Capturing Correlational Structure in Russian Paradigms: a Case Study in Logistic Mixed-Effects Modeling

Laura A. Janda[1], Tore Nesset[2] & R. Harald Baayen[3]
University of Tromsø[1,2] and University of Alberta[3]
August 17, 2009

Abstract

This study addresses the statistical analysis of a phenomenon in Russian verbal paradigms, a suffix shift that is spreading through the paradigm and making it more regular. A problem that arises in the analysis of data collected from the Russian National Corpus is that counts documenting this phenomenon are based on repeated observations of the same verbs, and, moreover, on counts for different parts of the paradigms of these same verbs. Unsurprisingly, individual verbs display consistent (although variable) behavior with respect to the suffix shift. The non-independence of the elementary observations in our data has to be taken into account in the statistical evaluation of the patterns in the data. We show how mixed-effects modeling can be used to do this in a principled way, and that it is also necessary to do so in order to avoid anti-conservative evaluation of significance.

1 Introduction

A group of Russian verbs is undergoing a diachronic change in which the suffix *-a* is being replaced by the productive suffix *-aj*. The Russian suffix shift is recognized in reference works such as Zaliznjak (1977) and Švedova (1980), and has been investigated in the contexts of language acquisition, psycholinguistics, stylistic variation, sociolinguistics and dialectology (cf. e.g. Andersen, 1980; Gagarina, 2003; Gor and Chernigovskaya, 2001, 2003a,c,b; Kiebzak-Mandera et al., 1997; Krysin, 1974; Tkachenko and Chernigovskaya, 2006). The suffix shift is evident in present tense, imperative, present active participle and gerund (verbal adverb) paradigm slots, where the *-a* suffixed forms show suffix truncation usually accompanied by alternation of the root final consonant, whereas the *-aj* suffixed forms lack such alternations. This Russian suffix shift thus yields a regularization among verbs comparable to the shift of English verbs from the weak to strong pattern. Table 1 presents the relevant forms of *maxat′* 'wave', showing that the *-a* suffixed forms have a $x \sim \check{s}$ alternation, while the *-aj* suffixed forms preserve both the suffix and the x throughout the present paradigm (phonemically, there is a /j/ between the vowels in the orthographic sequences in the 2sg, 3sg, 1pl and 2pl forms). This table also includes the infinitive and masculine singular past forms in addition to the forms relevant to the suffix shift.

Corpus data show that the Russian suffix shift is not taking place uniformly, but is dependent upon two factors: paradigm slot and root final consonant. Verbs undergoing the Russian suffix shift have root final consonants with three different places of articulation: labial, which most favors the innovative *-aj* suffix; dental, which most favors the conservative *-a* suffix; and velar, which is intermediate in implementation of the suffix shift. Turning to the paradigm slots, the gerund appears to be the most innovative in replacing *-a* with *-aj*

Table 1: Forms of *maxatʹ*, 'wave'.

	forms suffixed with *-a*	form suffixed with *-aj*
infinitive	*maxatʹ*	*maxatʹ*
masculine sg past	*maxal*	*maxal*
1sg present	*mašu*	*maxaju*
2sg present	*mašešʹ*	*maxaešʹ*
3sg present	*mašet*	*maxaet*
1pl present	*mašem*	*maxaem*
2pl present	*mašete*	*maxaete*
3pl present	*mašut*	*maxajut*
imperative	*maši(te)*	*maxaj(te)*
present active participle	*mašuščij*	*maxajuščij*
gerund	*maša*	*maxaja*

approximately 50% of the time, and other relevant forms follow a cline, ending with the 3sg present as the most conservative form, resisting suffix shift by maintaining *-a* most strongly.

Our hypothesis is that prototypicality plays a major role in the ordering of paradigm slots. Nesset & Janda (in prep.) discuss in more detail the Paradigm Structure Hypothesis, according to which paradigms have the structure of radial categories with a central prototype related to more peripheral members. The known markedness and prototypicality relationships among members of the verbal paradigm make it possible to establish the following structure, with more prototypical members toward the left:

3sg > 3pl > 1&2 > Imperative > Gerund/Participle.

This hypothesis predicts that the most prototypical forms resist the suffix shift, while the less prototypical forms are more likely to implement it. The model presented in this paper shows that paradigm slot is indeed a robust predictor of the implementation of language change, and the overall order of the slots is confirmed with the exception of the participle. Interestingly, the present active participle is a "parasitic" form derived from the 3pl form. This formal relationship may have reduced the implementation of suffix shift among participles. In sum, the language change documented here provides empirical evidence for the internal structure of paradigms since this language change does not take place uniformly, but is most pronounced among the peripheral forms of a paradigm.

The issue addressed in the present study is what the best way is to analyse counts of *-a* and *-aj* in the Russian National Corpus (`www.ruscorpora.ru`), obtained for a number of different verbs with varying root-final consonants across different paradigm slots.

The Russian National Corpus contains approximately 140 million words collected from a wide variety of genres and authors. Though the bulk of material is written and recent (post 1950), spoken Russian and earlier sources are also represented. Unlike the BNC, the RNC contains entire works instead of excerpts. Search options make it possible to target lexical items, morphological forms, and semantic groupings, however with decreasing reliability over

these domains. Approximately 5% of the corpus has been manually tagged for morphology and semantics, whereas the remainder depends upon an automated system and only a fraction of the words are semantically tagged. For further information about morphological tagging in the RNC, we refer the reader to `http://ruscorpora.ru/en/corpora-morph.html`. For a fuller description and critique of the RNC, the reader is referred to Kopotev and Janda (2006).

