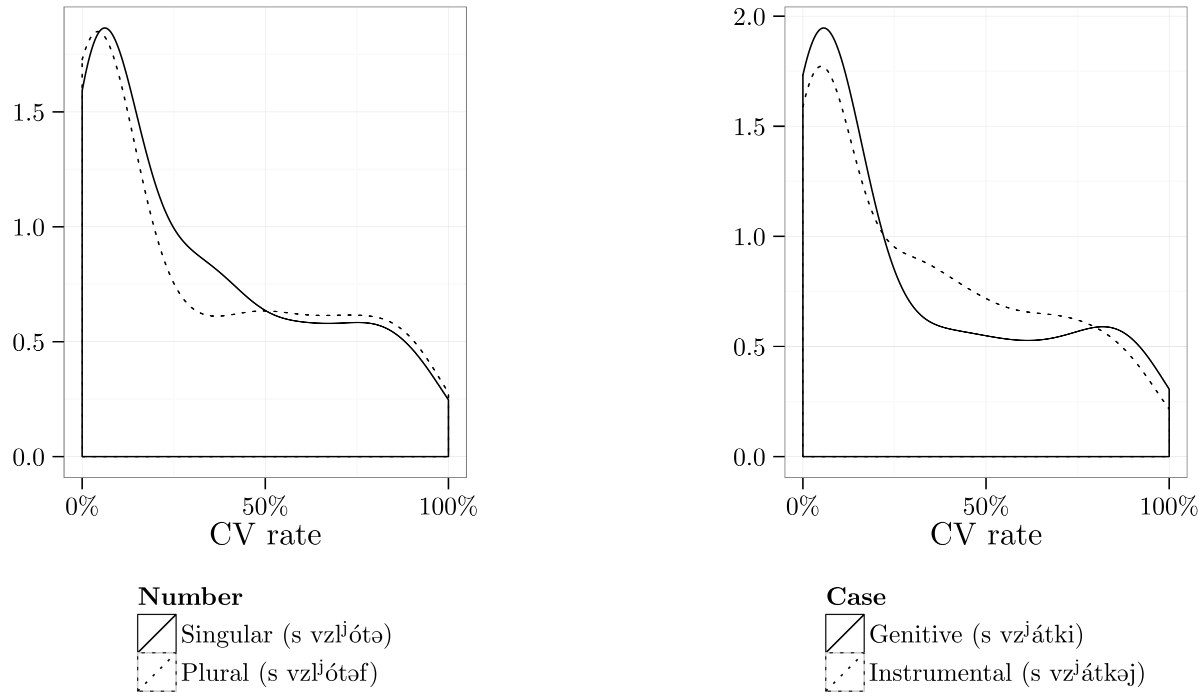


Figure 9: Sonority slope effect on the [s]/[so] alternation before [v]-initial clusters: vK clusters condition the CV form more often than vR clusters (vK: v + non-sibilant obstruent; vR: v + sonorant). (Glosses: 'with harms', 'with a widower')

3.2.5 Morphology

Recall from Section 2.3.3 that we found some differences between inflected forms in the same paradigm—for example, [z dn^{j}óm] 'with the day (inst)' vs. [sa dn^{j}á] 'from the day (gen)'. A potential explanation for this difference is that the two homophonous prepositions [s] 'with' and [s] 'from' condition the alternation differently. We looked for similar morphological differences in the Yandex corpus. Figure 10 compares forms in the genitive and instrumental cases. The difference in CV rate of [s] between the two grammatical cases is very small. This suggests that the contrast between 'with the day' and 'from the day' cannot be attributed to a difference in the overall behavior of the two homophonous prepositions. Similarly, CV rates do not seem to vary between singular and plural nouns.



(a) (b)

Figure 10: Effect of case and number on the [s]/[so] alternation before [v]-inital clusters. (Glosses: (a) 'with a takeoff', 'with takeoffs'; (b) 'from a bribe', 'with a bribe')

3.3 Statistical analysis

We analyzed the behavior of the [s]/[so] alternation before [v]-initial clusters ($n = 420$) using a logistic regression model. The outcome variable was the proportion of the CV form. Based on our exploratory investigation, we included two phonological variables of interest: stress position (word-initial or elsewhere) and the type of consonant following [v] (vR, vS or vK). We included two phonological control variables: the number of vowels in the word (which is identical to the number of syllables), and the number of consonants in the word-initial cluster (e.g. [fsV] compared to [fstV]).[11] In addition, we included three non-phonological variables: case (instrumental or genitive), number (singular or plural) and log-transformed lemma frequency (based on the Russian National Corpus).

The results of the regression are shown in Table 4. Factors that increased the likelihood of the CV form appear with a positive regression coefficient. We found that both of the main phonological predictors of interest had a significant effect on the alternation. Compared to the vK base level, vS clusters conditioned the CV form more often. This non-adjacent OCP effect was extremely robust ($t = 6.25, p < 0.001$), confirming the qualitative pattern shown in Figure 8 above. The CV form was also less common before vR than vK clusters ($t = -2.17, p = 0.03$), consistent with a sonority slope effect, whereby the C form is discouraged before falling sonority clusters.

Non-initial stress decreased the likelihood of the CV form. This effect was again highly significant

[11] We thank an anonymous reviewer for suggesting the two phonological control variables.

($t = -2.94, p = 0.003$). We also found that longer words were less likely to condition the CV form ($t = -3.44, p < 0.001$). This effect is partially correlated with the effect of stress, since short words are naturally more likely to have stress on their first syllable than long words; however, follow-up stepwise regression analyses suggest that the two effects are independent. Stress position was also correlated with cluster size in our sample: words beginning with longer clusters were more likely to be stress-initial (in line with Ryan to appear). Again, we verified that the effect of stress persisted after controlling for cluster size. Log lemma frequency, case, number and cluster size did not reach significance.

In a follow-up analysis, we divided the class of sonorants into nasals ([m] and [n]) and liquids ([l] and [r]). This analysis did not reveal any difference between the two types of sonorants ($t = 0.1$, n.s.). The absence of a difference between the two types of sonorants may be due to the small overall effect of sonority slope: even collapsing across the two classes of sonorants it was only marginally significant. Alternatively, the lack of significant difference may indicate that speakers only distinguish obstruents from sonorants and do not make finer distinctions within the class of sonorants, as has been suggested for Slavic (Bethin 1992 and others).

Predictor	Coefficient	Standard Error	t	p-value	
Second is sibilant (vS)	1.5	0.24	6.25	<0.001	***
Second is sonorant (vR)	−0.85	0.39	−2.17	0.03	*
Non-initial stress	−0.71	0.24	−2.94	0.003	**
Number of vowels	−0.21	0.06	−3.44	<0.001	***
Cluster size	−0.15	0.14	−1.04	0.3	
Log lemma frequency	−0.06	0.06	−1.03	0.3	
Case (inst. > gen.)	−0.12	0.25	−0.46	0.6	
Number (pl. > sg.)	−0.26	0.2	−1.29	0.2	

Table 4: Logistic regression results for the [s]/[so] alternation before [v]-initial clusters in Yandex. The baseline level is [vK]; predictors with positive regression coefficients increase the likelihood of the CV form [so].

In conclusion, the Yandex corpus study provides more quantitatively robust evidence for the phonological factors that contribute to the Russian prepositional alternation; see Table 1 for an overview of the results of the two corpus studies.

4 Nonce Word Experiment

Do the phonological pattern we identified constitute a productive phonological grammar? To address this question, we designed a nonce word experiment. The experiment was conducted in the auditory modality to ascertain that the patterns we found in the orthographic corpora extend to the spoken language. We set out to assess the productivity of the following generalizations:

(16) Generalizations tested in the experiment

- *Adjacent OCP:* The CV form is strongly favored when the first consonant in the following word agrees in place and manner of articulation with the preposition, even when it is not identical to it: [sa ʂrímәm] ≻ [s ʂrímәm].
- *Non-adjacent OCP:* The CV form is favored when the *second* consonant in the following cluster agrees in place and manner of articulation with the preposition: [sa fsímәm] ≻ [s fsímәm].

- *Stress:* Stress-initial words favor the CV form more than stress-final words: [sa fsívəm] ≻ [sə fsivóm].

- *Sonority slope:* Falling sonority clusters favor the CV form more than rising sonority clusters, for example: [sa rdíməm] ≻ [sa dríməm]. Likewise, if the first consonant in the cluster is held constant (e.g. [r]), the CV form is more acceptable when the second consonant is an obstruent than when it is a sonorant: [sa rdíməm] ≻ [sa rníməm].

4.1 Methods

4.1.1 Participants

There were 86 participants in the study.[12] They participated anonymously and were not compensated for taking the experiment. Participants were recruited online through various Russian language online communities and groups on social networks.[13] We assume they were native speakers of Russian, since we asked only native speakers to participate.

The participants volunteered the following information after completing the experiment: gender (53 females, 33 males), age (range 19–60, mean of 28.6, standard deviation 7.7, median 25), and location (18 from Moscow, 33 from St. Petersburg, 12 from other cities in Central Russia, and the remaining 23 from elsewhere in Russia and the former Soviet Union).

4.1.2 Materials

The list of test items included 80 monosyllabic nonce stems that started with the following 10 clusters: the v-initial clusters [vz], [vd], [vs] (pronounced as [fs]), and [vn], the sibilant-initial clusters [ʂr] and [ʂt], the fricative-initial cluster [xs], and the sonorant-initial clusters [rn], [rd] and [ms]. There were 8 items for each cluster, four of them with the vowel [u], and the remaining 4 with the vowel [i]. The rimes for the items that had [u] as their vowel were randomly selected out of following list: [uf], [ub], [up], [um], [un], [uk], [ug]. Similarly, the rimes for the items containing the vowel [i] were selected out of the following list: [ib], [id], [iz], [iv], [in], [ix], and [ig]. We chose high vowels because they show minimal phonetic variability based on stress position, as opposed to low and mid vowels, which undergo qualitative changes when stress is manipulated. In addition to the 80 experimental items mentioned above, the list included 8 control items starting with [xr] and 8 control items starting with [k].

The corpus studies did not yield any evidence that morphological factors such as case and agreement features influence the alternation (Section 3.2.5). To constrain the variability among the items to phonological rather than morphological factors, all of the nonce stems were therefore presented with the singular masculine instrumental/comitative case suffix [-om]. For each word-initial cluster, 4 items were stressed on the stem and 4 items were stressed on the suffix, evenly distributed between the vowels [i] and [u]. In total, the stimuli comprised 96 sentence pairs, contrasting in the presence of the vowel "o", pronounced as [ə] or [a] (see Appendix A).

The experimental items were divided into two versions, such that each participant was only presented with half of the items, to keep the experiment running time at around 10 minutes. Each version of the experiment contained exactly one item from each combination of the 12 word-initial consonant sequences (see above), 2 stress positions (final and initial) and 2 vowels ([i] and [u]), totaling 48 items per version.

[12] We collected data from 90 participants but excluded four from analysis because they did not report their age.
[13] The sites were http://odnoklassniki.ru, http://livejournal.com, http://vkontakte.ru, and http://facebook.com.

The nonce nouns, the frame sentences and the prepositions [s], [sə], and [sa] were recorded by a female native speaker of Moscow Russian. The recordings were made at a sampling rate of 44.1 kHz in a sound-attenuated booth on a Marantz PMD-660 solid state recorder using a head-mounted Audio Technica ATM75 microphone. Each nonce word audio file was concatenated with both [s] and the context-appropriate allomorph, either [sə] or [sa], depending on the position of stress in the nonce noun. The concatenation was done by script in Praat (Boersma and Weenink 2009); a Russian-speaking experimenter then edited the files manually so they would sound more natural. The intensity of all the sound files was normalized by Praat script. The files were converted to the MP3 format using the LAME encoder (http://lame.sourceforge.net). The audio files are posted at http://tallinzen.net/projects/russian_preps/.

