Phonology **30** (2013) 453–515. © Cambridge University Press 2013
doi:10.1017/S0952675713000225

*Lexical and phonological variation in Russian prepositions**

Tal Linzen, Sofya Kasyanenko and Maria Gouskova
New York University

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. 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-constraints analysis. In addition to 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 analyse this pattern by augmenting the weighted-constraints approach with lexical scaling factors.

1 Introduction

Much 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, Pater 2000, 2007). For example, Chomsky & Halle (1968) observe that derived forms in English conserve the stress pattern of the base, as in *im'port* ~ *ˌimˌpor'tation*. 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, *con'sult*, which

* For feedback that has 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, 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, Yevgenia Gouskova, Stephanie Harves, Barbara Partee and Kevin Roon for help with recruiting the participants.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

ends in a sonorant–stop sequence just like *im'port*, shifts its stress position in the derived word *ˌconsul'tation* (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, 1991, Labov 1989, Raymond *et al*. 2006, Coetzee & Kawahara 2013). 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 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 & Pater 2011). Recently, Coetzee & Kawahara (2013) have analysed lexical frequency effects on stochastic variation using a weighted-constraints approach (Legendre *et al*. 1990, Smolensky & Legendre 2006, Pater 2009, Coetzee & 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] 'from, with', [v] 'into, in' and [k] 'towards' (Matushansky 2002, Timberlake 2004, Katz 2005, Steriopolo 2007, Gribanova 2009a, Blumenfeld 2012). These prepositions either surface as a single consonant (C) or as a consonant–vowel (CV) sequence ([sa/sə], [va/və] and [ka/kə], spelled in Russian orthography as *so*, *vo* and *ko* respectively). The vowel is [a] when the following syllable is stressed and [ə] otherwise (see Crosswhite 1999, Padgett & Tabain 2005, Bethin 2006, among others). Other prepositions, such as [za] 'behind', [po] 'along' and [na] 'on', do not alternate.[1]

[1] All transcriptions are in IPA, except where noted. Stress is represented with ˈ immediately preceding the vowel. We do not transcribe palatalisation before [e] and [i]; to indicate palatalisation 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 strings [so], [vo] and [ko], 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]. Likewise, we transcribe unstressed front vowels (reduced /i/ and /e/) as [i] rather than the phonetically more accurate [ɪ].

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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 the 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).

(1) *C form conditioned by clusters with rising sonority*

s prikˈazəm	*sə prikˈazəm	'with the order'
v brˈofʲ	*va brˈofʲ	'in the eyebrow'
k slanˈu	*kə slanˈu	'towards the elephant'

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).

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

*s stərikˈom	sə stərikˈom	'with the old man'
*v vrˈemʲə	va vrˈemʲə	'in time'

The generalisations 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).

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

s mnˈozʲəstvəm ~ sa mnˈozʲəstvəm	'with a large amount, (mathematical) set'

Variation is not always possible, however. Timberlake (2004: 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 mnˈoj	sa mnˈoj	'with me'
*s fsˈemi	sa fsˈemi	'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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

Phonological factors	
Obligatory Contour Principle Adjacent sibilants (s + {s z ʂ ʐ}) and adjacent labial fricatives (v + {f v}) tend to be avoided in the beginning of clusters (see §2.2).	[sə stərikˈom] ≻ [s stərikˈom]
Non-adjacent Obligatory Contour Principle 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 vzˈorəm] ≻ [z vzˈorəm]
Stress CV forms are more likely before stressed syllables than before unstressed ones (CV + *σ́* ≻ CV + *σ̌*) (see §3.2.2).	[sa vdˈoxə] ≻ [sə vdafʦˈa]
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 prikˈazəm] ≻ *[sə prikˈazəm] [sə vdaf˂ˈa] ≻ [s vridˈami]
Lexical factors	
Lexical idiosyncrasies Lexical items with similar phonological properties differ widely in their tendency to appear with CV forms (see §2.3).	[sə dvarˈa], [və dvarˈe] *vs.* [z dvajnikˈa], [v dvajnikˈe]
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 dnʲˈom] ‘with the day’ *vs.* [sa dnʲˈa] ‘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 snˈu] ‘towards sleep’ *vs.* [k psˈu] ‘towards the dog’
Conservative contexts Ecclesiastical and archaic contexts, as well as certain frozen expressions, favour CV forms (see §2.3.1, §2.3.4).	[liʐˈatʲ va prˈaxi] ‘be dead (lit. lie in ashes)’ *vs.* [prˈax k prˈaxu] ‘ashes to ashes’

Table I

Overview of the results of the two corpus studies.
X ≻ Y indicates that X is more well-formed than Y.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

more extensive than previously known, and exhibits several surprising morphological properties (see Table I 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 favours CV forms turns out to be more widespread than previously thought: it applies across intervening consonants (e.g. in [s+fs]) 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). 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+vd] is more likely than [sa+vr]).

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 the vowel is to surface. We model the results in a maximum entropy grammar with weighted constraints and probabilistic outputs (Goldwater & Johnson 2003, Coetzee & Pater 2008, Hayes & Wilson 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 2007). Our implementation takes the scaling factors approach proposed by Coetzee & Kawahara (2013) as its starting point, but modifies it in several ways. The scaling factors analysis allows us to straightforwardly analyse a register effect, whereby ecclesiastical words as a group (such as [kr'est] '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 [dv'or] '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 [dv'or] is 62%, whereas for [v] before [dv'or] it 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. §2 examines the alternation in the Russian National Corpus. §3 reports an additional corpus study, performed using a web search engine. §4 reports the results of a nonce-word experiment designed to test the productivity of the phonological patterns we found in the corpus studies. §5 proposes a set of constraints

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

that account for the phonological patterns we found, and models the results of the nonce-word experiment using a maximum entropy grammar. §6 takes up lexical variation, and §7 concludes.

2 Corpus study 1

This section presents the results of a study of the prepositional vowel–zero alternation in the Russian National Corpus (RNC).[2] 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 §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 (§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 (§2.3). Finally, §2.4 addresses the connection between prepositional vowel–zero alternations and Russian yer deletion (e.g. [lʲˈon] ~ [lʲn-ˈa] 'linen (NOM SG ~ GEN SG)'). Traditional accounts treat the vowel–zero alternation in these prepositions as a case of yer alternation, thus predicting that yer words should categorically select the CV form. Contrary to this 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 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 twelve forms – six 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 [mnˈoj] 'me (INSTR)' or [minʲˈa] 'me (GEN)'. Finally, we excluded lemmas that had a frequency of fewer than five 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

[2] Available at ruscorpora.ru.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

queries sent to the RNC site, we performed this preliminary elimination offline, using a frequency database (Sharoff 2005).[3]

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 ten 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 six times with [k] and seven times with [ko] was included in the sample for [k], but one that had six [k] tokens and three [ko] tokens was not. In total, 601 inflected forms met the threshold for [k]/[ko], 1133 for [v]/[vo] and 1017 for [s]/[so].

(5) *Case forms of nouns in preposition phrases with* [s v k]*: sample queries for* [dvˈor] *'yard'*

	singular	*plural*	
[k/ko] + DAT	g/kə dvar-ˈu	g/kə dvar-ˈam	'towards the yard(s)'
[s/so] + GEN	z/sə dvar-ˈa	z/sə dvar-ˈof	'from the yard(s)'
[s/so] + INSTR	z/sə dvar-ˈom	z/sə dvar-ˈami	'with the yard(s)'
[v/vo] + ACC	v/va dvˈor	v/və dvar-ˈɨ	'into the yard(s)'
[v/vo] + PREP	v/və dvar-ˈe	v/və dvar-ˈax	'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 & Pearson (1934) method, implemented as *binom.test* in R (R Development 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%.

[3] All of the data discussed in the article are available at is available in the supplementary online materials at journals.cambridge.org/issue_Phonology/Vol30No03.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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; these 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	postalveolar/ palatal	retroflex	dorsal
stops	p pʲ b bʲ	t tʲ d dʲ			k kʲ g gʲ
fricatives	f fʲ	s sʲ z zʲ	ʃ ʃʲ	ʂ ʐ	x xʲ
affricates		ts	tʃʲ		
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. [ftˈor] 'fluorine', from Greek *phthoros*, or [fˈol] 'foul (N)', 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 & 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_2 t_6] has a sonority slope of −4, and [t_6 n_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. [svˈet] 'light' *vs.* [zvˈon] 'ringing'), but with obstruents in undergoing the rule itself (e.g. /v + bar/ → [v bˈar] 'into the bar' *vs.* /v + par/ → [f pˈar] 'into the steam'). We therefore assume the sonority scale given in (7), while acknowledging that [v] may occasionally show ambiguous patterning.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

2.2 Phonological factors

The corpus results confirmed the generalisation 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 §2.3 below). We also confirmed the generalisation 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 stərikˈom] 'with an old man'), and [vo] is used before the labial continuants [v] and [f] (e.g. [va frˈantsii] '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 ($p < 0{\cdot}05$ 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 §3.2.1 below for similar results and discussion). The strong but non-categorical preference for the CV form before sibilants is also corroborated in our nonce-word experiment (§4). Thus the alternation is gradient even before sibilants: while there is a strong preference for [so] before [svintsˈa], occasionally a speaker will produce [s svintsˈa] 'from lead'.

We also 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 favoured when the *second* consonant of the word-initial cluster is a sibilant, as in the lemma [vzˈor] 'gaze': 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 (Fig. 1b): when the second consonant of the following word is a labial continuant, there is a higher proportion of CV forms.[4]

[4] Anticipating the discussion in §2.4, it should be noted that some of the examples in Fig. 1a have prefixes that have yer alternations, although they are quite limited in productivity (Pesetsky 1979, Svenonius 2004, Gribanova 2009b, 2010, Gouskova 2012). Examples with prefixes include /v + zor/ [vzˈor] 'gaze' and /v + xod/ [fxˈot] 'entrance'. The initial clusters in [vrˈet] 'damage', [ftˈornik] 'Tuesday' and [vrˈemjə] 'time' are not morphologically analysable. 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 ([fxˈot], [fkˈus] and [fklʲutʃˈenii]), 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 (*à la* the Lower analysis described in §2.4).

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

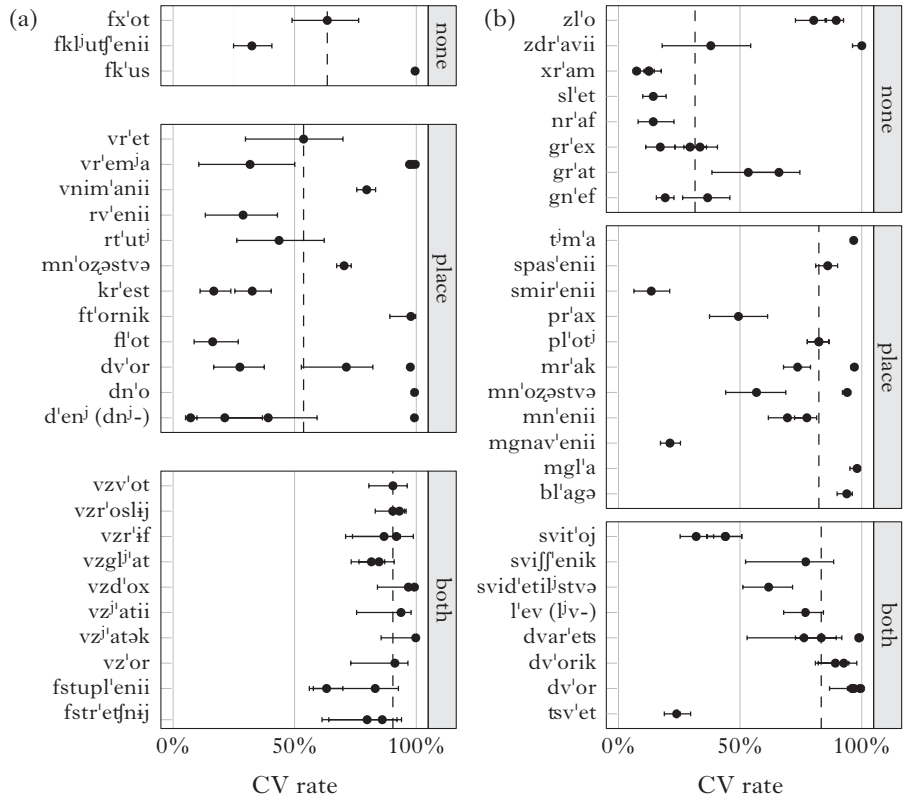


Figure 1

Effect of similarity between the preposition and the consonants in the word-initial cluster. Each row in the plot represents a lemma. A lemma can appear with each of the prepositions (a) [s/so] and (b) [v/vo] in four different forms (two cases, singular, plural); all relevant forms which met our inclusion criteria are shown in the plot. 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. Glosses for the forms in Figs 1–4 can be found in Appendix B.