4.1.3 Procedure

The experiment was conducted on the web, using the Flash platform (Adobe Systems, San Jose, CA). During the training phase, we presented the participants with written instructions in Russian that asked them to listen to an example of a nonce word in a context that did not require a preposition, and then two examples of the same word in a context that did require the preposition: one example with [s], and one with [sə/sa] (see Appendix B for an example of a trial). They were asked to choose which sound file they preferred. There were four training items, which the participants received feedback on. For example, they were told that most Russian listeners preferred [s tib-óm] to [sa tib-óm]. This was done to clarify the task to the listeners. The other training words were [(sa ≻ s) stíd-əm], [(s ≻ sa) vúm-əm], and [(sa ≻ s) zvúf-əm].

The nonce words were first presented to the participants in the context of one of the Type 1 frame sentences (see Appendix B). Type 1 frame sentences contained verbs that select for an object in the instrumental case, so as to familiarize the participants with the relevant form of the nonce word in a context where it is not preceded by a preposition. In addition to hearing the audio file, the participants were visually presented with the entire frame sentence (including the word) in the Russian orthography. Next, the nonce word was auditorily presented in the context of one of the Type 2 frame sentences, where it was preceded by the preposition [s]. The Type 2 sentence was played twice: once with the C form of the preposition ([s]) and once with the CV form ([sə/sa]). Only the first two words of the sentence (e.g. [ja pridú], 'I will come') were visually presented, followed by a blank in place of the prepositional phrase.

The order in which the items were presented was randomized in a way that no cluster appeared in more than two consecutive trials. Within each trial, the order of presentation of the two variants of the preposition was randomized, with the constraint that the sound file containing the variant [s] was never first one played in more than four consecutive trials.

4.2 Results

As expected, participants preferred the C form before [k]-initial controls almost categorically (>99%): [s kíbəm] ≻ [sa kíbəm]. These items were removed from further consideration. An inspection of by-item average responses to the cluster-initial items revealed that the item [xrix] was suspiciously different from other items beginning with [xr], including the minimally different [xriv]. For the three other [xr]-initial items, the C form was rated as more acceptable 90% of the time, as would be expected for a rising sonority cluster. By contrast, participants showed only a slight preference of 63% for the C form in [xrix]. A logistic mixed effects model with by-item and by-subject intercepts (see below) yielded a random intercept for this item that was 4.3 standard deviations from 0, indicating that it is an outlier. An inspection of the relevant sound file suggested that there was a recording artifact in the word-initial cluster [xr], which may have obscured the rising sonority of the cluster and caused uncertainty as to the preferred form of the preposition. We therefore excluded this item from further analysis. A similar test for participants did not reveal any obvious outliers (maximal difference of random intercept from the mean = 2.8 s.d.).

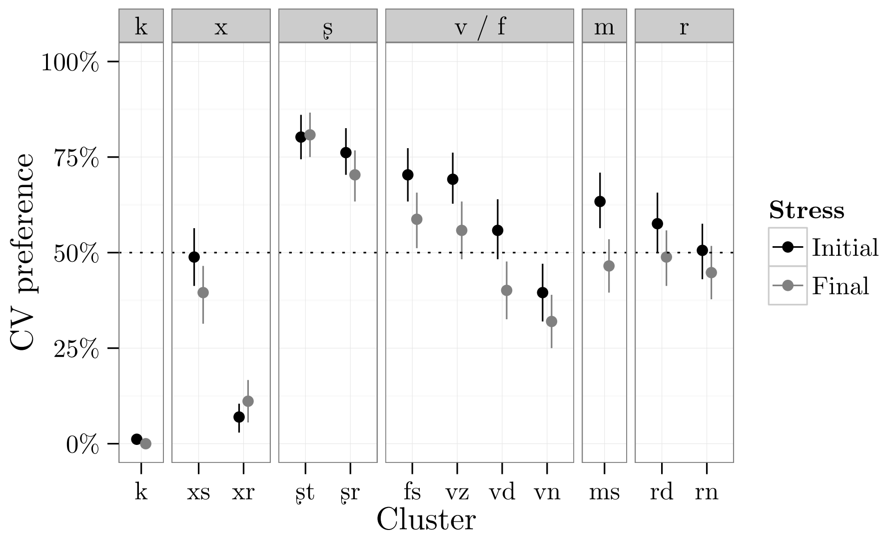


Figure 11: Experiment results: acceptability of [so+CC] broken down by CC cluster type and stress position. The dashed line at 50% indicates chance performance (guessing).

After removing outliers, we fitted a logistic mixed effects model in R using the *lme4* package. The binary response variable was 1 if the CV form was preferred and 0 if the C form was preferred. Stress and cluster were entered as fixed effects, with treatment (dummy) coding, such that the base level for stress was initial stress, and the base level for cluster was [fs]. We selected the maximal random effect structure justified by the data using model comparison, as follows: our base model included only a by-subject intercept. Adding a by-item intercept did not improve model fit ($p = 0.25$). A by-subject slope for the stress predictor did improve the model fit ($p < 0.001$) and was therefore included in the model. Due to the large number of clusters, we were not able to fit a by-subject slope for the cluster predictor, as the model fitting procedure did not converge. With the random effect structure of a by-subject intercept and a by-subject stress slope, the demographic variables did not improve the model fit (age: $p = 0.8$; sex: $p = 0.76$; geographic region: $p = 0.52$, model comparison).[14] The interaction between stress and cluster type did not reach significance ($p = 0.06$, model comparison) and was therefore not included in the model either. The marginal interaction was such that stress did not have any effect for the [xr]-initial items, before which there was a very strong preference for the C form; no difference in the effect of stress was found among any of the other clusters. The resulting model is presented in Table 5.

In what follows, comparisons between individual clusters were performed by resetting the base level of the cluster predictor and using the p-value corresponding to the resulting Wald statistic (z-value).

[14]The lack of an age effect contrasts with two previous reports. Pająk and Bakovic (2010) report an age effect for Polish vowel-zero alternations: younger speakers were less likely to use CV forms in their production study. Our results do not provide evidence for a similar change in progress in Russian. Steriopolo (2007) suggests that Russian speakers exhibit similar age differences; however, her report is based on a relatively small sample of two younger speakers and two older speakers.

Predictor	Coefficient	Standard error	z-score	p-value
Stress is final	−0.50	0.11	−4.52	<0.001
[ms]	−0.51	0.17	−2.95	<0.001
[rd]	−0.61	0.17	−3.48	0.003
[rn]	−0.90	0.17	−5.12	<0.001
[ʂr]	0.51	0.18	2.80	<0.001
[ʂt]	1.02	0.19	5.23	0.005
[vd]	−0.88	0.17	−5.04	<0.001
[vn]	−1.55	0.18	−8.59	<0.001
[vz]	−0.11	0.17	−0.63	0.520
[xr]	−3.81	0.26	−14.25	<0.001
[xs]	−1.08	0.17	−6.15	<0.001

Table 5: Experiment results: logistic mixed effects regression coefficients. Positive coefficients indicate higher acceptability of the CV form. Baseline level: initial stress and word initial cluster = [fs].

We confirmed the stress effect that we found in the Yandex corpus study (Section 3.2.2): participants preferred CV prepositions before words with initial stress ([sa vzíbəm] ≻ [sə vzibóm]). We also confirmed an OCP effect between adjacent sibilants: participants preferred the CV form before [ʂ]-initial clusters. There was a sonority slope effect: for clusters that started with the same consonant, the CV forms were more likely to be accepted when the second consonant was less sonorous. Thus, participants chose [so] more often before the falling sonority cluster [ʂt] than in the rising sonority cluster [ʂr] ($p = 0.01$). There was a small effect in the same direction in [r]-initial clusters: participants chose [so] before the steeply falling sonority cluster [rd] more often than before the moderately falling sonority [rn], though the effect was only marginally significant ($p = 0.09$). A similar effect of sonority slope can be seen in the comparison between [vd] and [vn]: [so] was more acceptable with falling sonority, and the effect was highly significant ($p < 0.001$). The effect of the sonority of the first consonant in the cluster can be seen in the three-way comparison between [fs], [ms] and [xs]: the sonority fall is steepest in [fs] (underlyingly and orthographically [vs]), milder in [ms] and flat in [xs], and accordingly the CV form becomes less acceptable ([ms] < [xs]: $p = 0.001$; [xs] < [fs]: $p < 0.001$).

Participants accepted [so] before [vz] more often than before both [vd] and [vn] ($p < 0.001$). Since the sonority fall in [vz] is milder than in [vd] and greater than in [vn], this is the opposite of what we would expect based on sonority alone. Thus, this pattern confirms an OCP effect between non-adjacent sibilants. The clusters [vz] and [fs] behaved identically ($p = 0.52$), which is consistent with treating the devoiced [f] in [fs] as a sonorant [v]. Since this is a null result, we do not want to make stronger conclusions here.

There was a sharp difference between [xr] and [xs]: [s] is overwhelmingly preferred for [xr] but much less so for [xs]. This could be due to a combination of two factors: sonority slope (rising in [xr], flat in [xs]), and non-adjacent OCP, disfavoring the cluster [s+xs].

The difference between [rn] and [vn] ($p < 0.001$) is in the opposite direction than would be expected based on sonority slope: since the sonority in [vn] is more steeply falling, we would expect the CV form to be more acceptable before [vn] than before [rn], which is not the pattern we see in the data. One possible explanation for this is that [v] is patterning both with sonorants and obstruents rather than just sonorants. The alternative explanation is that this unexpected pattern is due to technical issues with the recording of the [r]-initial cluster, pointed out by some of our participants. A third possible reason may be the fact that [r]-initial clusters are rare in Russian. The last two factors may have caused our participants to have less robust judgments about [r]-initial clusters than about [v]-initial clusters, causing the difference between [rn] and [rd] to be less pronounced than the difference between [vn] and [vd].

4.3 Summary

The experiment confirmed that the CV preposition [so] was more acceptable when the stress was on the initial syllable, when the sonority slope in the initial cluster was decreasing, and when there was a sibilant anywhere in the initial cluster (OCP). The OCP effect was stronger when the sibilant was the first consonant in the cluster than when it was the second one. For the most part, the constraints interact in a cumulative way: while the CV form is favored in sibilant-initial clusters across the board, this tendency is even stronger when the OCP constraint is amplified by initial stress and falling sonority.

5 Analysis: Stochastic Variation

The experiment established that several generalizations about Russian prepositional vowel-zero alternations are applied productively to novel words. The goal of the present section is to formally define the phonological grammar that speakers are applying. We model the stochastic and additive nature of the variation using MaxEnt (Goldwater and Johnson 2003, Hayes and Wilson 2008). We start with the basic phonological question of whether the vowel-zero alternation is deletion or insertion (Section 5.1). Section 5.2 defines the constraints we use in our analysis, and Section 5.3 presents the main features of MaxEnt and develops the analysis. One of the key features of the analysis is that the constraints themselves are fairly simple and similar to those proposed for other languages, but they can interact in a cumulative way to generate the complex trade-off patterns we found in the corpora and in the experiment, as discussed in Section 5.4. The analysis presented in the present section deals with phonological variation; we take up lexical variation in Section 6.

5.1 Deletion or epenthesis?

Vowel-zero alternations can be analyzed as either deletion or epenthesis. There are several arguments for both analyses in the case of the alternating Russian prepositions. On the epenthesis side, it has been observed that all Russian prepositions have vowel-final variants ([bés∼bézə̠] 'without', [t͡ʃéris∼t͡ʃérizə̠], [périt∼péridə̠], etc.—see Matushansky 2002, Gribanova 2009a).[15] The alternating vowels are always [+back], unlike yer vowels in lexical morphemes, which can be either back or front (see Section 2.4.1 for an introduction to yer vowels). This predictable quality would follow from an epenthetic account. Moreover, Gribanova (2008, fn. 5) notes that prepositions without alternations, such as [po] and [do], can appear phrase-finally, whereas C prepositions cannot. Yer words always surface with a vowel when no other vowel is available (Yearley 1995, Gouskova 2012), so if prepositions are part of the same deletion pattern as yers, they ought to surface to satisfy the requirement that all phonological words must have a vowel in Russian. An epenthetic account captures the observation that vowels tend to appear in prepositions only when they are needed to break up illicit clusters. Indeed, Pająk and Bakovic (2010) analyze analogous alternations in Polish prepositions as epenthesis. Hayes (2009) suggests that Polish vowel-zero alternations in other contexts are more compatible with an epenthesis analysis than with a deletion analysis because only one vowel, [ɛ], alternates productively (Czaykowska-Higgins 1988, Yearley 1995).

There are also arguments for deletion. In the Moscow dialect of Russian, the alternating vowel in the prepositions is [ə] or [a], and vowels of both qualities are typologically common epenthetic vowels (Kitto and

[15] We did not study these longer prepositions in any detail, but our impression is that these only have vowel-zero alternations when followed by function words (e.g., [bizə fséx] 'without all'; cf. (4)). Even when followed by [st] clusters, however, [bes] normally surfaces without a final vowel: [bis stakánə] 'without a glass'.

de Lacy 2000, Hall 2011a). In dialects with other vowel reduction patterns, however, the vowel may surface as [o], and this is a far less common epenthetic vowel, possibly unattested (Hall 2011b). Even in Moscow Russian, prepositions have [o] forms in fixed expressions with preposition stress, such as [só svitu] 'from the world' and [vó pəli] 'in a field' (Zaliznjak 1985, Ukiah 1998).

Our study presents another argument for deletion: we found that the CV forms of the prepositions are more likely to show up as [a] than [ə] (both in the Yandex study and in the experiment). This is easy to analyze in a deletion account: assuming iambic footing ($\sigma\acute{\sigma}$, Halle and Vergnaud 1987, Alderete 1999, Crosswhite 1999), the vowel is retained in a prominent footed position: /so vdóxa/ → [(savdó)xə] 'from an inhale'. It is deleted to improve the alignment of the lexical word and the phonological word, /so vdovet͡sa/→ [(svdaft͡sá)]. Alternatively, the more prominent vowel [a] is less likely to delete than a schwa, in a typologically common pattern (Steriade 2001, Gouskova 2003, Howe and Pulleyblank 2004). The epenthesis analysis could assume something like foot binarity as the driver of epenthesis, but then it would have to explain why this vowel shows up only in prepositions.

Another argument for deletion comes from the pattern of lexical variation in our corpus: conservative contexts, such as idioms and ecclesiastical words, tend to favor the CV form of the prepositions, not the C form. In fact, we were unable to find words that take the C form *less* than would be predicted by their phonological forms (e.g., a sibilant-initial word which only appears with [s] and never with [so]). The register effect fits with the situation in other languages if the alternation involves deletion: in this analysis, conservative registers are more faithful to the historical forms, whereas informal registers tend towards more deletion (see Coetzee and Kawahara in press, inter alia).

In summary, while we do not have a serious quarrel with the epenthetic analysis, the deletion analysis is simpler and more plausible historically and typologically, and we adopt it in what follows.

5.2 The constraints

This section presents the constraints we used to analyze the alternation. The syntactic distribution of prepositions suggests that they are clitics—they can attach to nouns, adjectives and adverbs (see Gouskova 2010, Padgett 2012 for a discussion of the phonological properties of prepositions, and Sekerina 1997, Fanselow and Ćavar 2002 on the syntactic behavior of prepositions in scrambling). In our analysis, clitichood is the reason why vowels delete in these prepositions: there is a requirement for the lexical word edge to coincide with the phonological word edge, discouraging word-internal syllabic prosodic clitics. We formalize this requirement as an alignment constraint:

(17) Align(L, Pwd, L, LxWd): 'The left edge of the phonological word coincides with the left edge of the lexical word edge.' Assign a violation mark for every syllable that stands between the left edge of a lexical word and the left edge of a phonological word. (After Selkirk 1995)

Align conflicts with Max-V, the familiar correspondence-theoretic faithfulness constraint. Vowels do not delete in Russian prepositions such as [po] 'along', [za] 'behind', and [u] 'by', so Align must have an effect in the grammar only when the syllable causing the misalignment contains the exponent of the prepositions [so], [vo] and [ko] (following the locality convention for indexed constraints introduced in Pater 2006). We assume that for non-alternating prepositions, Max-V outweighs Align (see Section 6.2).

(18) Max-V: 'Assign a violation mark for every vowel in S1 (input) that has no correspondent in S2 (output).' (McCarthy and Prince 1995)

Deletion is blocked by several markedness constraints on the resulting consonant clusters. One of these constraints is the Sonority Sequencing Principle (SSP). We will assume a gradient SSP cast in Harmonic Grammar, which assigns a violation mark of 0 to the best sequence of a stop followed by a glide, a mark of 1 to fricative-glide sequences and stop-rhotic sequences, a mark of 5 to clusters of stops with stops or nasals with nasals (e.g., [mn] or the fake geminate [nn]), and so on. The worst sequence gets a violation score of 10, as shown in (21). The sonority scale is repeated in (20) for reference.

(19) Sonority Sequencing Principle (SSP): Gradient HG formulation (Pater 2012; cf. Prince and Smolensky 1993/2004).

(20) Sonority scale:

$v_1 > r_2 > l_3 > n_4 > s_5 > t_6$

(21) Clusters and their violations of SSP:

10 {vt}; **9** {vs, rt}; **8** {vn, rs, lt}; **7** {vl, rn, ls, nt}; **6** {vr, rl, ln, ns, st}, **5** {vv, rr, ll, nn, ss, tt}; **4** {rv, lr, nl, sn, ts}, **3** {lv, nr, sl, tn}; **2** {nv, sr, tl}; **1** {sv, tr}; **0** {tv}.

To capture the fact that only cluster-initial words condition the CV prepositions, we define the blanket constraint (22) against triconsonantal clusters, which is blind to the particular phonological properties of the cluster. As discussed in detail in Section 6.2, indexing this constraint to individual morphemes allows us to best model the lexical variation, which is to some extent independent of the phonology of the cluster.

(22) *#CCC: 'Assign a violation mark for every sequence of three consonants.' (Cover constraint for *$[_\omega$C.CC and *$[_\omega$CC.C, see Gouskova 2012 and Steriopolo 2007.)

We now turn to similarity avoidance. Our constraints need to capture the following two facts:

1. Two sibilants are dispreferred within the same cluster.
2. Adjacent OCP is stronger than non-adjacent OCP ([ss̢r] is worse than [srs̢]).

To account for fact 1, we posit a general OCP constraint on sibilants within the same cluster, regardless of whether they are adjacent or separated by a consonant:

(23) OCP[sibilant] (*s(C)S): 'Assign a violation mark for every pair of consonants in a cluster that have an identical specification for [sibilant].' (penalizes [ss], [ss̢], [zz̢], [sCs], [sCs̢], etc. but not [sVs], [sVs̢]).

We account for fact 2 by introducing a specific constraint against adjacent sibilants, defined in (24). The dispreference for adjacent sibilants followed by a consonant (e.g. [sst]) is extremely strong. Preliminary MaxEnt modeling revealed that this preference is not adequately captured by the additive effects of *#CCC and a general constraint against two adjacent sibilants: either the dispreference is underestimated, or the general constraint against two adjacent sibilants is weighted very high, resulting in the undesirable penalization of CCV sequences such as [s+sa]. Since existing MaxEnt models do not learn multiplicative (superadditive) interaction terms, as would be necessary to obtain a dispreference for [sst] which is stronger than the combined dispreference for [ss] and [CCC], we manually combined the two conditions into a single constraint.

(24) *SimC-C: 'Assign a violation mark for every pair of adjacent sibilants if they are followed by another consonant.' (penalizes [sst], [zzg], etc. but not [zgz], [ssa], [kkr])

Since the experiment did not include the labial fricative preposition [v], we have only defined sibilant OCP constraints. A full set of constraints would have to account for the somewhat different behavior of labial OCP discussed in Section 2.2. Likewise, we do not introduce a gradient OCP constraint penalizing [s] + sibilant sequences in proportion to the featural overlap between [s] and that sibilant (see Sections 2.2 and 3.2.1), since we only had one type of sibilant in our experiment (ş).

Finally, we attribute the effect of stress to the constraint PARSE-σ, which is violated by unfooted syllables. We explain the role of this constraint in the analysis in the following section.

(25) PARSE-σ: 'Assign a violation mark to every syllable that does not belong to a foot.' (McCarthy and Prince 1993)

5.3 MaxEnt analysis of the nonce word ratings

A MaxEnt grammar works as follows. Constraints have positive numerical weights rather than rankings. Suppose there are m constraints and n candidates. The harmony H_j of the j-th candidate is determined by summing its violations for each constraint times the weight of the respective constraint, as in Harmonic Grammar (Goldwater and Johnson, 2003, Legendre et al., 1990, Smolensky and Legendre, 2006, Pater et al., 2007, Hayes and Wilson, 2008). Formally, if w_i is the weight of the i-th constraint and v_{ji} is the number of times that candidate j violates constraint i, then

$$H_j = \sum_{i=1}^{n} w_i v_{ji}$$

The probability of candidate j is then calculcated as the exponent of H_j, normalized by the sum of the exponents of all harmony scores:

$$p_j = \frac{e^{H_j}}{\sum_{k=1}^{m} e^{H_k}}$$

We fitted the constraint weights to the observed frequencies of outputs with the C and CV forms from our experiment using the MaxEnt Grammar Tool.[16] The weights are shown in Table 6, arranged from highest to lowest.

Constraint	Weight
ALIGN	6.24
*#CCC	4.26
*SIMC-C	1.72
*s(C)S	0.74
PARSE-σ	0.37
SSP	0.24
MAX	0.00

Table 6: MaxEnt constraint weights for experiment results

The rest of this section demonstrates how the resulting MaxEnt grammar derives the patterns we identified in the results of the experiment. We start with the simplest case, the near-categorical deletion in words that start with a single consonant. As shown in (26), ALIGN favors deletion in all prepositions: [sa kíbəm]

[16] http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/. We assumed the default priors mean $\mu = 0.0$, and standard deviation $\sigma^2 = 100000.0$ for the bias term.

violates ALIGN once, so its harmony score is the weight of ALIGN (6.24) times −1 for the violation. MAX is weighted at 0 for /so/, so it cannot block deletion. The MaxEnt grammar in Table 6 predicts the input /so kíb-om/ to map to [s kíbəm] 99.8% of the time (calculated as $p_1 = \frac{e^0}{e^0+e^{-6.24}} = 0.998$), compared to the 99% we observed in the experiment.

(26) Deletion very likely before simple onsets

/so kíb-om/	ALIGN	*#CCC	*SIMC-C	*s(C)S	PARSE-σ	SSP	MAX	$\mathcal{H}$	*Prob*
s kíbəm							−1 × 0	0	99.8%
sa kíbəm	−1×6.24							−6.24	0.2%

When deletion yields a cluster that violates several of the markedness constraints, the predicted outcome varies depending on the severity of the violations of the cluster constraints. In /so ṣtíb-om/ (see (27)), the resulting cluster violates *#CCC, the constraint against adjacent sibilants *SIMC-C, and *s(C)S, and it has a relatively poor sonority profile. The harmony of the deletion candidate [s ṣtíbəm] is a summation of its performance on *#CCC, the constraints on sibilants, and the SSP, which it violates six times. Thus, even though none of these constraints are weighted more than ALIGN, their cumulative action makes deletion a lot less probable. Deletion is predicted to be blocked 87.4% of the time. In our experiment, people preferred the CV form 80% of the time.