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, favour

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

the CV form [vo] more than words that do not have a labial consonant (Fig. 1b). For example, a word like [mr'ak] 'darkness', which starts with the labial [m], will be more likely than a word like [nr'af] '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 & 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 & 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 regardless of place of articulation (Frisch *et al*. 2004, Coetzee & Pater 2008). Thus 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.

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]. Both models had a predictor for the similarity between the preposition and the following cluster, with one of the following three levels: no features in common (baseline level), shared major place, or shared place and manner.[5] 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 & Hinkley 1997, Canty & Ripley 2012). In each iteration of the bootstrap procedure, a random subset of the words was sampled 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 10,000 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. Statistical models

[5] Recall from §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.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

	predictor	coefficient	standard error	t	p-value
s/so	place similarity only	−0·38	0·26	−1·46	0·14
	place and manner	0·86	0·09	9·50	<0·001
v/vo	place similarity only	1·95	0·32	6·00	<0·001
	place and manner	2·09	0·43	4·85	<0·001

Table II

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.

confirm the graphical impression (Table II). 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 likelihood of the CV form. Manner similarity did not give an additional boost to the CV form for [v].

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

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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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 [krˈest] 'cross' is one such example. The cluster [kr] has a steeply rising sonority profile and, accordingly, [kr]-initial words such as [krˈeslə] 'armchair', [kravˈatʲ] 'bed' and [krɨlˈo] 'wing' always appear with C prepositions. However, we find a substantial number of occurrences of [krˈest] with CV forms, as shown in (9).

(9) *CV preference in occurrences of* [krˈest]

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

The word [krˈest] is not an isolated case. Other examples are given in (10).

(10) *CV preference in other ecclesiastical words*

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

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).[6] 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 [g] 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 [ɣˈospədi] 'Lord' and [jej bˈoɣu] 'I swear to God' (Timberlake 2004: 23, Gouskova 2012). The phoneme [lˤ] in Arabic is another example: it only appears in the word [ʔalˤlˤaːh] 'God' (Ferguson 1956).

[6] 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 judgements.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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, [lʲef] can either mean 'lion' or be the name 'Lev', 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* [lʲef]

a.	s lʲvʲom	4%	'with the lion' ($n = 85$)
	sa lʲvʲom	96%	
b.	s lʲvʲom	55%	'with Lev' ($n = 73$)
	sa lʲvʲom	45%	

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 of distinguishing the two homophones (see §6.6 and Blumenfeld 2012).

Another case of homonym-related variation is [mnʲozʲəstvə]. Gribanova (2009a) notes that [s mnʲozʲəstvəm], with the C form of the preposition, means 'with a mathematical set', while [sa mnʲozəstvəm] means 'with a large amount'. We found some support for this characterisation, though it was not clear-cut. Out of the 340 occurrences of [s mnʲozʲəstvəm], six were used with the meaning 'with a set'.[7] Of the 561 occurrences of [sa mnʲozʲəstvəm], only one was used with this meaning. This asymmetry is statistically significant ($p = 0{\cdot}01$, Fisher's exact test), but it is not the case that [s mnʲozʲəstvəm] *has to* refer to mathematical sets. Alternatively, this may be a difference in register: formal registers favour CV forms more than informal ones (recall the ecclesiastic register effect in §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 [dvʲor] strongly favours CV prepositions throughout the paradigm (Fig. 2a; see also Matushansky 2002: n. 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 [dvʲigətilʲ] 'engine' and [dvizʲenijə] 'movement' invariably appear with C prepositions.

Most lexical items behave like [dvʲor], in that they show fairly consistent CV rates across inflected forms. There are, however, cases of variable behaviour within the inflectional paradigm of a word, which cannot be explained by the phonological properties of preposition–word junctures.

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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

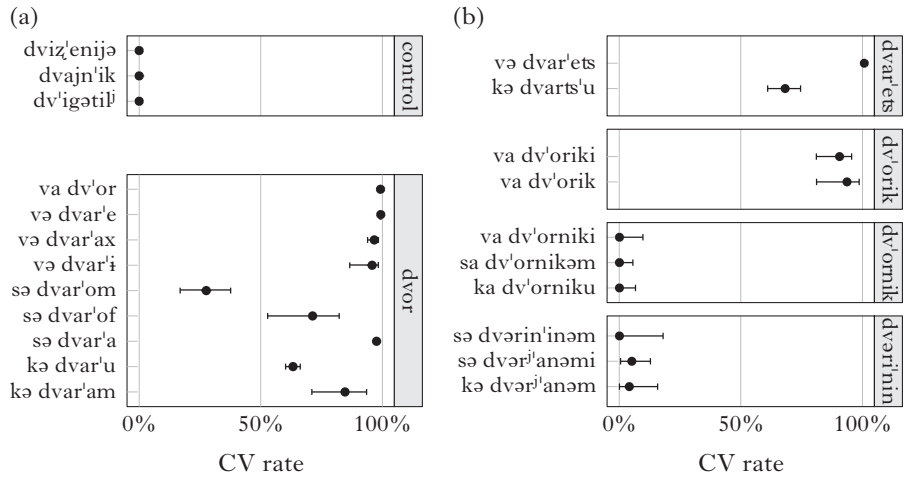


Figure 2

(a) Idiosyncratic CV-preference throughout the paradigm in [dv'or] 'yard'. Other lemmas beginning with [dv] virtually never occur with CV prepositions (the control items are collapsed across prepositions and cases), but [dv'or] takes the CV prepositions fairly consistently (seven out of the eight 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 §2.2. (b) Behaviour of words morphologically derived from [dv'or]. The words [dvar'ets] 'palace' and [dv'orik] 'yard-DIM' behave like the base form, taking the CV form exclusively after [v] and frequently after [k]. Conversely, [dv'ornik] 'janitor' and [dvərin'in] 'nobleman', also derived from [dv'or], almost always take C forms. This contrast shows that variation can be affected by suffixes.

This is particularly evident in the case of the two homophonous prepositions [s] in Russian, meaning 'with' and 'from', which take different case forms. Some nouns only show a CV preference with one of the homophones of the preposition: [sa dnʲ'a] 'from the day' but [z dnʲ'om] 'with the day' (Fig. 3a). The word [dn'o] 'bottom' behaves similarly (Fig. 3b).

Upon closer examination, it turns out that the difference between [sa dnʲ'a] and [z dnʲ'om] lies not in the homophone itself ('from' *vs.* 'with'), but in the sense of the word 'day' which is typically used with each homophone. The phrase [z dnʲ'om] 'with the day' occurs primarily in salutatory expressions referring to holidays (as in [pəzdravl'j'aju z dnʲ'om raʐd'enijə] 'I wish you a happy birthday', [z dnʲ'om pab'edɨ] 'happy Victory Day'). On the other hand, [sa dnʲ'a] 'from the day' usually refers to days in the general sense of the word (e.g. [sa dnʲ'a prinʲ'atijə kanstit'utsii] 'from the day of the enactment of the constitution'). The 'holiday' use is not limited to 'with': of the 19 hits for [z dnʲ'a] in the RNC,

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

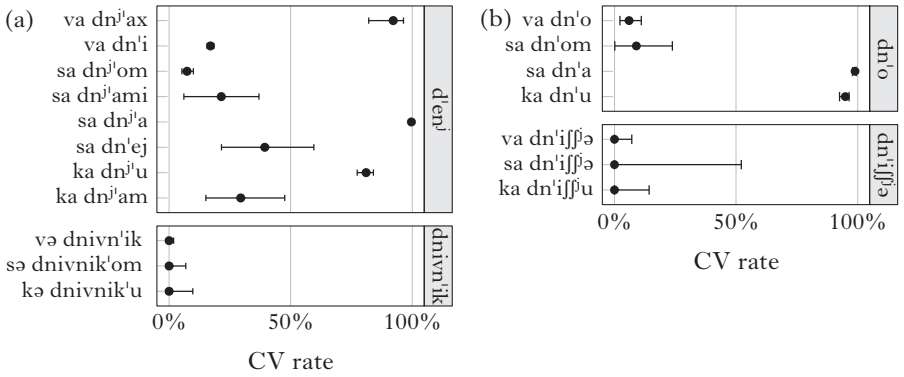


Figure 3

Variable CV rates within the paradigm: the two homophones [s] 'with' and [s] 'from' behave differently before both [dˈenʲ] ([dnʲ-]) 'day' and [dnˈo] 'bottom'. The morphologically derived [dnivnˈik] 'diary' (from [dˈenʲ]) and [dnˈiʃʲə] 'bottom of a ship' (from [dnˈo]) always take C forms, as expected from their phonological form.

seven are used in the birthday/holiday context; there are 'non-holiday' uses, but they come from older texts. Of the 50 uses of [sa dnʲˈom] in the corpus, one is a quote from an Old Church Slavonic text, and many refer to an ecclesiastical holiday ([sa dnʲˈom ˈangilə] 'with one's name day'). There are also uses in the birthday context. Thus, the 'holiday' use tends to favour the C preposition, whereas the 'non-holiday' use mostly conditions CV prepositions.

There is a similar asymmetry between [z dnˈom] 'with the bottom' and [sa dnˈa] '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 to be discussed below (see §3.2.5). It is possible that [dnˈo] 'bottom' is analogising to [dˈenʲ] 'day', which is considerably 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 (Fig. 2b). For instance, [dvˈor-ik] 'yard-DIM' has a strong CV preference, whereas [dvˈor-nik] 'janitor' does not. Certain suffixes (e.g. [-ik] (DIM)) have a systematic effect on the CV preference of derived words, either maintaining it, reducing it or eliminating it. For reasons of space, we do not discuss derivational context effects any further here; see Gouskova & Linzen (in preparation).

2.3.4 *Fixed expressions.* Certain words show a CV preference only in the context of a specific idiom (Blumenfeld 2011). This phenomenon appears

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

to be particularly common in the ecclesiastical register. The word [prˈax] 'ash', for example, begins with a rising sonority cluster that almost always favours the C form of prepositions. For example, the word [pridlazʲˈenijəm] 'proposal (INSTR)', 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, [prˈax] also behaves as expected from its initial cluster, taking the C form. With the preposition [v], however, almost 50% of the tokens in the Russian National Corpus are the CV form [va]. Closer inspection reveals that most of these tokens are part of the fixed expression [lizʲˈatʲ va prˈaxi] 'be dead (lit. lie in ashes)'. The fixed expression [prˈax k prˈaxu] '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 prˈaxu], and 100,000 for [k prˈaxu]).

A similar case is the word [plˈotʲ] 'flesh', which appears with a CV preposition in the idiom [və platˈi] 'in the flesh (fig)'. 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 plˈotʲju] 'with the flesh', not *[sa plˈotʲju]. Thus, unlike the two uses of [dnʲ] 'day' discussed in §2.3.3, the two allomorphs of the root [plˈotʲ] are not even fully homophonous: they have different properties with respect to stress and to the preposition phonology.

2.4 Sonority and lexical idiosyncrasy in monosyllabic yer stems

Vowel–zero alternations in Russian prepositions are sometimes analysed as a special case of a more general rule known as yer deletion (Lightner 1965, 1972, Halle 1973, Pesetsky 1979, Melvold 1989, Farina 1991, Yearley 1995, and many others). In §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 containing a yer vowel. We contrast the yer-deletion account with an alternative account, which attributes the behaviour 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 [lʲˈon] 'linen (NOM)', for example, deletes when a vowel-initial affix is added in the genitive: [lʲn-ˈa]. The alternation is analysed 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 & Becker 2013). In (12a) we list a few examples of monosyllabic yer nouns, and in (b) we list similar nouns that do not exhibit such alternations.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

(12) a.

nom sg	*gen pl*	
lˈef	lʲvˈof	‘lion’
rˈof	rvˈof	‘ditch’
ʂˈof	ʂvˈof	‘seam’
lʲˈon	lʲnˈof	‘linen’

b.

nom sg	*gen pl*	
lˈof	lˈovəf	‘catch’
rʲˈof	rʲˈovəf	‘roar’
ʂˈef	ʂˈefəf	‘chief’
lˈes	lisˈof	‘forest’

The dominant phonological analysis of yer deletion (Lightner 1972, Pesetsky 1979, Kenstowicz & Rubach 1987, Melvold 1989, Scheer 2006) marks alternating vowels as underlyingly different from regular vowels and posits a rule called Lower, which realises 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 underlyingly a yer (represented as /O/ in (13)); it triggers the realisation of the stem yer and then deletes. Non-yer suffixes such as the genitive plural condition yer deletion.

(13) *Traditional analysis: yers are only realised if followed by other yers* (*Lower rule*)

/rOt + O/ → rˈot ‘mouth (NOM SG)’
/rOt + ov/ → rtˈof ‘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 should be realised if the following vowel is also a yer. There are indeed some examples where this prediction is confirmed: words such as [rˈot] ‘mouth’ condition the CV form of alternating prepositions.[8]

(14) *Yer realisation in prepositions*

/sO + rOt + a/ → sa rtˈa ‘from the mouth’
/sO + rabˈot + ɨ/ → s rabˈotɨ ‘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 realised to prevent a geminate ([s+s]). Some examples of lexical variation are pointed out by Steriopolo (2007) and Blumenfeld (2012): there are yer monosyllables that condition C prepositions (e.g. [s psˈom] ‘with the dog’, [k pnʲˈu] ‘towards the stump’), and there are non-yer words that condition CV prepositions

[8] In order to work for cases such as [rˈot], the analysis requires the assumption that the preposition and the following noun are phonologised 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 §6.6 for discussion.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

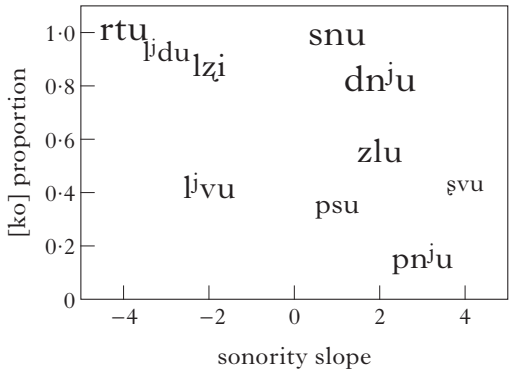


Figure 4

Sonority-slope and frequency effects on the likelihood of the CV preposition [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 ʂvˈu] 'towards the seam' to 570 for [k/ko dnʲˈu] 'towards the day'.

(Steriopolo cites [k rtˈuti] 'towards mercury', but notes that there is a [ka rtˈuti] 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 shape of the prepositions will depend on the following cluster. To test these predictions, we considered the preposition [ko]; the results are shown in Fig. 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 §2.1.3 for the sonority scale used here, and §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 Fig. 4, monosyllabic yer words indeed often appear with CV prepositions, though to different extents. For example, the form *[k rtˈu] 'towards the mouth' is categorically disallowed, the two forms [g zlˈu] and [ka zlˈu] 'towards evil' are equally frequent, and the C form [k pnʲˈu] 'towards the stump' is much more common than [ka pnʲˈu]. Words in our sample that started with a falling sonority cluster such as [rt] or [lʲd] were more likely to appear with the CV preposition [ko] than words starting with a rising sonority cluster.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

High-frequency words had a higher chance of being associated with CV prepositions; for example, the high-frequency form [dnʲˈu] 'day (DAT SG)' had a higher CV rate than [zlˈu] 'evil (DAT SG)' (medium frequency) and [psˈu] 'dog (DAT SG)' (low frequency) – even though their sonority slopes are similar. For most of the words in Fig. 4, the proportion of [ko] decreases as sonority slope increases – the outliers are [snˈu] 'sleep (DAT SG)' and [dnʲˈu], which are also the highest-frequency words (frequency is shown by font size).

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 the two forms of the preposition, [k] and [ko]. Log-transformed frequency was mildly correlated with sonority slope ($r = -0{\cdot}37$). To reduce collinearity in the model, we regressed out sonority slope from frequency. Residualised frequency was highly correlated with frequency ($r = 0{\cdot}92$), 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: [lʲvˈu] 'lion (DAT SG)', [lʲdˈu] 'ice (DAT SG)' and [snˈu] '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 ($p < 0{\cdot}001$ in both cases). The two factors interacted such that the sonority-slope effect was larger for more frequent words ($p = 0{\cdot}01$).

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 snˈu]) appears with CV more often than would be expected from its phonotactics. The categorical preference of [ka rtˈu] over *[k rtˈu] may have been a productive pattern in older stages of the language, but it is no longer, as indicated by the behaviour of roots such as [ʂv-] and [ps-]. There is no evidence suggesting that monosyllabic yer words condition prepositional vowel alternations in a way that is qualitatively different from other nouns: they condition CV forms based on their phonotactics and on lexically idiosyncratic CV preference.

In summary, this section has 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 generalisations as prepositions 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 realisation, which is a categorical phenomenon that applies across the board, but rather a combination of stochastic phonotactic rules and idiosyncratic properties of individual nouns.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

3 Corpus study 2

The results from corpus study 1 enabled us to discover new stochastic phonotactic 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.[9]

3.1 Methods

We generated a set of search queries using the methods described in §2.1. In contrast with the queries sent to the Russian National Corpus, we did not impose a predetermined 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 zʲ z l m n s f x ʦ ʧ 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.[10]

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 estimate the frequency more accurately 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 fewer 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 ten 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 1,120 tokens on average for the lemmas used in corpus study 2, compared to 14,760 tokens for the lemmas used in corpus study 1.

[9] Available at www.yandex.ru.

[10] The results can be downloaded as a CSV file from the online supplementary materials (see note 3).

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

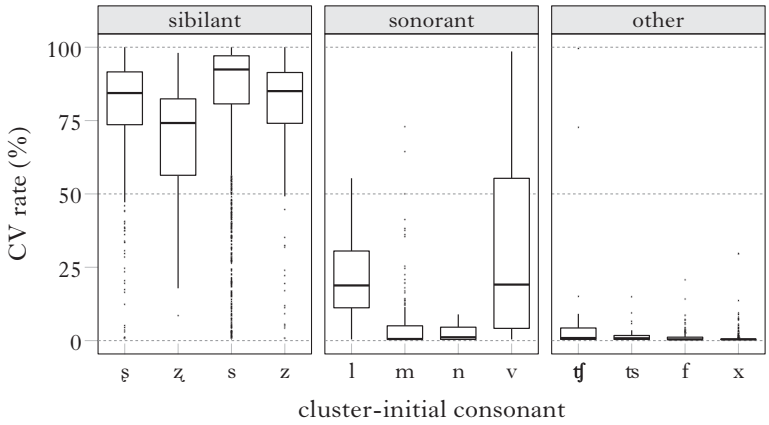


Figure 5
The probability of the CV form [so] in the Yandex corpus, by cluster-initial consonant.

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 in many phonological contexts we expect to find a great deal of variability. 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 Fig. 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 further 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. [fr] or [xv]) almost always condition the C form. There is more variability before sibilant-initial clusters, though the tendency is to favour the CV form. Finally, sonorant-initial clusters show a non-uniform pattern: the nasals [n] and [m] essentially pattern with non-sibilant obstruents in favouring the C form, whereas clusters that start with [v] and [l] show a great deal of variability.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

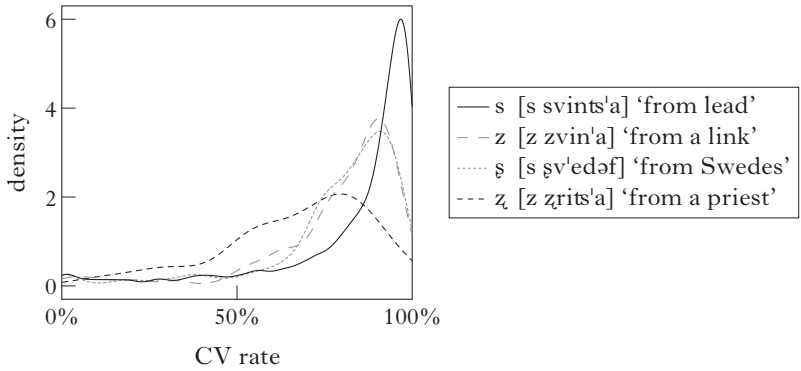


Figure 6

The probability of the CV form [so] before sibilant-initial clusters increases the more features are shared between that sibilant and [s].

3.2.1 *Variation among sibilants.* The next series of figures we present consists of kernel-density plots, rescaled and smoothed versions of histograms. 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, 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 §2.2.

This gradient similarity-avoidance effect is not unique to Russian. Pająk & Baković (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 (Fig. 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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

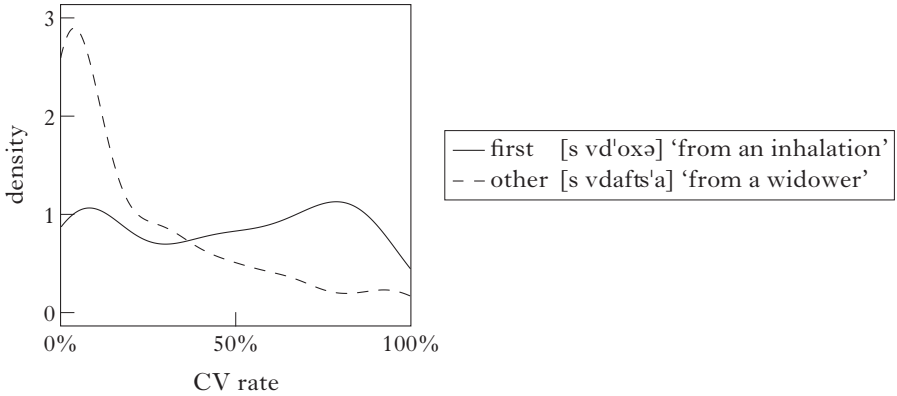


Figure 7
Words stressed on the first syllable are more likely to condition the CV form than words stressed on other syllables.

stress-initial words (i.e. [sa vd'oxə]) than in words with non-initial stress (e.g. [sə vdafts'a]). 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 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 *ǿ*] is more likely than [sə *ǿ*]. 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 centralised or bearing a high tone – than vowels in other positions (Crosswhite 1999, Padgett & Tabain 2005, Bethin 2006). Our analysis in §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.

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 non-sibilant obstruents in the sample as [vS], [vR] and [vK] respectively, as shown in (15).