(27) Deletion unlikely if a #CCC cluster with a sibilant OCP violation and poor sonority profile results

/so ṣtíb-om/	ALIGN	*#CCC	*SIMC-C	*s(C)S	PARSE-σ	SSP	MAX	$\mathcal{H}$	*Prob*
sa ṣtíbəm	−1 × 6.24							−6.24	87.2%
s ṣtíbəm		−1 × 4.26	−1 × 1.72	−1 × 0.74		−6 × 0.24	−1 × 0	−8.16	12.8%

A slightly higher sonority rise makes deletion less likely: in /so ṣríb-om/, deletion is predicted only 72% of the time. This is shown in (28).

(28) Deletion more likely if the CCC cluster with an OCP violation has a better sonority profile

/so ṣríb-om/	ALIGN	*#CCC	*SIMC-C	*s(C)S	PARSE-σ	SSP	MAX	$\mathcal{H}$	*Prob*
sa ṣríbəm	−1 × 6.24							−6.24	72%
s ṣríbəm		−1 × 4.26	−1 × 1.72	−1 × 0.74		−2 × 0.24	−1 × 0	−7.2	28%

A sufficiently bad sonority profile will make deletion less likely even in words that have no sibilants near the preposition. Deleting the preposition's vowel in /so rdíb-om/ results in a #CCC cluster that violates the SSP nine times, which in tandem with the *#CCC violation is sufficient to override ALIGN 58% of the time.

(29) Deletion in words with clusters that have a marked sonority profile

/so rdíb-om/	ALIGN	*#CCC	*SIMC-C	*s(C)S	PARSE-σ	SSP	MAX	$\mathcal{H}$	*Prob*
sa rdíbəm	−1 × 6.24							−6.24	58%
s rdíbəm		−1 × 4.26				−9 × 0.24	−1 × 0	−6.42	42%

We capture the stress effect through PARSE-σ. Deletion is more likely when the vowel of the preposition is reduced to [ə], which we take to be due to its unfooted status (assuming iambic footing for Russian; see Alderete 1999, Crosswhite 1999, Halle and Vergnaud 1987, Gouskova 2010, though cf. Revithiadou 1999, Bethin 2006). As shown in (30), the retained vowel cannot be footed in [sə (ʂribóm)], so its deletion is encouraged more compared to the footed vowel in [(sa ʂrí)bəm] in (28).

(30) Stress effect: schwa deletes more readily than footable [a]

/so ʂrib-óm/	Align	*#CCC	*SimC-C	*s(C)S	Parse-σ	SSP	Max	$\mathcal{H}$	*Prob*
sə (ʂribóm)	−1 × 6.24				−1 × 0.37			−6.61	64%
s (ʂribóm)		−1 × 4.26	−1 × 1.72	−1 × 0.74		−2 × 0.24	−1 × 0	−7.20	36%

In clusters that have non-adjacent sibilants in the potential #CCC cluster, deletion is only weakly encouraged, since the weight of *s(C)S is smaller than that of *SimC-C. The violation scores of the two candidates for the input /so xsib-om/ are very close (see (31)), and the predicted frequencies (60% and 40%) match our experimental results.

(31) Triconsonantal clusters with non-adjacent sibilants

/so xsib-óm/	Align	*#CCC	*SimC-C	*s(C)S	Parse-σ	SSP	Max	$\mathcal{H}$	*Prob*
sə (xsibóm)	−1 × 6.24				−1 × 0.37			−6.61	40%
s (xsibóm)		−1 × 4.26		−1 × 0.74		−5 × 0.24	−1 × 0	−6.20	60%

Finally, we did not include /so sV.../ wugs in our experiment, because they pattern virtually categorically: they surface with the C form of the preposition. This result is generated in the MaxEnt grammar.

(32) No vowels in prepositions that precede sV-initial words

/so síb-om/	Align	*#CCC	*SimC-C	*s(C)S	Parse-σ	SSP	Max	$\mathcal{H}$	*Prob*
(sa sí)bəm	−1 × 6.24				−1 × 0.37			−6.61	0.03%
(s síbəm)				−1 × 0.74			−1 × 0	−0.74	99.7%

To sum up, this analysis allows us to capture the cumulative effects of the OCP, stress, and sonority sequencing through constraint weighting. Even though the weight of Align exceeds that of any other constraint here, the lower-weighted constraints can override its preference for deletion. The cumulative effect of these constraints is captured directly through the summing of their weights, and the frequencies of the outputs we got in the experiment are matched fairly closely.

Finally, we address an alternative analysis that we rejected for Russian. In their analysis of a similar pattern in Polish, Pająk and Bakovic (2010) argue that the CV preposition surfaces whenever the alternative is to assimilate sibilants or labials, which would result in a violation of the constraint they call NoGem/NVA 'no non-vowel-adjacent geminates'. The evidence for this analysis is that the alternative to variable C/CV alternation in Polish is voicing and coronal place assimilation, which applies in z+sV clusters but not in z+sCV clusters. Thus, either assimilation fails to apply or the vowel shows up to break up the clusters—a kind of conspiracy to avoid preconsonantal geminates. Russian as well has variable coronal place assimilation,[17] but this analysis would not work for the language. The problem is that in Russian vowels show up even in

[17] To our knowledge, the phonetics of assimilation in such clusters have not been systematically studied in Russian, though there is some experimental evidence for long-distance interactions between sibilants (Kochetov and Radišić 2009).

clusters that cannot assimilate completely to make a geminate: labial-labial-C clusters such as [vmn] and [fpr] are avoided, as well (recall Figure 1). There is also an interaction between non-adjacent sibilants, which cannot form a geminate across an intervening consonant (even if they could be argued to be articulatorily local, as in Gafos 1999). Russian also has a preposition Polish lacks, [k], which exhibits sonority effects but not antigemination (e.g., /ko ɡrob-u/ → [ɡ ɡrabú] 'towards a coffin'). While the place difference between dorsal stops and other consonants is essentially stipulated in our analysis, it is simply not the case in Russian that geminates before consonants are avoided across the board, or that preposition vowels appear only to avoid preconsonantal geminates.

Pająk and Bakovic (2010) and Bakovic (2005) do shed light on a feature of the Russian alternation that we do not analyze in detail, however: the gradient similarity effect among sibilants (see Fig. 6). The direction of the effect suggests that faithfulness is at play: assuming that there is at least some assimilation between the preposition consonant and the following sibilant, deletion is less likely when it implies having to make more changes to the preposition's consonant. It is additive in the way we predict: /so z̢rit͡som/ is less likely to have deletion than /so s̢vedov/ because [z/z̢ z̢rit͡sóm] requires more featural unfaithfulness than [s/s̢ s̢védəv]. We do not analyze this directly because our experiment tested only one sibilant, but the gradient similarity effect can be captured through the interaction of faithfulness to the features of the preposition's sibilant with the other markedness constraints violated by the resulting cluster.

5.4 Complex conditioning contexts for alternations as gang effects

The experimental results clearly show that the probability of the CV form is jointly determined by multiple constraints simultaneously rather than by the highest-ranked one. Even the near-categorical OCP constraint against adjacent sibilants does not overshadow the weaker sonority sequencing constraint: sibilant-initial clusters are more likely to occur with the CV preposition if they have a relatively bad sonority profile ([s̢t] occurs with CV more often than [s̢r] does). Similarly, stress on the following syllable makes a CV preposition slightly more likely across the board, on top of the phonotactic constraints related to the cluster.

Casting the analysis in a weighted constraint theory allows us to capture these additive effects while keeping the constraints themselves relatively simple compared to previous analyses in standard OT. For example, to explain why the OCP applies only to [ssC] clusters rather than [ssV], Steriopolo (2007) assumes that pre-consonantal sibilants are unsyllabified appendices rather than syllable onsets—a structure often posited for Slavic (Bethin 1992, Yearley 1995, and others). The constraint that blocks deletion in Steriopolo's analysis is the OCP locally conjoined with the constraint against unsyllabified consonants. This structural definition of the context for the OCP does not explain our findings, however. We found an OCP effect even with sibilant-liquid clusters such as [s̢r], and these are usually assumed to be tautosyllabic. The OCP applies across an intervening consonant (as in [s̲a m̲s̲íbəm]), where the second sibilant is syllabified with the following vowel. Our solution to this analytic problem is to keep the OCP constraint itself relatively simple, deriving the complex conditioning from its interaction with the constraint against three-consonant clusters. Neither three-consonant clusters nor OCP violations by themselves are sufficient to condition a high rate of CV prepositions; the effect arises through cumulative action. Such "gang" effects have been observed in other gradient phenomena (Keller 2000, Jäger and Rosenbach 2006, Hayes and Wilson 2008, Pater 2009, Potts et al. 2010).

As Boersma and Hayes (2001) point out, much of the success in modeling variation in constraint-based theories depends on the choice of constraints. A detailed comparison of constraint-based theories of variation would require comparing analyses that use different constraint sets and definitions, which would take us too far afield. We want to make one point that bears on the issue of how to analyze the stress effect in particular. In principle, gang effects do not require constraint weighting: some limited gang effects are possible in Stochastic OT (Boersma 1997, Boersma and Hayes 2001), and standard OT can capture certain types of gang effects through Local Conjunction (Smolensky 1995 et seq.). Local conjunction is a way to allow two low-ranked constraints to override a higher-ranked constraint when they are violated simultaneously: for example, in some languages, word-final consonants are allowed and voiced consonants are allowed, but word-final consonants cannot be voiced. Ito and Mester (2003) capture this by conjoining NoCoda and *VoicedObstruent in the domain of a segment; the resulting constraint is violated by segments that violate both constraints, and the conjoined constraint can compel unfaithful mappings that neither of the individual markedness constraints can compel. To extend this approach to the stress effect we found, one could conjoin Align with Parse-σ: the resulting constraint would be violated only by a syllable that is unfooted and causes misalignment of lexical and prosodic word edges, as in [sə (ṣribóm)]. But conjunction would not work with other possible accounts of the stress effect, such as a licensing-by-cue analysis (Steriade 2001). The idea behind such an analysis is that schwa is more likely to delete than [a] because its deletion is less noticeable.[18] Implementing this would require conjoining a constraint that specially protects [a] but not [ə] (Max-[a]) with each of the markedness constraints that disfavor deletion (*#CCC, *s(C)S, and so on). Recall tableaux (28) and (30): the goal is to make deletion slightly worse when it involves [a] than [ə], which means that Max-[a] would work in tandem with other constraints that disfavor deletion—all of them markedness constraints. The problem is that conjunction of Max with markedness is impossible: these constraints do not share a locus of violation (Moreton and Smolensky 2002). In a weighted constraint theory, on the other hand, Max-[a] can interact with markedness constraints through cumulativity, so this analysis can be implemented. In (33), we show the weights of the constraints and the harmony scores for two inputs that differ in stress position. The crucial interaction is in (33b), where the weight of Max-[a] is combined with that of the rest of the markedness constraints to make deletion less likely in pretonic position than in a position with greater reduction.