(15) *Notation used for* vC *clusters*

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)

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

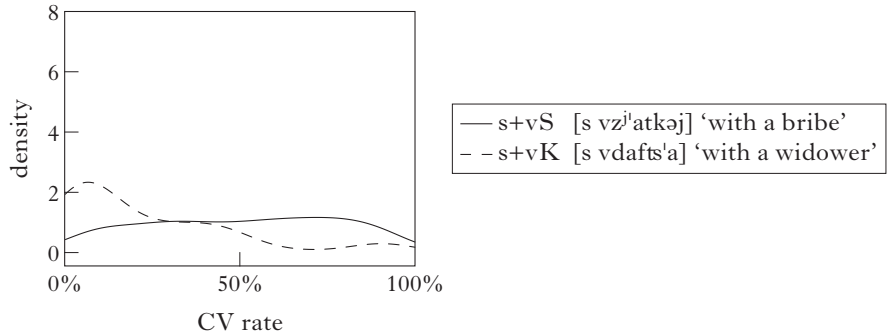


Figure 8
Non-adjacent OCP effect on the [s/so] alternation before [v]-initial clusters in the Yandex corpus: [vS] clusters condition the CV form more often than [vK] clusters.

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].

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 & 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 §2.1.3); [vf] is not a possible cluster, which leaves only [fx]. There is only one lemma in our data set that starts with this cluster: [fxazˌdˈenijə] '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 on the basis of 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 [dvˈor] and words derived from it (see Fig. 2). The same example also indicates that the two sibilants need not be adjacent for the OCP effect to hold.

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 Fig. 9, the likelihood of [so] increases as the second consonant in the cluster becomes less sonorous. In other words, [svC] clusters are more strongly dispreferred when there is a steeper drop in sonority between [v]

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

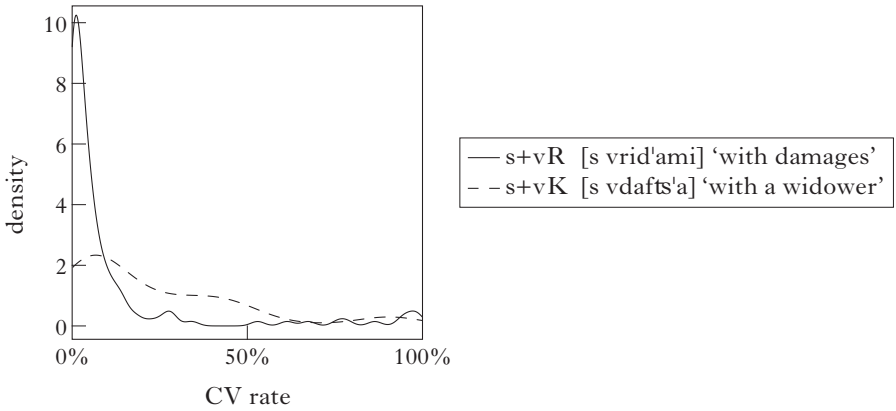


Figure 9
Sonority-slope effect on the [s/so] alternation before [v]-initial clusters: [vK] clusters condition the CV form more often than [vR] clusters.

and the following consonant. Words that start with a [vR] cluster, such as [vrid'ami] 'damages (INSTR)', almost always appear with [s]. On the other hand, there is more variation among words that start with a [vK] cluster, such as [vdafts'a] '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 §3.3 below).

3.2.5 *Morphology*. Recall from §2.3.3 that we found some differences between inflected forms in the same paradigm – for example, [z dnʲ'om] 'with the day (INSTR)' *vs.* [sa dnʲ'a] '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 10a 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 behaviour of the two homophonous prepositions. Similarly, CV rates do not seem to vary between singular and plural nouns (Fig. 10b).

3.3 Statistical analysis

We analysed the behaviour 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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

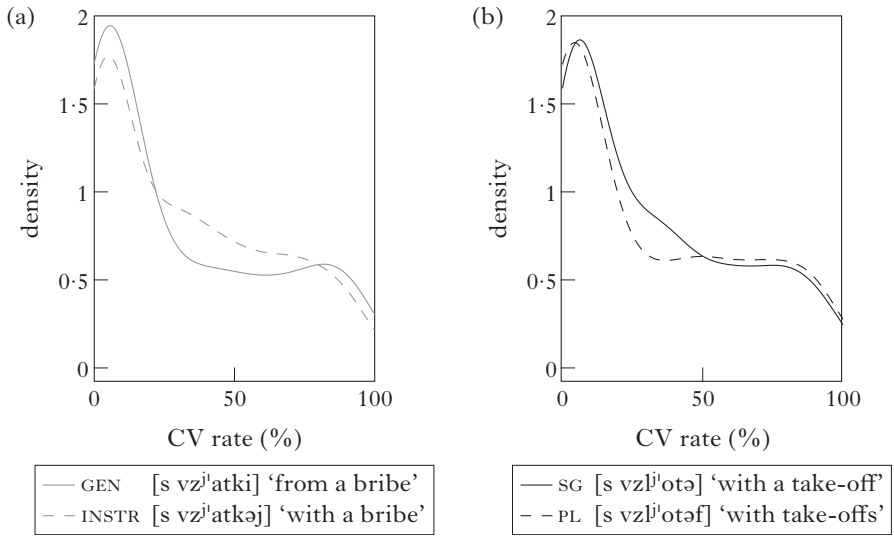


Figure 10
Effect of case and number on the [s/so] alternation before [v]-initial clusters.

(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 III. 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 ($t = 6{\cdot}25$, $p < 0{\cdot}001$), confirming the qualitative pattern shown in Fig. 8 above. The CV form was also less common before [vR] than [vK] clusters ($t = -2{\cdot}17$, $p = 0{\cdot}03$), 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 ($t = -2{\cdot}94$, $p = 0{\cdot}003$). We also found that longer words were less likely to condition the CV form ($t = -3{\cdot}44$, $p < 0{\cdot}001$). This effect is partially correlated with the effect of stress,

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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

predictor	coefficient	standard error	*t*	*p*-value
second is sibilant (vS)	1·50	0·24	6·25	<0·001***
second is sonorant (vS)	−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 (INSTR > GEN)	−0·12	0·25	−0·46	0·6
number (PL > SG)	−0·26	0·20	−1·29	0·2

Table III

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

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 n]) and liquids ([l r]). This analysis did not reveal any difference between the two types of sonorants ($t = 0{\cdot}1$, 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 barely 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).

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

4 A nonce-word experiment

Does the phonological pattern we identified constitute a productive phonological grammar? To address this question, we designed a nonce-word experiment. We used auditory stimuli to ascertain that the patterns

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

we found in the orthographic corpora extend to the spoken language. We set out to assess the productivity of the generalisations in (16).

(16) *Generalisations tested in the experiment*

a. *Adjacent OCP:* The CV form is strongly favoured 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 ʂrˈiməm] ≻ [s ʂrˈiməm].

b. *Non-adjacent OCP:* The CV form is favoured when the second consonant in the following cluster agrees in place and manner of articulation with the preposition: [sa fsˈiməm] ≻ [s fsˈiməm].

c. *Stress:* Stress-initial words favour the CV form more than stress-final words: [sa fsˈivəm] ≻ [sə fsivˈom].

d. *Sonority slope:* Falling sonority clusters favour the CV form more than rising sonority clusters: [sa rdˈiməm] ≻ [sa drˈiməm]. 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 rdˈiməm] ≻ [sa rnˈiməm].

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 gave the following information after completing the experiment: gender (53 females, 33 males), age (range 19–60, mean 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 ten clusters: the [v]-initial clusters [vz vd vn vs (pronounced as [fs])], the sibilant-initial clusters [ʂr ʂt], the fricative-initial cluster [xs] and the sonorant-initial clusters [rn rd ms]. There were eight items for each cluster, four of them with the vowel [u], and the remaining four with the vowel [i]. The rhymes for the items that had [u] as their vowel were randomly selected from the following list: [uf ub up um un uk ug]. Similarly, the rhymes for the items containing the vowel [i] were selected from [ib id iz iv in ix ig]. We chose high vowels because they show minimal phonetic variability based on stress position,

[12] We collected data from 90 participants, but excluded four from analysis because they did not report their age.

[13] The sites were odnoklassniki.ru, livejournal.com, vkontakte.ru and facebook.com.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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 eight control items starting with [xr] and eight starting with [k].

The corpus studies did not yield any evidence that morphological factors such as case and agreement features influence the alternation (§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, four items were stressed on the stem and four 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 running time of the experiment at around ten minutes. Each version of the experiment contained exactly one item from each combination of the twelve word-initial consonant sequences (see above), two stress positions (final and initial) and two vowels ([i] and [u]), totalling 48 items per version.

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 & Weenink 2009); a Russian-speaking experimenter then edited the files manually to make them sound more natural. The intensity of all the sound files was normalised by a Praat script.[14]

4.1.3 *Procedure.* The experiment was conducted on the web, using the Adobe Systems Flash platform. 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]. (17) gives an example of a trial.

[14] The files were converted to MP3 format using the LAME encoder (lame.sourceforge.net). The audio files can be found in the online supplementary materials (see note 3).

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

(17) *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 vzunom'

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'

2 [audio file]

Какое предложение кажется более правильным? [1] [2]
'Which sentence seems more correct?'

Participants were asked to choose which sound file they preferred. There were four training items, which they received feedback on. For example, they were told that most Russian listeners preferred [s tib-ˈom] to [sa tib-ˈom]. This was done to clarify the task to the listeners. The other training words were [(sa≻s) stˈid-əm], [(s≻sa) vˈum-əm] and [(sa≻s) zvˈuf-əm].

The nonce words were first presented to the participants in the context of one of the Type 1 frame sentences shown in (18a).

(18) *Frame sentences for words with and without prepositions*

a. *Type 1*
ja intirisˈujusʲ__ 'I am interested in'
ja xvˈastəjusʲ__ 'I boast of'
ja risˈuju__ 'I paint with'
ja davˈolʲnə__ 'I am pleased by'

b. *Type 2*
ja pridˈu s(o)__ 'I will come with'
ja gulʲˈaju s(o)__ 'I walk with'
ja druzˌˈu s(o)__ 'I am friends with'
ja igrˈaju s(o)__ 'I play with'

These sentences contained verbs that select for an object in the instrumental case, so as to familiarise the participants with the relevant form of the nonce word in a context where it was 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 Russian orthography. Next, the nonce word was auditorily presented in the context of one of the Type 2 frame sentences in (18b), 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 pridˈu] 'I will come') were presented visually, followed by a blank in place of the prepositional phrase.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

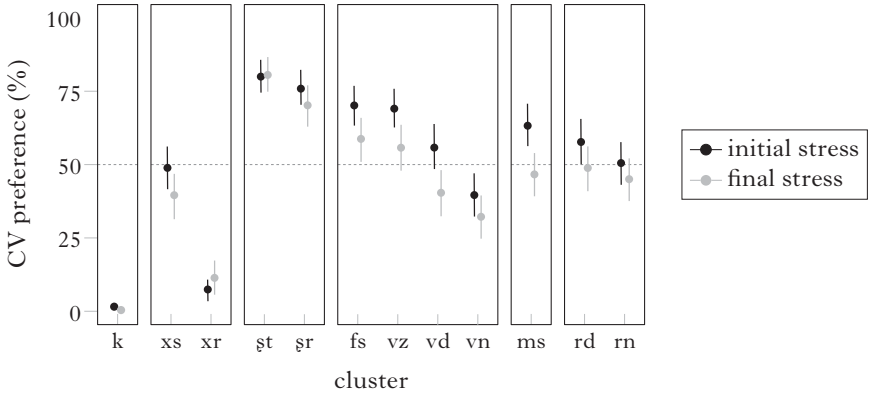


Figure 11
Experiment results: acceptability of [so + CC] broken down by CC cluster type and stress position. The dashed line at 50 % indicates chance performance (guessing).

The order in which the items were presented was randomised in such 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 randomised, with the constraint that the sound file containing the variant [s] was never the 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 kˈibəm] ≻ [sa kˈibəm]. 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 artefact 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.). The results are shown in Fig. 11.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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 ($p = 0{\cdot}25$). A by-subject slope for the stress predictor did improve the model fit ($p < 0{\cdot}001$), 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: $p = 0{\cdot}8$; sex: $p = 0{\cdot}76$; geographic region: $p = 0{\cdot}52$, model comparison).[15] The interaction between stress and cluster type did not reach significance ($p = 0{\cdot}06$, 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 IV.