(33) A licensing-by-cue analysis in a weighted constraints model: [ə] vs. [a]

/so ṣríb-om/	Align 6.28	*#CCC 3.93	*SimC-C 1.72	*s(C)S 0.74	Max-[a] 0.37	SSP 0.24	Max-[ə] 0.0	$\mathcal{H}$	*Prob*
a. sa ṣríbəm	−1 × 6.28							−6.28	72%
b. s ṣríbəm		−1 × 3.93	−1 × 1.72	−1 × .74	−1 × .37	−2 × .24		−7.24	28%
/so ṣrib-óm/									
c. sə (ṣribóm)	−1 × 6.28							−6.28	64%
d. s (ṣribóm)		−1 × 3.93	−1 × 1.72	−1 × .74		−2 × .24	−1 × 0	−6.87	36%

If analyzed this way, the Russian stress effect supplies a possible candidate of a gang effect between markedness and faithfulness, in addition to markedness-markedness (Pater 2009, Potts et al. 2010) and faithfulness-faithfulness gang effects (Farris-Trimble 2008).

While the additive approach generally allowed us to obtain a good fit to the experimental results, it did not suffice in the case of the near-categorical OCP constraint against adjacent sibilants in clusters.

[18]The vowel that is deleting is not necessarily either [a] or [ə] underlyingly; it could be /o/, as we suggested in Section 5.1. To work technically, this analysis would have to include intermediate stages of representation where the prepositional vowels are present but reduced, as outlined by McCarthy (2011), with the vowel-specific Max constraints referring to reduced vowels rather than UR vowels.

As mentioned in Section 5.2, modeling this phenomenon as an additive effect of *#CCC and a general OCP constraint against adjacent sibilants (*SS) would require assigning a very large weight to *SS, thereby incorrectly predicting dissimilation in simple onsets (e.g. s + sa). To get around this problem, we posited the constraint *SimC-C, which is essentially a conjoined version of *SS and *#CCC. A potential alternative to hand-crafting conjoined constraints such as this one is expanding the expressive power of the MaxEnt model by allowing it to learn multiplicative interactions between constraints, such that the joint effect of two constraint is allowed to be larger than the sum of their individual effects.

6 Analysis: Lexical Variation

In our corpus studies, some words conditioned different rates of the CV prepositions even though they started with the same cluster and had the same stress pattern—that is, they were indistinguishable from the perspective of our phonological constraints. For example, [mnénijem] 'opinion (inst)' appeared with [s] almost exclusively, whereas [mnóz̧əstvəm] 'a large amount (inst)' showed considerable variation between [s] and [so]. Crucially, the morphemes in the lexicon cannot be simply divided into two classes, regular morphemes and exceptions. Among [mn]-initial words, for example, there are at least three classes: words that categorically occur with the C form [s], such as [mnénijem]; words that categorically occur with the CV form [so], such as [mnój] 'me (inst)'; and words that occur with both forms, such as [mnóz̧əstvəm] 'a large amount'. Even more problematically for a dichotomous distinction between exceptions and regular words, the words that occur with both forms vary in the *rates* at which they occur with each form; for instance, the rate of the CV form [vo] before [gréx] 'sin' is 20%, whereas before [grát] 'city' it is 60%. This suggests that people keep track of fine-grained distributional information pertaining to the co-occurrence of each individual lexical item with C and CV forms.

This section presents a scaled constraints analysis that accounts for this pattern. We start by presenting the scaled constraints mechanism (6.1). We then demonstrate how this mechanism captures lexical variation both among prepositions and among the nouns that condition prepositional alternations (6.2). We then discuss effects of register (6.3) and morphological context (6.4). Finally, we discuss how our findings bear on the relationship between variation and frequency (6.5) and the syntax-phonology interface (6.6).

6.1 Introducing scaled constraints

There are several approaches to lexical variation in Optimality Theory and related frameworks. In cophonology theory (Anttila 2002, Inkelas et al. 1996, Orgun 1996, Inkelas and Zoll 2007), lexical variation is analyzed by assigning different constraint rankings to different morphemes. By contrast, the indexed constraints approach (Benua 1997, Ito and Mester 1999, Pater 2006) assumes a single constraint ranking for the entire language and captures lexical variation by positing two or more instantiations of the same constraint, each indexed to a different set of morphemes. When applied to Harmonic Grammar or a Maximum Entropy grammar, this approach translates into morpheme-specific constraint weights. Finally, Coetzee and Pater (2011) and Coetzee and Kawahara (in press) implement morpheme-specific weights in Harmonic Grammar using *scaling factors*. When a faithfulness constraint is evaluated on a given word, the weight of the contrsaint is adjusted by an additive scaling factor, which is determined by extra-grammatical properties of the word (e.g. its frequency). In the scaling factor approach, the constraint is not multiply instantiated in the grammar; instead, its weight varies depending on the word being evaluated.

We adopt the scaling factors approach, but extend it in three ways. First, Coetzee and Kawahara (in press) assume that all faithfulness constraints are scaled to the same degree by a morpheme's scaling factor; we depart from this assumption and instead allow the scaling factors to vary by constraint. Second, we assume that scaling factors can apply both to markedness and faithfulness constraint (Pater 2006, Gouskova 2007, Flack 2007). Finally, we introduce scaling factors that are lexically listed and vary idiosyncratically from morpheme to morpheme, in addition to the scaling factors that reflect properties such as frequency and register. Using the notation introduced in Section 5.3, if s_{ji} is the sum of all scaling factors relevant to evaluating constraint i with respect to candidate j, including frequency, register and lexically specific factors, then the harmony of candidate j is given by:

$$H_j = \sum_{i=1}^{n} (w_i + s_{ji}) v_{ji}$$

While our modified constraint scaling approach is in some respects a notational variant of constraint indexing, we prefer it for two reasons. First, a constraint scaling mechanism is independently needed to capture the register effect we have observed. Second, as we show elsewhere, constraint weights are determined not only by the individual morpheme immediately following the preposition, but also by the morphological structure of the derived word that the morpheme is part of, which is not necessarily listed in the lexicon. It is not obvious how this would be handled using lexical indexing. A flexible constraint scaling factor approach can straightforwardly predict the CV rate of the derived word from the stem and its suffixes (Gouskova and Linzen, in preparation).

6.2 A scaled constraint analysis

The three prepositions [k/ko], [s/so] and [v/vo] are exceptional in that other prepositions which end in a vowel, such as [na] 'on' and [po] 'by/along', never alternate with monoconsonantal forms (*[n] or *[p]). In our analysis, the constraint that needs to be weighted differently for these three prepositions is Max-V, which blocks the deletion of the vowel. By default, the weight of Max-V exceeds that of Align enough to never allow vowel-zero alternations, which achieves the categorical resistance to deletion in most prepositions. Under the scaling factors approach, the alternating prepositions need to be assigned a negative scaling factor, which would bring the weight of Max-V down enough to enable Align to significant impact the form of the preposition. Indeed, in the MaxEnt model fit shown in Section 5.3, Max-V is the lowest-weighted constraint. Assuming that the language-wide weight of Max-V is very high, say 25, the scaling factor for the prepositions should be $s_{k/s/v} = -25$. Of course, each of the prepositions may have a different scaling factor, indicating a lower overall tendency for deletion in e.g. [s] than in [k].[19]

Moving on to the variation among the morphemes following the three prepositions, recall that words that can condition the CV form of the prepositions are overwhelmingly cluster-initial. At the same time, the specific word-initial cluster is not restricted to any phonologically defined subset of clusters (e.g. falling sonority clusters only). This suggests that the constraint that should be assigned morpheme-specific weights is *#CCC, which encourages the retention of the vowel in the preposition when it is followed by a cluster-initial morpheme. We assume that the nonce word experiment reflects speakers' phonological knowledge about regular lexical items; the basic weight of *#CCC should therefore be the weight obtained from our

[19]In our MaxEnt model of the results of the experiment, the weight assigned to Max-V was exactly 0. This was due to the fact that our data did not allow the model to distinguish Max-V from Align: whenever the former was violated, the latter was not, and vice versa.

MaxEnt model, namely 4.26. Lexical items that show a higher CV preference than would be expected from their phonological properties are assigned positive scaling factors, such that the weight of *#CCC is higher than the default for these words. For instance, the contrast between the near-categorical deletion of the vowel in [s dvígətilʲə] 'from the engine' and the tendency to retain the vowel in [sə dvar-á] 'from the yard' will be captured by assigning a positive scaling factor for *#CCC to [dvór].

As mentioned above, the gradient variation in CV rates between lexical items does not lend itself to a dichotomous distinction between regular items and exceptions: in principle, each lexical item may require a different scaling factor for *#CCC, although there may well be some sub-regularities (Zuraw 2000, Albright and Hayes 2003). See (34) for a rough estimate of the range of scaling factors required to derive the range of CV rates in the corpus (abstracting away from the words' phonological shape, for illustration purposes). The effect of scaling factors is illustrated in (35) for [dvór], under the assumption that $s_{dvor} = 2.98$.

(34) A scaling factors approach to gradient exceptionality

	Scaling factor	Weight of *#CCC	CV rate
Most words	0	4.26	depends on cluster
[gréx]	0.85	$0.85 + 4.26$	20%
[grát]	1.7	$1.7 + 4.26$	60%
[dvór]	2.98	$2.98 + 4.26$	60-96%
[mnój]	8.5	$8.5 + 4.26$	≈ 100%

(35) A scaling factors treatment of lexical variation

/so dvor-a/	ALIGN	*#CCC	*SIMC-C	*v(C)LAB	PARSE-σ	SSP	$\mathcal{H}$	*Prob*
sə(dvará)	-1×6.24				-1×0.37	0	−6.61	66%
(zdvará)		$-1 \times (s_{dvor} + 4.26) = -1 \times 7.24$				0	−7.24	34%

Even when a morpheme conditions idiosyncratic CV rates, the distribution of each of the specific prepositions is affected by the phonological constraints we have identified. For instance, [dvór] shows a non-adjacent OCP effect: the mean CV rate for [s] and [k] is 62%, whereas for the preposition [v], which is identical to one of the consonants in the word-initial cluster, the CV rate is an almost categorical 96%. The interaction with language-wide phonological constraints suggests that learners do not simply store the deletion rate for each morpheme in their lexicon (as proposed, for example, by Guy 2007), but rather consider both language-wide and lexically-specific constraint weights in the same additive weighting. The greater preference of [dvór] for the CV form of [v], compared to [s] and [k], falls out naturally from the combination of the scaled weight of *#CCC and the labial dissimilation constraint *v(C)LAB. There is no need to index any other constraint to [dvór] to derive this behavior, as shown in (36):

(36) The additive interaction between phonological and lexical variation

/vo dvor-e/	ALIGN	$*\#CCC_{dvor}$	*SIMC-C	*v(C)LAB	PARSE-σ	SSP	$\mathcal{H}$	*Prob*
və(dvaré)	-1×6.24				-1×0.37	0	−6.61	96%
(vdvaré)		$-1 \times (s_{dvor} + 4.26) = -1 \times 7.24$		-1×2.74		0	−7.98	4%

Tying the lexical variation to *#CCC makes a potential prediction for other contexts: if another phonological rule involves the prohibition on #CCC clusters, then morphemes indexed to the constraint should exhibit variation proportional to their scaling factor.[20] Word-initial CCC clusters can be created in several ways: through morpheme concatenation /C + CC.../ or deletion in the preposition /CV + CC/ → #CCC, or by deleting a vowel morpheme-internally between three consonants: /CVCC/ → CCC... or /CCVC/→ CCC.... The latter kind of deletion is not found in Russian nouns—Gouskova and Becker (to appear) show that CVCC words never have yer deletion in Russian, and CCVC words resist it, as well. Since only #CCV... words condition the prepositional alternation variably, the lack of such words in the yer sublexicon makes the prediction impossible to test.

6.3 Register effects

Many of the morphemes we found that showed a CV preference were ecclesiastical words (see Section 2.3.1). An overall register bias of this sort is straightforwardly analyzed using the scaling factor mechanism: a register-wide scaling factor increases the weight of faithfulness to the underlying vowel of the preposition (Max-V), joining the existing preposition-specific scaling factor (Boersma and Hayes 2001, Coetzee and Kawahara in press). For illustration purposes, consider the word [krést] 'cross', which has a CV rate of 68%. Non-ecclesiastical words that start with the same cluster, such as [kravát^j] 'bed', tend to have a very low CV rate. Recall that we stipulated that the global weight of Max-V is 25, such that $s_{k/s/v} = -25$. Tableaux (37) and (38) show how the difference between [krést] 'cross' and the non-ecclesiastical [kravát^j] 'bed' can be accounted for by assuming that the register-wide scaling factor is $s_{ecc} = 3$:

(37) Analyzing the register effect using scaling factors: ecclesiastical words

/so krest-om/	Align	*#CCC	Parse-σ	Max-V	SSP	$\mathcal{H}$	*Prob*
sə(kristóm)	-1×6.24		-1×0.37		0	-6.61	65%
(skristóm)		-1×4.26		$-1 \times (25 + s_{k/s/v} + s_{ecc}) = -1 \times 3$	0	-7.26	35%

(38) Analyzing the register effect using scaling factors: other words

/so kravát^j-ju/	Align	*#CCC	Parse-σ	Max-V	SSP	$\mathcal{H}$	*Prob*
sə(kravát^{j}ju)	-1×6.24		-1×0.37		0	-6.61	8%
(skravát^{j}ju)		-1×4.26		$-1 \times (25 + s_{k/s/v}) = -1 \times 0$	0	-4.26	92%

Alternatively, it may be that the ecclesiastical words we found constitute an isolated set of lexical items that emerged through a diachronic process and are therefore not qualitatively different from other high-CV words like [dvór]. If that is the case, the higher CV rate should be handled by morpheme-specific scaling of *#CCC as in (35), rather than by register-wide scaling of Max-V as in (37). This alternative explanation predicts that a putative religion-related nonce word should not exhibit register effects in an experiment such as the one described in Section 4.

[20] We would like to thank the Associate Editor for pointing out this prediction.

6.4 Morphological context

The problem of CV rates varying by morphological context has two sides: variation within the inflectional paradigm and conditioning by derivational suffixes (see Section 2.3.3). We will focus on the inflectional cases; for a full treatment of derivational effects, see Gouskova and Linzen (in preparation). The most prominent example of CV rates varying within the inflectional paradigm is [dnʲ-] 'day', which conditions different CV rates for the two homophones of [s], 'with' and 'from': [z dnʲóm] 'with the day (birthday)' but [sa dnʲá] 'from the day' (see Figure 3). Similarly, the root /mnog-/ 'much' in [mnóʐ-əstv-ə] conditions slightly more CV prepositions when used to mean 'a large amount' and more C prepositions when used to mean 'a mathematical set'. This pattern was specific to these roots—we did not find an overall difference in CV rates between the two homophones of the preposition [s] in the Yandex corpus (Section 3.2.5). The explanation we suggested for these cases is that the root, e.g. 'day', has two homophonous allomorphs, which differ slightly in their contextual meaning and in how they condition prepositional phonology. The allomorph selected in the "holiday" context, [z dnʲóm], behaves like most other [dn]-initial words. The allomorph selected in other contexts has a higher scaling factor for *#CCC, favoring CV prepositions. The context for allomorph selection is semantic/morphosyntactic (Bobaljik 2008), but it has consequences for the phonology. The analysis is no different otherwise, using a scaling factor for allomorphs just as for other exponents of morphemes, as shown in (39):

(39) Allomorphy and variation in special semantic contexts

$\sqrt{\text{DAY}}$	$\leftrightarrow$	/$\text{den}^{\text{j}}{}_{s(*\#CCC)=0}$/	/__'birthday'
		/$\text{den}^{\text{j}}{}_{s(*\#CCC)=2}$/	
$\sqrt{\text{MUCH}}$	$\leftrightarrow$	/$\text{mnog}_{s(*\#CCC)=0}$/	/__'mathematical set'
		/$\text{mnog}_{s(*\#CCC)=0.5}$/	

6.5 Variation and frequency

We now turn to the role of word frequency in the alternation. High token frequency is generally known to encourage deletion: when faced with the choice between a longer and a shorter form of a word, people are often more likely to use the shorter form when the word is frequent or predictable. For example, the schwa in the frequent word *memory* is more likely to delete than the schwa in the infrequent word *mammary* (Hooper 1976, Dalby 1984, Patterson et al. 2003). Similar results have been demonstrated for English t/d deletion (Bybee 2000, Coetzee and Kawahara in press), Dutch schwa epenthesis (Tily and Kuperman 2012), and other phenomena. In contrast with these studies, we did not find evidence that speakers were less likely to use the longer CV form of the preposition before high frequency words. This may be because only phonetically gradient reductions increase in likelihood when the frequency of the word increases; the Russian vowel-zero alternation appears to be a case of categorical deletion rule rather than a gradient phonetic process of the sort described in the studies mentioned above. However, in the absence of phonetic data on the realization of the vowel in CV prepositions, we can not draw definite conclusions from this fact.

Our study also differs from previous studies in that the deleted material is part of the preposition rather than the content word. The CV rate should therefore be modulated by the frequency of the preposition, rather than the frequency of the content word. There is evidence that frequency does not affect the pronunciation of function words, at least not to the same extent as content words (Bell et al. 2009). In addition, our sample of prepositions is very small and inherently confounded with phonological factors, making it difficult to detect effects of the frequency of the preposition on the CV rate. On the other hand, Bell et al. (2009) show that most function words are more likely to be reduced in duration when they are predictable from the following context word. This predicts that the vowel in a given preposition, say [s/so], should be less likely to be realized before a word that always occurs with that particular preposition (for example, in the context of a frequent collocation). We leave the testing of this prediction for future research.

The lemma frequency effects we did find went in the opposite direction: in monosyllabic yer words, higher frequency words were more likely to appear with CV prepositions, retaining a residue of the historical yer rule (see Section 2.4). These monosyllabic roots pattern with high frequency function words, which always appear with CV prepositions ([sa mnój] 'with me' and [sa fsém] 'with all'). Frequency thus has a *conserving* rather than *reducing* effect in our data (Bybee 2006): after systematic yer alternations were lost, the CV rate of each monosyllabic root had to be learned on a morpheme-by-morpheme basis. Morpheme-specific rates for low frequency morphemes did not have enough evidence in their favor; as a consequence, they were not learned by the the generation of speakers, and disappeared from the language. Our data join previous examples of sound changes that affect infrequent words first, such as glide deletion in south American English (Phillips 1981, 2006).

We found an interaction between lemma frequency and the degree of sonority sequencing principle violation, such that frequent words showed a slightly larger effect of sonority sequencing. This is the opposite of what would be expected from a conserving effect of frequency: frequent words should be less affected by phonological pressure and should therefore be more likely to retain the older pattern (i.e. preference for CV).[21] It is possible that sonority slope has two kinds of effects: both the synchronic effect on phonologically "regular" words that we built into our MaxEnt model, and a diachronic filter on exceptions, which makes it more likely for words with SSP violations to become exceptions in the first place. The diachronic effect may be much larger than the synchronic one, which could explain the surprising direction of the interaction.

6.6 Syntax-phonology interface

We now return to the architecture of the syntax-phonology interface. In lexicalist theories of the syntax-phonology interface such as Lexical Phonology (Mohanan 1982, Kiparsky 1982, 1985, Kaisse and Shaw 1985), only lexical rules are allowed to have lexical exceptions, and only postlexical (e.g., phrasal) rules can be variable. Coetzee and Pater (2011) offer an extensive and convincing critique of this division, but their examples focus on lexical rules that are variable. The case of Russian shows that the so-called postlexical rules can indeed have lexical exceptions. Through the corpus study, we identified a number of morphemes that systematically appear with CV prepositions, even though the productive phonological pattern that is extended to nonce words calls for a C preposition in similar phonological environments.

There has been some disagreement as to whether Russian prepositions are really phonologized postlexically. According to Kiparsky (1985), Russian prepositions attach in the lexicon; his arguments are based on rule ordering in voicing assimilation (see also Hayes 1984). But, as we noted earlier, Russian prepositions are not affixes—they can phonologically cliticize onto nouns, adjectives, quantifiers, or adverbs. It seems doubtful that all of these combinations are derived in the lexicon. There are alternative analyses of voicing assimilation that do not assume either opaque rule ordering or this architecture, relying on the prosodic structure of prepositions instead (Padgett 2002, Gouskova 2010).

Russian is not an isolated example; many syntactically conditioned mutation rules have lexical exceptions (e.g., Mende, Hayes 1990, and mutations in Celtic languages, Green 2006). Our corpus studies and

[21] We thank the Associate Editor for pointing out this fact.

experiment supply another case of a variable phonological rule that is part of phrasal/sentence phonology[22] and that shows lexical variability. This has consequences for the organization of the grammar and the syntax/phonology interface. Phonological theory must allow for lexical exceptions in phrasal phonology, not just in the treatment of word phonology. The distinction between phrasal and word syntax is difficult to motivate on syntactic grounds alone (Marantz 1997, 2008, though cf. Bermudez-Otero 2011, 2012), so the elimination of the analogous distinction in phonology is independently motivated. Phrasal and word phonology are not different: both can show lexical variation.

7 Conclusion

Our study of Russian prepositional C/CV alternations confirmed the similarity avoidance patterns reported in previous studies: sequences of sibilants (s+sC) and labial continuants (v+[v/f]C) are systematically avoided in preconsonantal position, but adjacent dorsals are allowed (k+kC). We also discovered three new generalizations about this phenomenon. First, the OCP is gradient: the avoidance of C prepositions is strongest for adjacent identical consonants ([s] + [s]), but weakens when the consonants differ in one or more features (e.g. [s] + [z]). Second, the OCP applies to non-adjacent consonants in the same cluster: sequences such as [s] + [fs] and [v] + [dv] are disfavored. Third, in the case of [v], it applies even when the preposition differs in manner of articulation from the following consonant (e.g. [v] + [m]).

We reported two other novel phonological findings. First, the alternation is affected by the position of the stress in the following word: words stressed on the first syllable are more likely to condition the CV form than words stressed on other syllables. Second, the sonority slope of the word-initial cluster also affects the alternation: the CV form is more likely to be used to break up the sequence [s] + [vd] than the sequence [s] + [vr]. All of these patterns are attested in the corpora, and were extended to novel words by participants in our experiment, suggesting that they form part of a productive grammar. We analyzed these patterns in a weighted constraint framework, MaxEnt, in which constraints interact additively: two constraints with low weights can override a single constraint with a higher weight.

We also found pervasive lexical variation: a large number of words appeared with CV prepositions more frequently than would be predicted based on their phonological properties. We showed that the selection of the CV form cannot be analyzed using the traditional *yer* rule (Lightner 1965 et seq.): many of the high-CV words do not contain underlying yer vowels, and some monosyllabic words that do contain a *yer* vowel do not cause the vowel in the preposition to be realized. Most striking is the example of the root [dn^{j}-] 'day', which contains a *yer*, but behaves differently after the preposition [s] depending on whether that preposition means 'with' or 'from', ruling out any purely phonological explanation.

The lexical variation we found would not be well characterized as an all-or-nothing distinction between regular and exceptional words. There is a wide range of CV rates, with some individual items favoring the CV forms more than others. In addition, the phonological constraints active in the general grammar enhance the tendency of high-CV words to select the CV form: for example, [dvór] tends to favor [so] and [ko], but its preference for [vo]—a labial—is even stronger. This suggests that CV rates for individual lexical items are not simply listed; they interact with phonological regularities, supporting the framework we adopted, which combines weighted constraints with morpheme-specific constraint weights, implemented using scaling factors.