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).

We confirmed the stress effect that we found in the Yandex corpus study (§3.2.2): participants preferred CV prepositions before words with initial stress ([sa vzˈibəm] ≻ [sə vzibˈom]). 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 before the rising sonority cluster [ʂr] ($p = 0{\cdot}01$). 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 ($p = 0{\cdot}09$). 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 ($p < 0{\cdot}001$). The effect of the sonority of the

[15] The lack of an age effect contrasts with two previous reports. Pająk & Baković (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.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

predictor	coefficient	standard error	t	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 IV

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].

first consonant in the cluster can be seen in a 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]: $p = 0{\cdot}001$; [xs] ≺ [fs]: $p < 0{\cdot}001$).

Participants accepted [so] before [vz] more often than before both [vd] and [vn] ($p < 0{\cdot}001$). 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. This pattern thus confirms an OCP effect between non-adjacent sibilants. The clusters [vz] and [fs] behaved identically ($p = 0{\cdot}52$), 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 draw stronger conclusions here.

There was a marked 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, disfavouring the cluster [s+xs].

The difference between [rn] and [vn] ($p < 0{\cdot}001$) is in the opposite direction than would be expected based on sonority slope: since 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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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 favoured in sibilant-initial clusters across the board, this tendency is even stronger when the OCP constraint is reinforced by initial stress and falling sonority.

5 Analysis: stochastic variation

The experiment established that several generalisations about Russian prepositional vowel–zero alternations apply 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 & Johnson 2003, Hayes & Wilson 2008). We start with the basic phonological question of whether the vowel–zero alternation is deletion or insertion (§5.1). §5.2 defines the constraints we use in our analysis, and §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 §5.4. The analysis presented in the present section deals with phonological variation; we take up lexical variation in §6.

5.1 Deletion or epenthesis?

Vowel–zero alternations can be analysed 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 ([bʲes ~ bʲezə] 'without', [ʧʲeris ~ ʧʲerizə] 'through', [pʲerit ~ pʲeridə] 'before', etc. – see Matushansky 2002, Gribanova 2009a).[16] The alternating vowels are

[16] We did not study these longer prepositions in any detail, but our impression is that they only have vowel–zero alternations when followed by function words

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

always [+back], unlike yer vowels in lexical morphemes, which can be either back or front (see §2.4.1 for a discussion of yer vowels). This predictable quality would follow from an epenthetic account. Moreover, Gribanova (2008: n. 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 & Baković (2010) analyse analogous alternations in Polish prepositions as epenthesis. Hayes (2009: 235ff) 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 & de Lacy 1999, Hall 2011). 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 2011). Even in Moscow Russian, prepositions have [o] forms in fixed expressions with preposition stress, such as [s'o svitu] 'from the world' and [v'o 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 analyse in a deletion account: assuming iambic footing (*σσ́*; Halle & Vergnaud 1987, Alderete 1999, Crosswhite 1999), the vowel is retained in a prominent footed position: e.g. /so vd'oxa/ → [(sa vd'o)xə]. It is deleted to improve the alignment of the lexical word and the phonological word: e.g. /so vdovetsa/ → [(s vdafts'a)]. Alternatively, the more prominent vowel [a] is less likely to delete than schwa, a typologically common pattern (Steriade 2001, Gouskova 2003, Howe & 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 favour the CV form of the prepositions rather than 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

(e.g. [bizə fs'ex] 'without all'; cf. (4)). Even when followed by [st] clusters, however, [bes] normally surfaces without a final vowel: [bis stak'anə] 'without a glass'.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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 & Kawahara 2013, *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 analyse the alternation. The syntactic distribution of prepositions suggests that they are clitics – they can attach to nouns, adjectives and adverbs (see Gouskova 2010 and Padgett 2012 for a discussion of the phonological properties of prepositions, and Sekerina 1997 and Fanselow & Ćavar 2002 on the syntactic behaviour 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 formalise this requirement as an alignment constraint, after Selkirk (1995), as in (19).

(19) ALIGN(PWd, L; LxWd, L)
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.)

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 2007). We assume that for non-alternating prepositions, MAX(V) outweighs ALIGN (see §6.2).

(20) MAX(V)
Assign a violation mark for every vowel in S1 (input) that has no correspondent in S2 (output). (McCarthy & 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 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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

violation score of 10, as shown in (21c). The sonority scale from (7) is repeated in (21b).

(21) a. *Sonority Sequencing Principle* (*SSP*)
Gradient HG formulation (Pater 2012; cf. Prince & Smolensky 1993)

b. *Sonority scale*
$v_1 > r_2 > l_3 > n_4 > s_5 > t_6$

c. *Scores for violations of the SSP by different clusters*

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

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 §6.2, indexing this constraint to individual morphemes allows the best modelling of 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 Steriopolo 2007, Gouskova 2012.)

We now turn to similarity avoidance. Our constraints need to capture the two facts in (23).

(23) a. Two sibilants are dispreferred within the same cluster.
b. Adjacent OCP is stronger than non-adjacent OCP ([sʂr] is worse than [srʂ]).

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

(24) OCP[sibilant] (*s(C)S)
Assign a violation mark for every pair of consonants in a cluster that have an identical specification for [sibilant].
(Penalises [ss], [sʂ], [zʐ], [sCs], [sCʂ], etc., but not [sVs], [sVʂ].)

We account for fact (b) by introducing a specific constraint against adjacent sibilants, defined in (25). The dispreference for adjacent sibilants followed by a consonant (e.g. [sst]) is extremely strong. Preliminary MaxEnt modelling revealed that this preference is not adequately captured by the additive effects of *#CCC and a general constraint against

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

two adjacent sibilants: either the dispreference is underestimated, or the general constraint against two adjacent sibilants is weighted very high, resulting in the undesirable penalisation 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.

(25) *SIMC-C
Assign a violation mark for every pair of adjacent sibilants if they are followed by another consonant.
(Penalises [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 behaviour of labial OCP discussed in §2.2. Likewise, we do not introduce a gradient OCP constraint penalising [s] + sibilant sequences in proportion to the featural overlap between [s] and that sibilant (see §2.2 and §3.2.1), since we only had one type of sibilant in our experiment ([ʂ]).

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.

(26) PARSE-σ
Assign a violation mark to every syllable that does not belong to a foot.
(McCarthy & Prince 1993)

5.3 A 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 jth candidate is determined by summing its violations for each constraint times the weight of the respective constraint, as in Harmonic Grammar (Legendre *et al*. 1990, Goldwater & Johnson 2003, Smolensky & Legendre 2006, Pater *et al*. 2007, Hayes & Wilson 2008). Formally, if w_i is the weight of the ith constraint and v_{ji} is the number of times that candidate j violates constraint i, then we have (27).

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

The probability of candidate j is then calculated as the exponent of H_j, normalised by the sum of the exponents of all harmony scores.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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

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.[17] The weights are shown in Table V, 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 V
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 (29), ALIGN favours deletion in all prepositions: [sa k'ibəm] 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 V predicts the input /so k'ib-om/ to map to [s k'ibəm] 99·8% of the time (calculated as $p_1 = e^0/(e^0 + e^{-6\cdot24}) = 0\cdot998$), compared to the 99% we observed in the experiment.

(29) *Deletion very likely before simple onsets*

/so k'ib-om/	ALIGN	*#CCC	*SIM C-C	*s(C)S	PARSE-σ	SSP	MAX	$\mathcal{H}$	*prob* (%)
s k'ibəm							−1×0	0	99·8
sa k'ibəm	−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 ʂt'ib-om/ (see (30)), the

[17] Available (August 2013) at www.linguistics.ucla.edu/people/hayes/Maxent GrammarTool. We assumed the default priors mean $\mu = 0\cdot0$, and variance $\sigma^2 = 100000\cdot0$ for the bias term.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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 ʂtˈibəm] 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 87·2 % of the time. In our experiment, people preferred the CV form 80 % of the time.

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

/so ʂtˈib-om/	Align	*#CCC	*Sim C-C	*s(C)S	Parse-σ	SSP	Max	$\mathcal{H}$	*prob* (%)
sa ʂtˈibəm	−1×6·24							−6·24	87·2
s ʂtˈibəm		−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 ʂrˈib-om/, deletion is predicted only 72 % of the time. This is shown in (31).

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

/so ʂrˈib-om/	Align	*#CCC	*Sim C-C	*s(C)S	Parse-σ	SSP	Max	$\mathcal{H}$	*prob* (%)
sa ʂrˈibəm	−1×6·24							−6·24	72·0
s ʂrˈibəm		−1×4·26	−1×1·72	−1×0·74		−2×0·24	−1×0	−7·20	28·0

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 rdˈib-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.

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

/so rdˈib-om/	Align	*#CCC	*Sim C-C	*s(C)S	Parse-σ	SSP	Max	$\mathcal{H}$	*prob* (%)
sa rdˈibəm	−1×6·24							−6·24	58·0
s rdˈibəm		−1×4·26				−9×0·24	−1×0	−6·42	42·0

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 Halle & Vergnaud 1987, Alderete 1999, Crosswhite 1999, Gouskova 2010, though cf. Revithiadou 1999, Bethin 2006). As shown in (33), the retained vowel cannot be footed in [sə (ʂribˈom)], so its deletion is more favoured in comparison with the footed vowel in [(sa ʂrˈi)bəm] in (31).

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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

/so ʂrib-ˈom/	ALIGN	*#CCC	*SIM C-C	*s(C)S	PARSE-σ	SSP	MAX	$\mathcal{H}$	*prob* (%)
sə (ʂribˈom)	−1×6·24				−1×0·37			−6·61	64·0
s (ʂribˈom)		−1×4·26	−1×1·72	−1×0·74		−2×0·24	−1×0	−7·20	36·0

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-ˈom/ are very close (see (34)), and the predicted frequencies (60 % and 40 %) match our experimental results.

(34) *Triconsonantal clusters with non-adjacent sibilants*

/so xsib-ˈom/	ALIGN	*#CCC	*SIM C-C	*s(C)S	PARSE-σ	SSP	MAX	$\mathcal{H}$	*prob* (%)
sə (xsibˈom)	−1×6·24				−1×0·37			−6·61	40·0
s (xsibˈom)		−1×4·26		−1×0·74		−5×0·24	−1×0	−6·20	60·0

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. The result in (35) is generated in the MaxEnt grammar.

(35) *No vowels in prepositions that precede* [sV]*-initial words*

/so sˈib-om/	ALIGN	*#CCC	*SIM C-C	*s(C)S	PARSE-σ	SSP	MAX	$\mathcal{H}$	*prob* (%)
(sa sˈi)bəm	−1×6·24				−1×0·37			−6·61	0·3
(s sˈibəm)				−1×0·74				−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 & Baković 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' (2010: 651). 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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

conspiracy to avoid preconsonantal geminates. Russian also has variable coronal place assimilation, but this analysis would not work for the language.[18] The problem is that Russian vowels show up even in clusters that cannot assimilate completely to make a geminate: labial–labial–C clusters such as [vmn] and [fpr] are also avoided (recall Fig. 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 which Polish lacks, [k], which exhibits sonority effects but not antigemination (e.g. /ko grob-u/ → [g grabˈu] '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 & Baković (2010) and Baković (2005) do shed light on a feature of the Russian alternation that we do not analyse 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 ʐritsom/ is less likely to have deletion than /so ʂvedov/ because [z/ʐ ʐritsˈom] requires more featural unfaithfulness than [s/ʂ ʂvˈedəf]. We do not analyse 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 ([ʂt] occurs with CV more often than [ʂr] 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-constraints theory allows us to capture these additive effects, while keeping the constraints themselves relatively simple compared to previous analyses in standard OT.

[18] 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 & Radišić 2009).