[22]The distinction between lexical and postlexical rules is reified in Lexical Phonology. We are not assuming any distinction of this sort.

From a methodological point of view, the extensive lexical variation we found highlights the importance of combining corpus studies and nonce word experiments. Examples cherry-picked by the analyst are unlikely to be a random sample of words in the language: salient examples tend to be frequent words, and as such are more often exceptional. This bias may lead to spurious phonological generalizations. Comprehensive corpus studies help chart the terrain of lexical variation in a systematic way. After exceptions have been identified, they can be excluded in order to detect general phonological patterns. The productivity of these patterns can then be verified using a nonce word experiment. Finally, the convergence among the findings of our three studies shows that orthographic corpora and Web search engines, which provide access to a much larger data set than phonologically transcribed corpora, can be useful for studying phonological phenomena, at least ones that are not governed by standardized or prescriptive orthographic rules.

Appendix A: Experiment results

Items used in the forced choice experiment, with the proportion of participants who preferred each variant. This table only lists items that had [ib] as their rime (see Section 4.1.2 for details).

Input	Candidate	Probability	Input	Candidate	Probability
/so vsíbom/	s fsíbom	0.3	/so ştíbom/	s ştíbom	0.2
/so vsíbom/	sa fsíbom	0.7	/so ştíbom/	sa ştíbom	0.8
/so vsibóm/	s fsibóm	0.41	/so ştibóm/	s ştibóm	0.19
/so vsibóm/	sə fsibóm	0.59	/so ştibóm/	sə ştibóm	0.81
/so kíbom/	s kíbom	0.99	/so vdíbom/	s vdíbom	0.44
/so kíbom/	sa kíbom	0.01	/so vdíbom/	sa vdíbom	0.56
/so kibóm/	s kibóm	1.0	/so vdibóm/	s vdibóm	0.6
/so kibóm/	sə kibóm	0.0	/so vdibóm/	sə vdibóm	0.4
/so msíbom/	s msíbom	0.37	/so vníbom/	s vníbom	0.6
/so msíbom/	sa msíbom	0.63	/so vníbom/	sa vníbom	0.4
/so msibóm/	s msibóm	0.53	/so vnibóm/	s vnibóm	0.68
/so msibóm/	sə msibóm	0.47	/so vnibóm/	sə vnibóm	0.32
/so rdíbom/	s rdíbom	0.42	/so vzíbom/	s vzíbom	0.31
/so rdíbom/	sa rdíbom	0.58	/so vzíbom/	sa vzíbom	0.69
/so rdibóm/	s rdibóm	0.51	/so vzibóm/	s vzibóm	0.44
/so rdibóm/	sə rdibóm	0.49	/so vzibóm/	sə vzibóm	0.56
/so rníbom/	s rníbom	0.49	/so xríbom/	s xríbom	0.93
/so rníbom/	sa rníbom	0.51	/so xríbom/	sa xríbom	0.07
/so rnibóm/	s rnibóm	0.55	/so xribóm/	s xribóm	0.89
/so rnibóm/	sə rnibóm	0.45	/so xribóm/	sə xribóm	0.11
/so şríbom/	s şríbom	0.24	/s xsíbom/	s xsíbom	0.51
/so şríbom/	sa şríbom	0.76	/s xsíbom/	sa xsíbom	0.49
/so şribóm/	s şribóm	0.3	/so xsibóm/	s xsibóm	0.6
/so şribóm/	sə şribóm	0.7	/so xsibóm/	sə xsibóm	0.4

Appendix B: Presentation

Type 1		Type 2	
ja intirisújus^{j}___	‘I am interested in’	ja pridú s(o)___	‘I will come with’
ja xvástəjus^{j}___	‘I boast of’	ja gul^{j}áju s(o)___	‘I walk with’
ja risúju___	‘I paint with’	ja druʐú s(o)___	‘I am friends with’
ja davól^{j}nə___	‘I am pleased by’	ja igráju s(o)___	‘I play with’

Table 8: Frame sentences for words with and without prepositions.

(40) An example of how nonce words were presented to the participants

Прослушайте пример употребления слова:
Please listen to an example of the word:
Я довольна взуном **[audio file]**
I am pleased by vzun-om
Теперь прослушайте еще два предложения с этим словом:
Now listen to two more sentences with this word:
Я приду **1 [audio file]**
I will come with
2 [audio file]
Какое предложение кажется более правильным? **[1] [2]**
Which sentence seems more correct/right?

References

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

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

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

Avanesov, Ruben I. 1968. *Russkoe literaturnoe proiznoshenie [Literary pronunciation of Russian]*. Moscow: Izdatelstvo Prosvescenie, 6th (1984) edition.

Bakovic, Eric. 2005. Antigemination, assimilation and the determination of identity. *Phonology*, 22:279–315.

Barnes, Jonathan. 2004. Vowel reduction in Russian: The categorical and the gradient. Boston, MA. Handout from LSA 78.

Bates, Douglas, Martin Maechler, and Ben Bolker. 2012. *lme4: Linear mixed-effects models using S4 classes*. URL `http://CRAN.R-project.org/package=lme4`, r package version 0.999999-0.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60:92–111.

Benua, Laura. 1997. Affixes classes are defined by faithfulness. In *University of Maryland working papers in linguistics 5. Selected phonology papers from Hopkins Optimality Theory workshop 1997 / University of Maryland Mayfest 1997*, ed. Viola Miglio and Bruce Moren, 1–27.

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

Bermudez-Otero, Ricardo. 2011. Cyclicity. In van Oostendorp et al. (2011), 2019–2048.

Bermudez-Otero, Ricardo. 2012. The architecture of grammar and the division of labor in exponence. In *The morphology and phonology of exponence: The state of the art*, ed. Jochen Trommer, 8–83. Oxford: Oxford University Press.

Bethin, Christina. 1992. *Polish syllables: The role of prosody in phonology and morphology*. Columbus: Slavica Publishers.

Bethin, Christina. 2006. Stress and tone in East Slavic dialects. *Phonology*, 23:125–156.

Blumenfeld, Lev. 2011. Vowel-zero alternations in Russian prepositions: Prosodic constituency and productivity. In *Russian language studies in North America*, ed. Veronika Makarova, 43–69. Anthem Press.

Bobaljik, Jonathan. 2008. Paradigms (Optimal and otherwise): A case for scepticism. In *Inflectional identity*, ed. Asaf Bachrach and Andrew Nevins, 29–54. Oxford: Oxford University Press.

Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 21:43–58. Available as ROA-221 on Rutgers Optimality Archive, http://roa.rutgers.edu.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, 32:45–86. Available as ROA-348 on Rutgers Optimality Archive, http://roa.rutgers.edu.

Boersma, Paul, and David Weenink. 2009. Praat: Doing phonetics by computer, version 5.1.

Bybee, Joan. 2000. The phonology of the lexicon: Evidence from lexical diffusion. In *Usage-based models of language*, ed. M. Barlow and S. Kemmer, 65–85. Stanford: CSLI.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language*, 82:711.

Canty, Angelo, and B. D. Ripley. 2012. *boot: Bootstrap R (S-Plus) functions*. R package version 1.3-4.

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

Clopper, CJ, and Egon S. Pearson. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26:404–413.

Coetzee, Andries, and Shigeto Kawahara. in press. Frequency biases in phonological variation. *Natural Language and Linguistic Theory*.

Coetzee, Andries, and Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory*, 26:289–337.

Coetzee, Andries, and Joe Pater. 2011. The place of variation in phonological theory. In *The handbook of phonological theory*, ed. John Goldsmith, Jason Riggle, and Alan Yu. Cambridge: Blackwell, 2nd edition.

Crosswhite, Katherine. 1999. Vowel reduction in Optimality Theory. Doctoral Dissertation, University of California, Los Angeles, Los Angeles, CA.

Czaykowska-Higgins, Ewa. 1988. Investigations into Polish morphology and phonology. Doctoral Dissertation, MIT, Cambridge, MA.

Dalby, Jonathan. 1984. Phonetic structure of fast speech in American English. Doctoral Dissertation, Indiana University, Bloomington.

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

Davison, Anthony C., and David V. Hinkley. 1997. *Bootstrap methods and their applications*. Cambridge: Cambridge University Press.

Fanselow, Gisbert, and Damir Ćavar. 2002. Distributed deletion. In *Theoretical approaches to universals*, ed. Artemis Alexiadou, 65–107. Amsterdam: John Benjamins.

Farina, Donna. 1991. Palatalization and yers in modern Russian phonology: An underspecification approach. Doctoral Dissertation, University of Illinois, Urbana, Urbana, IL.

Farris-Trimble, Ashley. 2008. Cumulative faithfulness effects: Opaque or transparent? In *Indiana university working papers in linguistics*, ed. Ashley Farris-Trimble and Daniel Dinnsen, volume 6, 119–145. Bloomington, IN: IULC Publications.

Ferguson, Charlesharles A. 1956. The emphatic L in Arabic. *Language*, 32:446–452.

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

Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory*, 22:179–228.

Gafos, Adamantios. 1999. *The articulatory basis of locality in phonology*. New York: Garland.

Gallagher, Gillian, and Jessica Coon. 2009. Distinguishing total and partial identity: Evidence from Chol. *Natural Language and Linguistic Theory*, 27:545–582.

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

Gouskova, Maria. 2003. Deriving economy: Syncope in Optimality Theory. Doctoral Dissertation, University of Massachusetts, Amherst, Amherst, MA.

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

Gouskova, Maria. 2010. The phonology of boundaries and secondary stress in Russian compounds. *The Linguistic Review*, 17:387–448.

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

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

Gouskova, Maria, and Tal Linzen. in preparation. Morphological effects in lexical variation: The case of Russian prepositions. Ms., NYU.

Green, Anthony Dubach. 2006. The independence of phonology and morphology: The Celtic mutations. *Lingua*, 116:1946–1985.

Gribanova, Vera. 2008. Russian prefixes, prepositions and palatalization in Stratal OT. In *Proceedings of WCCFL 26*, ed. Charles B. Change and Hannah J. Haynie, 217–225. Somerville, MA: Cascadilla Press.

Gribanova, Vera. 2009a. Phonological evidence for a distinction between Russian prepositions and prefixes. In *Studies in formal Slavic phonology, morphology, syntax, semantics and information structure: Proceedings of the 7th European conference on Formal Description of Slavic Languages*, ed. Gerhild Zybatow, Uwe Junghanns, Denisa Lenertová, and Petr Biskup, 383–396. Frankfurt am Main: Peter Lang Verlag.

Gribanova, Vera. 2009b. The phonology and syntax of sub-words. Presented at GLOW 32, 16 April 2009.

Gribanova, Vera. 2010. Composition and locality: The morphosyntax and phonology of the Russian verbal complex. Doctoral Dissertation, UC Santa Cruz, Santa Cruz, CA.

Guy, Gregory R. 1980. Variation in the group and the individual: The case of final stop deletion. In Labov (1980), 1–36.

Guy, Gregory R. 1991. Contextual conditioning in variable lexical phonology. *Language Variation and Change*, 3:223–39.

Guy, Gregory R. 2007. Lexical exceptions in variable phonology. *University of Pennsylvania Working Papers in Linguistics*, 13:9.

Hall, Nancy. 2011a. Vowel epenthesis. In *Companion to phonology*, ed. Elizabeth Hume Marc van Oostendorp, Colin J. Ewen and Keren Rice. Cambridge: Wiley-Blackwell. URL `http://www.companiontophonology.com`, blackwell Reference Online. 13 August 2012.