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

For example, to explain why the OCP applies only to [ssC] clusters rather than [ssV], Steriopolo (2007) assumes that preconsonantal sibilants are unsyllabified appendices rather than syllable onsets – a structure often posited for Slavic (Bethin 1992, Yearley 1995, among 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 [ʂr], and these are usually assumed to be tautosyllabic. The OCP applies across an intervening consonant (as in [sa ms'ibəm]), 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 & Rosenbach 2006, Hayes & Wilson 2008, Pater 2009, Potts *et al.* 2010).

As Boersma & Hayes (2001) point out, much of the success in modelling 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 analyse 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 & Hayes 2001), and standard OT can capture certain types of gang effects through local conjunction (Smolensky 1995). 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 & 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 (ʂrib'om)]. 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.[19] Implementing this

[19] The vowel that is deleting is not necessarily either [a] or [ə] underlyingly; it could be /o/, as we suggested in § 5.1. To work technically, this analysis would have to include

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

would require conjoining a constraint that specially protects [a] but not [ə] (MAX[a]) with each of the markedness constraints that disfavour deletion (*#CCC, *s(C)S, and so on). Recall tableaux (31) and (33): the goal is to make deletion slightly worse when it involves [a] rather than [ə], which means that MAX[a] would work in tandem with other constraints that disfavour 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 & Smolensky 2002). In a weighted-constraints theory, on the other hand, MAX[a] can interact with markedness constraints through cumulativity, so this analysis can be implemented. In (36), we show the weights of the constraints and the harmony scores for two inputs that differ in stress position. The crucial interaction is in (36b), 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.

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

	/so ʂrˈib-om/	ALIGN 6·28	*#CCC 3·93	*SIM C-C 1·72	*s(C)S 0·74	MAX [a] 0·37	SSP 0·24	MAX [ə] 0·00	$\mathcal{H}$	*prob* (%)
a.	sa ʂrˈibəm	−1×6·28							−6·28	72·0
	s ʂrˈibəm		−1×3·93	−1×1·72	−1×0·74	−1×0·37	−2×0·24		−7·24	28·0
b.	sə (ʂribˈom)	−1×6·28							−6·28	64·0
	s (ʂribˈom)		−1×3·93	−1×1·72	−1×0·74		−2×0·24	−1×0	−6·87	36·0

Analysed in this way, the Russian stress effect represents a possible 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 allows us to obtain a good fit to the experimental results, it does not suffice in the case of the near-categorical OCP constraint against adjacent sibilants in clusters. As mentioned in §5.2, modelling 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, so that the joint effect of

intermediate stages of representation where the prepositional vowels are present but reduced, as outlined by McCarthy (2011), with vowel-specific MAX constraints referring to reduced vowels rather than UR vowels.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

two constraints 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, [mnˈenijem] 'opinion (INSTR)' appeared with [s] almost exclusively, whereas [mnˈozʲəstvəm] 'a large amount (INSTR)' 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 [mnˈenijem], words that categorically occur with the CV form [so], such as [mnˈoj], and words that occur with both forms, such as [mnˈozʲəstvəm]. 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 [grˈex] 'sin' is 20%, whereas before [grˈat] 'city' it is 60%. This suggests that speakers 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 (Inkelas *et al.* 1996, Orgun 1996, Anttila 2002, Inkelas & Zoll 2007), lexical variation is analysed by assigning different constraint rankings to different morphemes. By contrast, the indexed-constraints approach (Benua 1997, Itô & 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 & Pater (2011) and Coetzee & Kawahara (2013)

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University 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 constraint is adjusted by an additive scaling factor, which is determined by extragrammatical properties of the word (e.g. its frequency). In the scaling factors 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 & Kawahara (2013) 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 constraints (Flack 2007, Gouskova 2007, Pater 2007). Finally, we introduce scaling factors that are lexically listed and vary idiosyncratically from morpheme to morpheme, in addition to scaling factors that reflect properties such as frequency and register. Using the notation introduced in §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 (37).

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

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 & Linzen in preparation).

6.2 A scaled-constraints 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] and [po], 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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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 §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].[20]

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 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 dvˈigətiʲə] 'from the engine' and the tendency to retain the vowel in [sə dvarˈa] 'from the yard' will be captured by assigning [dvˈor] a positive scaling factor for *#CCC.

	scaling factor	weight of *#CCC	CV rate
most words	0·00	4·26	depends on cluster
[grˈex]	0·85	0·85 + 4·26	20%
[grˈat]	1·70	1·70 + 4·26	60%
[dvˈor]	2·98	2·98 + 4·26	60–96%
[mnˈoj]	8·50	8·50 + 4·26	≈100%

Table VI
A scaling factors approach to gradient exceptionality.

[20] 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*.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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 subregularities (Zuraw 2000, Albright & Hayes 2003). Table VI gives 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 purposes of illustration).

The effect of scaling factors is illustrated in (38) for [dv'or], under the assumption that $s_{dvor} = 2{\cdot}98$.

(38) *A scaling factors treatment of lexical variation*

/so dvor-a/	ALIGN	*#CCC	*SIM C-C	*V(C) LAB	PARSE-σ	SSP	$\mathcal{H}$	*prob* (%)
sə (dvar'a)	−1×6·24				−1×0·37	0	−6·61	66·0
(zdvar'a)		$-1\times(s_{dvor}+4{\cdot}26) = -1\times7{\cdot}24$				0	−7·24	34·0

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, [dv'or] 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 [dv'or] 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 [dv'or] to derive this behaviour, as shown in (39).

(39) *The additive interaction between phonological and lexical variation*

/vo dvor-e/	ALIGN	$*\#CCC_{dvor}$	*SIM C-C	*V(C) LAB	PARSE-σ	SSP	$\mathcal{H}$	*prob* (%)
və (dvar'e)	−1×6·24				−1×0·37	0	−6·61	96·0
(vdvar'e)		$-1\times(s_{dvor}+4{\cdot}26) = -1\times7{\cdot}24$		−1×2·74		0	−7·98	4·0

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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

exhibit variation proportional to their scaling factor.[21] Word-initial CCC clusters can be created in several ways: through morpheme concatenation (/C + CC …/), deletion in the preposition (/CV + CC/ → #CCC) or deletion of a vowel morpheme-internally between three consonants (/CVCC/ → CCC … or /CCVC/ → CCC …). The latter kind of deletion is not found in Russian nouns – Gouskova & Becker (2012) show that CVCC words never display 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 §2.3.1). An overall register bias of this sort is straightforwardly analysed using the scaling factors 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 & Hayes 2001, Coetzee & Kawahara 2013). For illustration purposes, consider the word [krˈest] 'cross', which has a CV rate of 68%. Non-ecclesiastical words that start with the same cluster, such as [kravˈatʲ] '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 (40a) and (b) show how the difference between [krˈest] and the non-ecclesiastical [kravˈatʲ] can be accounted for by assuming that the register-wide scaling factor is $s_{ecc} = 3$.

(40) *Analysing the register effect using scaling factors*

	/so krest-om/	ALIGN	*#CCC	PARSE-σ	MAX(V)	SSP	$\mathcal{H}$	*prob* (%)
a.	sə (kristˈom)	−1×6·24		−1×0·37		0	−6·61	65·0
	(skristˈom)		−1×4·26		$-1\times(s_{k/s/v}+s_{ecc}+25)=-1\times3$	0	−7·26	35·0
b.	/so kravˈatʲ-ju/							
	sə (kravˈatʲ)ju	−1×6·24		−1×0·37		0	−6·61	8·0
	(skravˈatʲ)ju		−1×4·26		$-1\times(s_{k/s/v}+25)=-1\times0$	0	−4·26	92·0

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 [dvˈor]. If that is the case, the higher CV rate should be handled by morpheme-specific scaling of *#CCC, as in (38), rather than by register-wide scaling of MAX(V), as in (40a). 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 §4.

[21] We would like to thank the associate editor for pointing out this prediction.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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 §2.3.3). We will focus on the inflectional cases; for a full treatment of derivational effects, see Gouskova & 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 dnʲˈom] 'with the day (birthday)' but [sa dnʲˈa] 'from the day' (cf. Fig. 3). Similarly, the root /mnog-/ 'much' in [mnˈozˌəstvə] conditions slightly more CV prepositions when used to mean 'a large amount' and more C prepositions when used to mean 'a 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 (§3.2.5). The explanation we suggested for these cases is that a 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 dnʲˈom], behaves like most other [dn]-initial words. The allomorph selected in other contexts has a higher scaling factor for *#CCC, favouring CV prepositions. The context for allomorph selection is semantic/morphosyntactic (Bobaljik 2008), but it has consequences for the phonology. The analysis is otherwise no different, using a scaling factor for allomorphs just as for other exponents of morphemes, as shown in (41).

(41) *Allomorphy and variation in special semantic contexts*

a. $\sqrt{\text{DAY}} \leftrightarrow \begin{matrix}/\text{den}^{\text{j}}_{s(*\#\text{CCC})=0}/ \\ /\text{den}^{\text{j}}_{s(*\#\text{CCC})=2}/\end{matrix}$ / __ 'birthday'

b. $\sqrt{\text{MUCH}} \leftrightarrow \begin{matrix}/\text{mnog}_{s(*\#\text{CCC})=0}/ \\ /\text{mnog}_{s(*\#\text{CCC})=0{\cdot}5}/\end{matrix}$ / __ 'set'

6.5 Variation and frequency

We now turn to the role of word frequency in 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 & Kawahara 2013), Dutch schwa epenthesis (Tily & 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 a

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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 realisation of the vowel in CV prepositions, we cannot 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 realised 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 §2.4). These monosyllabic roots pattern with high-frequency function words, which always appear with CV prepositions ([sa mn'oj] 'with me' and [sa fs'em] '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 favour; as a consequence, they were not learned by the next 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 southern 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).[22] 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.

[22] We thank the associate editor for pointing out this fact.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

6.6 The 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 (Kiparsky 1982, 1985, Mohanan 1982, Kaisse & Shaw 1985), only lexical rules are allowed to have lexical exceptions, and only postlexical (e.g. phrasal) rules can be variable. Coetzee & 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 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 phonologised 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 cliticise phonologically 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. in Mende (Hayes 1990) and in Celtic languages (Green 2006)). Our corpus studies and experiment supply another case of a variable phonological rule that is part of phrasal/sentence phonology[23] and that shows lexical variability. This has consequences for the organisation 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. Bermúdez-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 has confirmed the similarity-avoidance patterns reported in previous studies: sequences of sibilants ([s+sC]) and labial continuants ([v+v/fC]) are systematically avoided in preconsonantal position, but adjacent dorsals are allowed

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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

([k+kC]). We also discovered three new generalisations 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+ʐ]). Second, the OCP applies to non-adjacent consonants in the same cluster: sequences such as [s+fs] and [v+dv] are disfavoured. 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 analysed these patterns in a weighted-constraints 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 analysed using the traditional yer rule: 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 realised. Most striking is the example of the root [dnʲ-] '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 could not be adequately characterised as an all-or-nothing distinction between regular and exceptional words. There is a wide range of CV rates, with some individual items favouring the CV form 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, [dvˈor] tends to favour [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.