Hall, Nancy. 2011b. Vowel epenthesis. In van Oostendorp et al. (2011), 1576–1596.

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

Halle, Morris, and Jean-Roger Vergnaud. 1987. *An essay on Stress*. Cambridge, MA: MIT Press.

Hayes, Bruce. 1984. The phonetics and phonology of Russian voicing assimilation. In *Language sound structure*, ed. Mark Aronoff and Richard T. Oehrle, 318–328. Cambridge, MA: MIT Press.

Hayes, Bruce. 1990. Precompiled phrasal phonology. In *The phonology-syntax connection*, ed. Sharon Inkelas and Draga Zec, 85–108. Chicago: University of Chicago Press.

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

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

Hooper, Joan. 1976. Word frequency in lexical diffusion and the source of morphophonological change. In *Current progress in historical linguistics*, ed. William M. Christie, 96–105. Amsterdam: North Holland.

Howe, Darin, and Douglas Pulleyblank. 2004. Harmonic scales as faithfulness. *Canadian Journal of Linguistics*, 49:1–49.

Inkelas, Sharon, C. Orhan Orgun, and Cheryl Zoll. 1996. Exceptions and static phonological patterns: Cophonologies vs. prespecification. Available as ROA-124-0496 on the Rutgers Optimality Archive, http://roa.rutgers.edu.

Inkelas, Sharon, Orhan Orgun, and Cheryl Zoll. 1997. The implications of lexical exceptions for the nature of grammar. In *Derivations and constraints in phonology*, ed. Iggy Roca, 393–418. Oxford: Oxford University Press.

Inkelas, Sharon, and 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, and Armin Mester. 1999. The phonological lexicon. In *The handbook of Japanese linguistics*, ed. Natsuko Tsujimura, 62–100. Oxford: Blackwell.

Ito, Junko, and Armin Mester. 2003. On the sources of opacity in OT: Coda processes in German. In *The syllable in Optimality Theory*, ed. Caroline Féry and Ruben van de Vijver, 271–304. Cambridge: Cambridge University Press.

Jäger, Gerhard, and Anette Rosenbach. 2006. The winner takes it all—almost: Cumulativity in grammatical variation. *Linguistics*, 44:937–971.

Kaisse, Ellen M., and Patricia Shaw. 1985. On the theory of lexical phonology. *Phonology*, 2:1–30.

Katz, Jonah. 2005. Russian consonant Cvljusters. Ms., MIT. Available for download from http://web.mit.edu/jikatz/www/KatzPhonSquib.pdf.

Keller, Frank. 2000. Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. Doctoral Dissertation, University of Edinburgh.

Kenstowicz, Michael, and Jerzy Rubach. 1987. The phonology of syllabic nuclei in Slovak. *Language*, 63:463–497.

Kiparsky, Paul. 1982. Lexical phonology and morphology. In *Linguistics in the morning calm*, ed. In-Seok Yang, volume 2, 3–91. Seoul: Hanshin.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology*, 2:85–138.

Kitto, Catherine, and Paul de Lacy. 2000. Correspondence and epenthetic quality. In *Proceedings of AFLA VI: The sixth meeting of the Austronesian Formal Linguistics Association (Toronto working papers in linguistics)*, ed. Catherine Kitto, Carolyn Smallwood, Catherine Kitto, and Carolyn Smallwood, 181–200. Toronto: Department of Linguistics, University of Toronto.

Kochetov, Alexei, and Milica Radišić. 2009. Latent consonant harmony in Russian: Experimental evidence for agreement by correspondence. In *Proceedings of FASL 17*, ed. Maria Babyonyshev, Darya Kavitskaya, and Jodi Reich, 111–130. Michigan Slavic Publications.

Labov, William, ed. 1980. *Locating language in time and space*. New York: Academic Press.

Labov, William. 1989. The child as linguistic historian. *Language variation and change*, 1:85–97.

Leben, Will. 1973. Suprasegmental Phonology. Doctoral Dissertation, MIT, Cambridge, MA.

Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic Grammar – a formal multi-level connectionist theory of linguistic wellformedness: Theoretical foundations. In *Proceedings of the twelfth annual conference of the Cognitive Science Society*, 388–395. Cambridge, MA: Lawrence Erlbaum.

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

Lightner, Theodore. 1972. *Problems in the theory of phonology*. Edmonton: Linguistic Research, Inc.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In *University of Pennsylvania working papers in linguistics: Proceedings of the 21st Annual Penn Colloquium*, ed. Alexis Dimitriadis, Laura Siegel, Clarissa Surek-Clark, and Alexander Williams, volume 4.2, 201–225.

Marantz, Alec. 2008. Phases and words. Ms. NYU.

Matushansky, Ora. 2002. On formal identity of Russian prefixes and prepositions. In *MIT Working Papers in Linguistics 42*, 217–253. Cambridge, MA: MIT.

McCarthy, John J. 1986. OCP Effects: Gemination and antigemination. *Linguistic Inquiry*, 17:207–263.

McCarthy, John J. 2011. Perceptually grounded faithfulness in Harmonic Serialism. *Linguistic Inquiry*, 42:171–183.

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

McCarthy, John J., and Alan Prince. 1995. Faithfulness and reduplicative identity. In *University of Massachusetts Occasional Papers in Linguistics 18*, ed. Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 249–384. Amherst, MA: GLSA Publications.

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

Mohanan, Karuvannur P. 1982. Lexical Phonology. Doctoral Dissertation, MIT, Cambridge, MA. Distributed by IULC Publications.

Moreton, Elliott, and Paul Smolensky. 2002. Typological consequences of local constraint conjunction. In *Proceedings of WCCFL 21*, ed. L. Mikkelsen and C. Potts, 306–319. Somerville, MA: Cascadilla Press.

Neu, Helene. 1980. Ranking of constraints on /t, d/ deletion in American English: A statistical analysis. In Labov (1980), 37–54.

Nunberg, Geoff. 2009. Climategate, Tiger, and Google hit counts: dropping the other shoe. URL `http://languagelog.ldc.upenn.edu/nll/?p=1943`.

Ohala, John J., and Haruki Kawasaki-Fukumori. 1997. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In *Language and its ecology: Essays in memory of Einar Haugen. Trends in Linguistics. Studies and Monographs, Vol. 100*, ed. Stig Eliasson and Ernst Hakon Jahr, 343–365. Berlin: Mouton de Gruyter.

van Oostendorp, M., C.J. Ewen, E.V. Hume, and K. Rice, ed. 2011. *The Blackwell companion to phonology*. Malden, MA and Oxford: Wiley-Blackwell.

Orgun, C. Orhan. 1996. Sign-based morphology and phonology, with special attention to Optimality Theory. Doctoral Dissertation, University of California, Berkeley.

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

Padgett, Jaye. 2008. Glides, vowels, and features. *Lingua*, 118:1937–1955.

Padgett, Jaye. 2012. The role of prosody in Russian voicing. In *Prosody matters: Essays in honor of Elisabeth Selkirk*, ed. Toni Borowsky, Takahito Kawahara, Shigeto Shinya, and Mariko Sugahara, 181–207. London: Equinox Publishing.

Padgett, Jaye, and Marija Tabain. 2005. Adaptive dispersion theory and phonological vowel reduction in Russian. *Phonetica*, 62:14–54.

Pająk, Bożena, and Eric Bakovic. 2010. Assimilation, antigemination, and contingent optionality: The phonology of monoconsonantal proclitics in Polish. *Natural Language and Linguistic Theory*, 28:643–680.

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 *Papers in Optimality Theory III*, ed. Leah Bateman, Michael O'Keefe, Ehren Reilly, and Adam Werle, 259–296. Amherst, MA: GLSA.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science*, 33:999–1035.

Pater, Joe. 2012. Serial Harmonic Grammar and Berber syllabification. In *Prosody matters: Essays in honor of Lisa Selkirk*, ed. Toni Borowsky, Shigeto Kawahara, Takahito Shinya, and Mariko Sugahara, 43–72. London: Equinox Press.

Pater, Joe, Christopher Potts, and Rajesh Bhatt. 2007. Harmonic Grammar with Linear Programming. Ms, University of Massachusetts, Amherst.

Patterson, David, Paul C. LoCasto, and Cynthia Connine. 2003. Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica*, 60:45–69.

Pesetsky, David. 1979. Russian morphology and lexical theory. Ms., MIT.

Phillips, Betty S. 1981. Lexical diffusion and Southern *tune, duke, news*. *American Speech*, 56:72–78.

Phillips, Betty S. 2006. *Word frequency and lexical diffusion*. New York: Palgrave Macmillan.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt, and Michael Becker. 2010. Harmonic Grammar with Linear Programming. *Phonology*, 27:1–41.

Prince, Alan, and 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 Core Team. 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`, ISBN 3-900051-07-0.

Raymond, William D, Robin Dautricourt, and Elizabeth Hume. 2006. Word-internal /t, d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, 18:55.

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

Ryan, Kevin. to appear. Onsets contribute to syllable weight: Statistical evidence from stress and meter. *Language*.

Scheer, Tobias. 2006. How yers made Lightner, Gussman, Rubach, Spencer and others invent CVCV. In *Studies in constraint-based phonology*, ed. Piotr Banski, Beata Lukaszewicz, and Monika Opalinska, 133–207. Warsaw: Wydawnictwa Uniwersytetu Warszawskiego.

Sekerina, Irina. 1997. The syntax and processing of scrambling constructions in Russian. Doctoral Dissertation, City University of New York.

Selkirk, Elisabeth. 1995. The prosodic structure of function words. In *University of Massachusetts Occasional Papers: Papers in Optimality Theory*, ed. Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 439–470. Amherst, MA: GLSA Publications.

Sharoff, Serge. 2005. Methods and tools for development of the Russian Reference Corpus. In *Corpus linguistics around the world*, ed. Andrew Wilson, Dawn Archer, and Paul Rayson, Language and Computers, 167–180. Amsterdam and New York: Rodopi. `http://www.comp.leeds.ac.uk/ssharoff/frqlist/frqlist-en.html`.

Smolensky, Paul. 1995. On the internal structure of the constraint component Con of UG. Handout of a talk given at UCLA on 4/7/95.

Smolensky, Paul, and Géraldine Legendre. 2006. *The harmonic mind: From neural computation to Optimality-Theoretic grammar*. Cambridge, MA: MIT Press.

Steriade, Donca. 2001. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. Ms., MIT.

Steriopolo, Olga. 2007. Jer vowels in Russian prepositions. In *Formal Approaches to Slavic Linguistics. The Toronto Meeting 2006*, 365–385. Ann Arbor, MI: Michigan Slavic Publications.

Svenonius, Peter. 2004. Slavic prefixes and morphology: An introduction to the Nordlyd volume. *Nordlyd*, 32:177–204.

Tily, Harry, and Victor Kuperman. 2012. Rational phonological lengthening in spoken dutch. *The Journal of the Acoustical Society of America*, 132:3935–3940.

Timberlake, Alan. 2004. *A reference grammar of Russian*. Cambridge: Cambridge University Press.

Ukiah, Nick. 1998. Stress retraction in phrases of the type *na den', za sorok, ne byl* in Modern Russian. *Russian Linguistics*, 22:287–319.

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

Vlasto, A. P. 1986. *A linguistic history of Russia at the end of the eighteenth century*. Oxford: Clarendon Press.

Yaschenko, Ivan. 2002. *Paradoksy teorii mnozhestv [Paradoxes of set theory]*. Moscow: Izdatelsvo Moskovskogo tsentra nepreryvnogo matematicheskogo obrazovania.

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

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

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

Zuraw, Kie. 2000. Patterned exceptions in phonology. Doctoral Dissertation, UCLA.