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 generalisations. Comprehensive corpus

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

studies help to 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 standardised 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 with [ib] as their rhyme (see §4.1.2 for details).

input	candidate	probability	input	candidate	probability
/so vs'ibom/	s fs'ibom	0·30	/so ʂt'ibom/	s ʂt'ibom	0·20
	sa fs'ibom	0·70		sa ʂt'ibom	0·80
/so vsib'om/	s fsib'om	0·41	/so ʂtib'om/	s ʂtib'om	0·19
	sə fsib'om	0·59		sə ʂtib'om	0·81
/so k'ibom/	s k'ibom	0·99	/so vd'ibom/	s vd'ibom	0·44
	sa k'ibom	0·01		sa vd'ibom	0·56
/so kib'om/	s kib'om	1·00	/so vdib'om/	s vdib'om	0·60
	sə kib'om	0·00		sə vdib'om	0·40
/so ms'ibom/	s ms'ibom	0·37	/so vn'ibom/	s vn'ibom	0·60
	sa ms'ibom	0·63		sa vn'ibom	0·40
/so msib'om/	s msib'om	0·53	/so vnib'om/	s vnib'om	0·68
	sə msib'om	0·47		sə vnib'om	0·32
/so rd'ibom/	s rd'ibom	0·42	/so vz'ibom/	s vz'ibom	0·31
	sa rd'ibom	0·58		sa vz'ibom	0·69
/so rdib'om/	s rdib'om	0·51	/so vzib'om/	s vzib'om	0·44
	sə rdib'om	0·49		sə vzib'om	0·56
/so rn'ibom/	s rn'ibom	0·49	/so xr'ibom/	s xr'ibom	0·93
	sa rn'ibom	0·51		sa xr'ibom	0·07
/so rnib'om/	s rnib'om	0·55	/so xrib'om/	s xrib'om	0·89
	sə rnib'om	0·45		sə xrib'om	0·11
/so ʂr'ibom/	s ʂr'ibom	0·24	/so xs'ibom/	s xs'ibom	0·51
	sa ʂr'ibom	0·76		sa xs'ibom	0·49
/so ʂrib'om/	s ʂrib'om	0·30	/so xsib'om/	s xsib'om	0·60
	sə ʂrib'om	0·70		sə xsib'om	0·40

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

Appendix B: Glosses for forms in Figs. 1–4

Figure 1

fxˈot	‘entry’	zlˈo	‘evil’
fklʲutʃˈenii	‘turning on’	zdrˈavii	‘health’
fkˈus	‘taste’	xrˈam	‘temple’
vrˈet	‘damage’	slˈet	‘footprint’
vrˈemʲa	‘time’	nrˈaf	‘character’
vnimˈanii	‘attention’	grˈex	‘sin’
rvˈenii	‘yearning’	grˈat	‘city’
rtˈutʲ	‘mercury’	gnˈef	‘wrath’
mnˈozɣəstvə	‘large amount, set’	tʲmˈa	‘darkness’
krˈest	‘cross’	spasˈenii	‘salvation’
ftˈornik	‘Tuesday’	smirˈenii	‘humility’
flˈot	‘fleet’	prˈax	‘ash’
dvˈor	‘yard’	plˈotʲ	‘flesh’
dnˈo	‘bottom’	mrˈak	‘darkness’
dˈenʲ (dnʲ-)	‘day’	mnˈozɣəstvə	‘large amount’
vzvˈot	‘platoon’	mnˈenii	‘opinion’
vzrˈoslɨj	‘adult’	mgnavˈenii	‘moment’
vzrˈɨf	‘explosion’	mglˈa	‘mist’
vzglʲˈat	‘glance, look’	blˈagə	‘good, blessing’
vzdˈox	‘sigh’	svitˈoj	‘saint’
vzʲˈatii	‘capture’	sviʃʃˈenik	‘priest’
vzʲˈatək	‘bribe’	svidˈetilʲstvə	‘evidence’
vzˈor	‘gaze’	lˈev (lʲv-)	‘lion’
fstupˈlenii	‘entry’	dvarˈets	‘palace’
fstrˈetʃnɨj	‘coming from the opposite side’	dvˈorik	‘yard-DIM’
		dvˈor	‘yard’
		tsvˈet	‘colour’

Figure 2

dvizɣˈenijə	‘movement’	və dvarˈets	‘into the palace’
dvajnˈik	‘double’	kə dvartsˈu	‘towards the palace’
dvˈigətilʲ	‘engine’	va dvˈoriki	‘in the yard-DIM’
va dvˈor	‘into the yard’	va dvˈorik	‘into the yard-DIM’
və dvarˈe	‘in the yard’	va dvˈorniki	‘in the janitor’
və dvarˈax	‘in the yards’	sa dvˈornikəm	‘with the janitor’
və dvarˈɨ	‘into the yards’	ka dvˈorniku	‘towards the janitor’
sə dvarˈom	‘with the yard’	sə dvərinˈinəm	‘with the nobleman’
sə dvarˈof	‘from the yards’	sə dvərʲˈanəmi	‘with the noblemen’
sə dvarˈa	‘from the yard’	kə dvərʲˈanəm	‘towards the noblemen’
kə dvarˈu	‘towards the yard’		
kə dvarˈam	‘towards the yards’		

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

Figure 3

va dnʲˈax	‘in days’	va dnˈo	‘into the bottom’
va dnˈi	‘into days’	sa dnˈom	‘with the bottom’
sa dnʲˈom	‘with the day’	sa dnˈa	‘from the bottom’
sa dnʲˈami	‘with days’	ka dnˈu	‘towards the bottom’
sa dnʲˈa	‘from the day’	va dnˈiʃʲə	‘into the bottom of a ship’
sa dnˈej	‘from days’	sa dnˈiʃʲə	‘from the bottom of a ship’
ka dnʲˈu	‘towards the day’	ka dnˈiʃʲu	‘towards the bottom of a ship’
ka dnʲˈam	‘towards days’		
və dnivnˈik	‘into the diary’		
sə dnivnikˈom	‘with the diary’		
kə dnivnikˈu	‘towards the diary’		

Figure 4

[rtˈu]	‘mouth (DAT SG)’	[dnʲˈu]	‘day (DAT SG)’
[lʲdˈu]	‘ice (DAT SG)’	[zlˈu]	‘evil (DAT SG)’
[lzˌˈi]	‘lie (DAT SG)’	[psˈu]	‘dog (DAT SG)’
[lʲvˈu]	‘lion (DAT SG)’	[ʂvˈu]	‘seam (DAT SG)’
[snˈu]	‘sleep (DAT SG)’	[pnʲˈu]	‘stump (DAT SG)’

REFERENCES

Albright, Adam & 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*. PhD dissertation, University of Massachusetts, Amherst.

Anttila, Arto (2002). Morphologically conditioned phonological alternations. *NLLT* **20**. 1–42.

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

Baković, 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. Handout of paper presented at the 78th Annual Meeting of the Linguistic Society of America, Boston.

Bates, Douglas M., Martin Maechler & Ben Bolker (2012). Package ‘lme4’ (Version 0.999999-0): linear mixed-effects models using S4 classes. Available (August 2013) at cran.r-project.org/web/packages/lme4/lme4.pdf.

Beckman, Jill, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) (1995). *Papers in Optimality Theory*. Amherst: GLSA.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & 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). Affix classes are defined by faithfulness. *University of Maryland Working Papers in Linguistics* **5**. 1–26.

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

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

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

Bermúdez-Otero, Ricardo (2012). The architecture of grammar and the division of labor in exponence. In Jochen Trommer (ed.) *The morphology and phonology of exponence*. Oxford: Oxford University Press. 8–83.
Bethin, Christina Y. (1992). *Polish syllables : the role of prosody in phonology and morphology*. Columbus: Slavica.
Bethin, Christina Y. (2006). Stress and tone in East Slavic dialects. *Phonology* **23**. 125–156.
Blumenfeld, Lev (2012). Vowel–zero alternations in Russian prepositions: prosodic constituency and productivity. In Veronika Makarova (ed.) *Russian language studies in North America : new perspectives from theoretical and applied linguistics*. London & New York: Anthem Press. 43–69.
Bobaljik, Jonathan David (2008). Paradigms (optimal and otherwise): a case for skepticism. In Asaf Bachrach & Andrew Nevins (eds.) *Inflectional identity*. Oxford: Oxford University Press. 29–54.
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 from the Rutgers Optimality Archive.
Boersma, Paul & Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *LI* **32**. 45–86.
Boersma, Paul & David Weenink (2009). *Praat : doing phonetics by computer* (version 5.1.34). www.praat.org.
Borowsky, Toni, Shigeto Kawahara, Takahito Shinya & Mariko Sugahara (eds.) (2012). *Prosody matters : essays in honor of Elisabeth Selkirk*. London: Equinox.
Bybee, Joan (2000). The phonology of the lexicon: evidence from lexical diffusion. In Michael Barlow & Suzanne Kemmer (eds.) *Usage-based models of language*. Stanford: CSLI. 65–85.
Bybee, Joan (2006). From usage to grammar: the mind's response to repetition. *Lg* **82**. 711–733.
Canty, Angelo & B. D. Ripley (2012). *boot : Bootstrap R (S-Plus) functions*. R package version 1.3–4. Available (August 2013) at cran.r-project.org/web/packages/boot/boot.pdf.
Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York: Harper & Row.
Clopper, C. J. & E. S. Pearson (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**. 404–413.
Coetzee, Andries W. & Shigeto Kawahara (2013). Frequency biases in phonological variation. *NLLT* **31**. 47–89.
Coetzee, Andries W. & Joe Pater (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *NLLT* **26**. 289–337.
Coetzee, Andries W. & Joe Pater (2011). The place of variation in phonological theory. In John Goldsmith, Jason Riggle & Alan Yu (eds.) *The handbook of phonological theory*. 2nd edn. Malden, Mass. & Oxford: Wiley-Blackwell. 401–431.
Crosswhite, Katherine (1999). *Vowel reduction in Optimality Theory*. PhD dissertation, University of California, Los Angeles.
Czaykowska-Higgins, Ewa (1988). *Investigations into Polish morphology and phonology*. PhD dissertation, MIT.
Dalby, Jonathan M. (1984). *Phonetic structure of fast speech in American English*. PhD dissertation, Indiana University.
Davidson, Lisa & Kevin Roon (2008). Durational correlates for differentiating consonant sequences in Russian. *Journal of the International Phonetic Association* **38**. 137–165.
Davison, A. C. & D. V. Hinkley (1997). *Bootstrap methods and their applications*. Cambridge: Cambridge University Press.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

Fanselow, Gisbert & Damir Ćavar (2002). Distributed deletion. In Artemis Alexiadou (ed.) *Theoretical approaches to universals*. Amsterdam & Philadelphia: Benjamins. 65–107.
Farina, Donna M. (1991). *Palatalization and jers in modern Russian phonology : an underspecification approach*. PhD dissertation, University of Illinois.
Farris-Trimble, Ashley W. (2008). Cumulative faithfulness effects: opaque or transparent? *Indiana University Working Papers in Linguistics* **6**. 119–145.
Ferguson, Charles A. (1956). The emphatic *l* in Arabic. *Lg* **32**. 446–452.
Flack, Kathryn (2007). Templatic morphology and indexed markedness constraints. *LI* **38**. 749–758.
Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe (2004). Similarity avoidance and the OCP. *NLLT* **22**. 179–228.
Gafos, Adamantios I. (1999). *The articulatory basis of locality in phonology*. New York: Garland.
Gallagher, Gillian & Jessica Coon (2009). Distinguishing total and partial identity: evidence from Chol. *NLLT* **27**. 545–582.
Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenador, Anders Eriksson & Östen Dahl (eds.) *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. Stockholm: Stockholm University. 111–120.
Gouskova, Maria (2003). *Deriving economy : syncope in Optimality Theory*. PhD dissertation, University of Massachusetts, Amherst.
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* **27**. 387–448.
Gouskova, Maria (2012). Unexceptional segments. *NLLT* **30**. 79–133.
Gouskova, Maria & Michael Becker (2013). Nonce words show that Russian yer alternations are governed by the grammar. *NLLT* **31**. 735–765.
Gouskova, Maria & Tal Linzen (in preparation). Morphological conditioning of phonological regularization. Ms, New York University.
Green, Anthony D. (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. *WCCFL* **26**. 217–225.
Gribanova, Vera (2009a). Phonological evidence for a distinction between Russian prepositions and prefixes. In Gerhild Zybatow, Uwe Junghanns, Denisa Lenertová & Petr Biskup (eds.) *Studies in formal Slavic phonology, morphology, syntax, semantics and information structure : Proceedings of the 7th European Conference on Formal Description of Slavic Languages*. Frankfurt am Main: Lang. 383–396.
Gribanova, Vera (2009b). The phonology and syntax of sub-words. Paper presented at GLOW 32, Nantes.
Gribanova, Vera (2010). *Composition and locality : the morphosyntax and phonology of the Russian verbal complex*. PhD dissertation, University of California, Santa Cruz.
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**. 223–239.
Guy, Gregory R. (2007). Lexical exceptions in variable phonology. *University of Pennsylvania Working Papers in Linguistics* **13:2**. 109–120.
Hall, Nancy (2011). Vowel epenthesis. In van Oostendorp *et al*. (2011). 1576–1596.
Halle, Morris (1973). The accentuation of Russian words. *Lg* **49**. 312–348.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

Halle, Morris & Jean-Roger Vergnaud (1987). *An essay on stress*. Cambridge, Mass.: MIT Press.
Hayes, Bruce (1984). The phonetics and phonology of Russian voicing assimilation. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*. Cambridge, Mass.: MIT Press. 318–328.
Hayes, Bruce (1990). Precompiled phrasal phonology. In Sharon Inkelas & Draga Zec (eds.) *The phonology–syntax connection*. Chicago: University of Chicago Press. 85–108.
Hayes, Bruce (2009). *Introductory phonology*. Malden, Mass. & Oxford: Wiley-Blackwell.
Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *LI* **39**. 379–440.
Hooper, Joan B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In William M. Christie, Jr. (ed.) *Current progress in historical linguistics*. Amsterdam: North Holland. 95–105.
Howe, Darin & Douglas Pulleyblank (2004). Harmonic scales as faithfulness. *Canadian Journal of Linguistics* **49**. 1–49.
Inkelas, Sharon, C. Orhan Orgun & Cheryl Zoll (1996). Exceptions and static phonological patterns: cophonologies vs. prespecification. Ms, University of California, Berkeley & University of Iowa. Available as ROA-124 from the Rutgers Optimality Archive.
Inkelas, Sharon, C. Orhan Orgun & Cheryl Zoll (1997). The implications of lexical exceptions for the nature of grammar. In Iggy Roca (ed.) *Derivations and constraints in phonology*. Oxford: Clarendon. 393–418.
Inkelas, Sharon & Cheryl Zoll (2007). Is grammar dependence real? A comparison between cophonological and indexed constraint approaches to morphologically conditioned phonology. *Linguistics* **45**. 133–171.
Itô, Junko & Armin Mester (1999). The phonological lexicon. In Natsuko Tsujimura (ed.) *The handbook of Japanese linguistics*. Malden, Mass. & Oxford: Blackwell. 62–100.
Ito, Junko & Armin Mester (2003). On the sources of opacity in OT: coda processes in German. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*. Cambridge: Cambridge University Press. 271–303.
Jäger, Gerhard & Anette Rosenbach (2006). The winner takes it all – almost: cumulativity in grammatical variation. *Linguistics* **44**. 937–971.
Kaisse, Ellen M. & Patricia A. Shaw (1985). On the theory of Lexical Phonology. *Phonology Yearbook* **2**. 1–30.
Katz, Jonah (2005). Russian consonant Cvljusters. Ms, MIT. Available (August 2013) at web.mit.edu/jikatz/www/KatzPhonSquib.pdf.
Keller, Frank (2000). *Gradience in grammar : experimental and computational aspects of degrees of grammaticality*. PhD dissertation, University of Edinburgh.
Kenstowicz, Michael & Jerzy Rubach (1987). The phonology of syllabic nuclei in Slovak. *Lg* **63**. 463–497.
Kiparsky, Paul (1982). Lexical phonology and morphology. In The Linguistic Society of Korea (ed.) *Linguistics in the morning calm*. Seoul: Hanshin. 3–91.
Kiparsky, Paul (1985). Some consequences of Lexical Phonology. *Phonology Yearbook* **2**. 85–138.
Kitto, Catherine & Paul de Lacy (1999). Correspondence and epenthetic quality. *Toronto Working Papers in Linguistics* **16:2**. 181–200.
Kochetov, Alexei & Milica Radišić (2009). Latent consonant harmony in Russian: experimental evidence for Agreement by Correspondence. In Maria Babyonyshev, Darya Kavitskaya & Jodi Reich (eds.) *Formal approaches to Slavic linguistics : the Yale Meeting 2008*. Ann Arbor: Michigan Slavic Publications. 111–130.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

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, William (1973). *Suprasegmental phonology*. PhD dissertation, MIT.
Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky (1990). Harmonic Grammar: a formal multi-level connectionist theory of linguistic well-formedness: an application. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*. Hillsdale: Erlbaum. 884–891.
Lightner, Theodore M. (1965). *Segmental phonology of Modern Standard Russian*. PhD dissertation, MIT.
Lightner, Theodore M. (1972). *Problems in the theory of phonology*. Vol. 1: *Russian phonology and Turkish phonology*. Edmonton: Linguistic Research.
McCarthy, John J. (1986). OCP effects: gemination and antigemination. *LI* **17**. 207–263.
McCarthy, John J. (2011). Perceptually grounded faithfulness in Harmonic Serialism. *LI* **42**. 171–183.
McCarthy, John J. & Alan Prince (1993). Generalized alignment. *Yearbook of Morphology 1993*. 79–153.
McCarthy, John J. & Alan Prince (1995). Faithfulness and reduplicative identity. In Beckman *et al*. (1995). 249–384.
Marantz, Alec (1997). No escape from syntax: don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* **4:2**. 201–225.
Marantz, Alec (2008). Phases and words. Ms, New York University.
Matushansky, Ora (2002). On formal identity of Russian prefixes and prepositions. *MIT Working Papers in Linguistics* **42**. 217–253.
Melvold, Janis L. (1989). *Structure and stress in the phonology of Russian*. PhD dissertation, MIT.
Mohanan, K. P. (1982). *Lexical Phonology*. PhD dissertation, MIT. Distributed by Indiana University Linguistics Club.
Moreton, Elliott & Paul Smolensky (2002). Typological consequences of local constraint conjunction. *WCCFL* **21**. 306–319.
Neu, Helene (1980). Ranking on 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. Available (August 2013) at languagelog.ldc.upenn.edu/nll/?p = 1943.
Ohala, John J. & Haruko Kawasaki-Fukumori (1997). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In Stig Eliasson & Ernst Håkon Jahr (eds.) *Language and its ecology : essays in memory of Einar Haugen*. Berlin & New York: Mouton de Gruyter. 343–365.
Oostendorp, Marc van, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.) (2011). *The Blackwell companion to phonology*. Malden, Mass.: Wiley-Blackwell.
Orgun, C. Orhan (1996). *Sign-based morphology and phonology with special attention to Optimality Theory*. PhD dissertation, University of California, Berkeley.
Padgett, Jaye (2002). Russian voicing assimilation, final devoicing, and the problem of [v] (*or*, The mouse that squeaked). Ms, University of California, Santa Cruz.
Padgett, Jaye (2008). Glides, vowels, and features. *Lingua* **118**. 1937–1955.
Padgett, Jaye (2012). The role of prosody in Russian voicing. In Borowsky *et al*. (2012). 181–207.
Padgett, Jaye & Marija Tabain (2005). Adaptive Dispersion Theory and phonological vowel reduction in Russian. *Phonetica* **62**. 14–54.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press

Pająk, Bożena & Eric Baković (2010). Assimilation, antigemination, and contingent optionality: the phonology of monoconsonantal proclitics in Polish. *NLLT* **28**. 643–680.
Pater, Joe (2000). Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints. *Phonology* **17**. 237–274.
Pater, Joe (2007). The locus of exceptionality: morpheme-specific phonology as constraint indexation. In Leah Bateman, Michael O'Keefe, Ehren Reilly & Adam Werle (eds.) *Papers in Optimality Theory III*. Amherst: GLSA. 259–296.
Pater, Joe (2009). Weighted constraints in generative linguistics. *Cognitive Science* **33**. 999–1035.
Pater, Joe (2012). Serial Harmonic Grammar and Berber syllabification. In Borowsky *et al*. (2012). 43–72.
Pater, Joe, Christopher Potts & Rajesh Bhatt (2007). Harmonic Grammar with Linear Programming. Ms, University of Massachusetts, Amherst.
Patterson, David, Paul C. LoCasto & Cynthia M. 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 & Michael Becker (2010). Harmonic Grammar with linear programming: from linear systems to linguistic typology. *Phonology* **27**. 77–117.
Prince, Alan & Paul Smolensky (1993). *Optimality Theory : constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published 2004, Malden, Mass. & Oxford: Blackwell.
R Development Core Team (2012). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at www.r-project.org.
Raymond, William D., Robin Dautricourt & 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**. 55–97.
Revithiadou, Anthi (1999). *Headmost accent wins : head dominance and ideal prosodic form in lexical accent systems*. PhD dissertation, University of Leiden.
Ryan, Kevin (to appear). Onsets contribute to syllable weight: statistical evidence from stress and meter. *Lg*.
Scheer, Tobias (2006). How yers made Lightner, Gussmann, Rubach, Spencer and others invent CVCV. In Piotr Bański, Beata Łukaszewicz & Monika Opalinska (eds.) *Studies in constraint-based phonology*. Warsaw: Wydawnictwa Uniwersytetu Warszawskiego. 133–207.
Sekerina, Irina (1997). *The syntax and processing of scrambling constructions in Russian*. PhD dissertation, City University of New York.
Selkirk, Elisabeth (1995). The prosodic structure of function words. In Beckman *et al*. (1995). 439–469.
Sharoff, Serge (2005). Methods and tools for development of the Russian Reference Corpus. In Andrew Wilson, Dawn Archer & Paul Rayson (eds.) *Corpus linguistics around the world*. Amsterdam & New York: Rodopi. 167–180.
Smolensky, Paul (1995). On the internal structure of the constraint component *Con* of UG. Handout of talk presented at the University of Arizona.
Smolensky, Paul & Géraldine Legendre (eds.) (2006). *The harmonic mind : from neural computation to optimality-theoretic grammar*. 2 vols. Cambridge, Mass.: MIT Press.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University 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 Richard Compton, Magdalena Goledzinowska & Ulyana Savchenko (eds.) *Formal approaches to Slavic linguistics : the Toronto Meeting 2006*. Ann Arbor: Michigan Slavic Publications. 365–385.
Svenonius, Peter (2004). Slavic prefixes and morphology: an introduction to the *Nordlyd* volume. *Nordlyd* **32**. 177–204.
Tily, Harry & Victor Kuperman (2012). Rational phonological lengthening in spoken Dutch. *JASA* **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 *на́ день*, *за́ сорок*, *не́ был* in Modern Russian. *Russian Linguistics* **22**. 287–319.
Usachev, Andrei (2004). *Fully accented paradigms from Zaliznjak's (1977) grammatical dictionary*. File available (August 2013) at www.speakrus.ru/dict/all_forms.rar.
Vlasto, A. P. (1986). A linguistic history of Russia to the end of the eighteenth century. Oxford: Clarendon.
Yaschenko, I. V. (2002). *Paradoksy teorii mnozhestv*. [*Paradoxes of set theory*.] Moscow: Izdatelsvo Moskovskogo Tsentra Nepreryvnogo Matematicheskogo Obrazovania.
Yearley, Jennifer (1995). Jer vowels in Russian. In Beckman *et al*. (1995). 533–571.
Zaliznjak, A. A. (1977). *Grammatičeskij slovar' russkogo jazyka*. [*A grammatical dictionary of the Russian language*.] Moscow: Izdatelstvo Russkij Jazyk.
Zaliznjak, A. A. (1985). Ot praslavjanskoj akcentuacii k russkoj. [From Proto-Slavic to Russian accentuation.] Moscow: Nauka.
Zuraw, Kie (2000). *Patterned exceptions in phonology*. PhD dissertation, University of California, Los Angeles.

https://doi.org/10.1017/S0952675713000225 Published online by Cambridge University Press