Although a straightforward examination of the probabilities of the two suffixes aggregated over verbs as observed in the Russian National Corpus suggests a clear pattern, a statistical evaluation of this pattern requires that we take into account the fact that the presence of repeated observations for these verbs renders inappropriate common tests (such as the chi-squared test) that presuppose the independence of the elementary observations. The solution we explore is to use logistic mixed-effects modeling, which allows us to bring under control strong verb-specific trends that are present in our data.

A mixed-effects model is a linear regression model that incorporates both fixed and random effects. Fixed-effect factors are factors with a usually small number of repeatable levels. In our study, the fixed factors are paradigm slot and the place of articulation of the root-final consonant. We model the fixed factors using contrast coding. One factor level is selected as the reference level, and the model's intercept will represent the group mean for this reference level (in our study, for paradigm slot, the active participle). The contrasts for the other factor levels represent the differences in the group means of those other factor levels and the reference level (e.g., between infinitives and present participles).

Random-effect factors are factors (usually with many levels) sampled from a population that is not exhaustively and repeatably sampled. In our study, the individual verbs constitute the levels of a random factor, henceforth referred to simply as 'verb'. We studied 37 verbs, which constitute a sample of a larger population of pertinent verbs. Random-effect factors are modeled as random variables with mean zero and some unknown variance to be estimated from the data. In this way, each individual verb comes to be associated with an adjustment to intercept (a kind of grand average), so that we allow for the possibility that some verbs have a greater preference for *-a* (or *-aj*) than others. Adding the adjustments to the intercept results in 'random intercepts', shorthand for intercepts that have been made precise for each individual verb.

As our dependent variable is binary, with as values *-a* versus *-aj*, we made use of a logistic mixed-effects model. This allows us to model the probability of the two variants with great precision for specific combinations of paradigm slot, place of articulation, and verb, without having to aggregate to obtain proportions, and at the same time avoiding technical problems associated with using a standard linear model for binary data. Technically, we do not model these probabilities directly, but indirectly, by considering the log odds ratio (the log of the ratio of *-a* versus *-aj* responses), and assuming that the variance can be modeled as binomial.

Statistical calculations in this study were carried out using `R`, version 2.9.1, an open source software package for statistical analysis, useful to linguists as both a programming language and a tool for corpus manipulation (Gries, 2009). `R` can be downloaded for free at the Comprehensive `R` Archive Network (CRAN) at `http://cran.r-project.org`. Additional `R` packages that are used in this study include `lme4` (Bates and Maechler, 2009) and `languageR`

Table 2: Verbs included in this study with their glosses and overall frequencies (based on 99 million word sample of the Russian National Corpus representing 1950–2007; cf. Lyashevskaya & Sharoff, forthcoming).

verb	gloss	frequency	verb	gloss	frequency
alkat'	hunger	107	*mykat'*	suffer	26
blistat'	shine	691	*paxat'*	plow	769
bryzgat'	spatter	364	*pleskat'*	splash	176
vcerpat'	scoop	567	*poloskat'*	rinse	218
dremat'	doze	1192	*prjatat'*	hide	2120
dvigat'	move	1244	*pryskat'*	spray	92
glodat'	gnaw	170	*pyxat'*	blaze	143
kapat'	drip	712	*ryskat'*	trot	305
klepat'	rivet; slander	58	*ščekotat'*	tickle	397
klikat'	call	184	*ščepat'*	chip	8
kloxtat'	cluck	7	*ščipat'*	pinch, pluck	310
kolebat'	rock	107	*stonat'*	moan	1110
kolyxat'	sway	102	*svistat'*	whistle	120
krapat'	dribble	5	*tykat'*	poke	878
kudaxtat'	cluck	69	*vnimat'*	perceive	537
kurlykat'	cry like a crane	35	*xlestat'*	whip	528
maxat'	wave	1789	*xnykat'*	whine	199
metat'	throw, sweep; baste	439	*žaždat'*	thirst	1328
murlykat'	purr	233			

(Baayen, 2008), which provides functions specially developed for application to linguistic problems.

2 Analysis

Our data set comprises the 37 verbs listed in Table 2, which also provides a measure of the overall frequency of each verb. The final consonant of the root is a dental for 11 verbs, a labial for 9 verbs, and a velar for 17 verbs. For each verb, counts are available of how often the rival suffixes *-a* and *-aj* are attested in the Russian National Corpus, for each of six slots in the Russian verbal paradigm: the first and second person (singular and plural), the third person singular, the third person plural, the imperative, the gerund and the active participle. The overall frequency of relevant verb forms in the RNC is approximately 10% for 1&2 person singular and plural forms (insufficient data makes it impossible to meaningfully distinguish among these four forms), 22% for 3sg forms, 10% for 3pl forms, 2% for imperative forms, 5% for gerund forms and 4% for active participles (cf. Lyashevskaya & Janda, in preparation; the remaining forms belong to the past tense and infinitive and do not participate in the suffix shift). Gerunds and participles are less characteristic of spoken than written registers (Zemskaja 1983: 116-117)

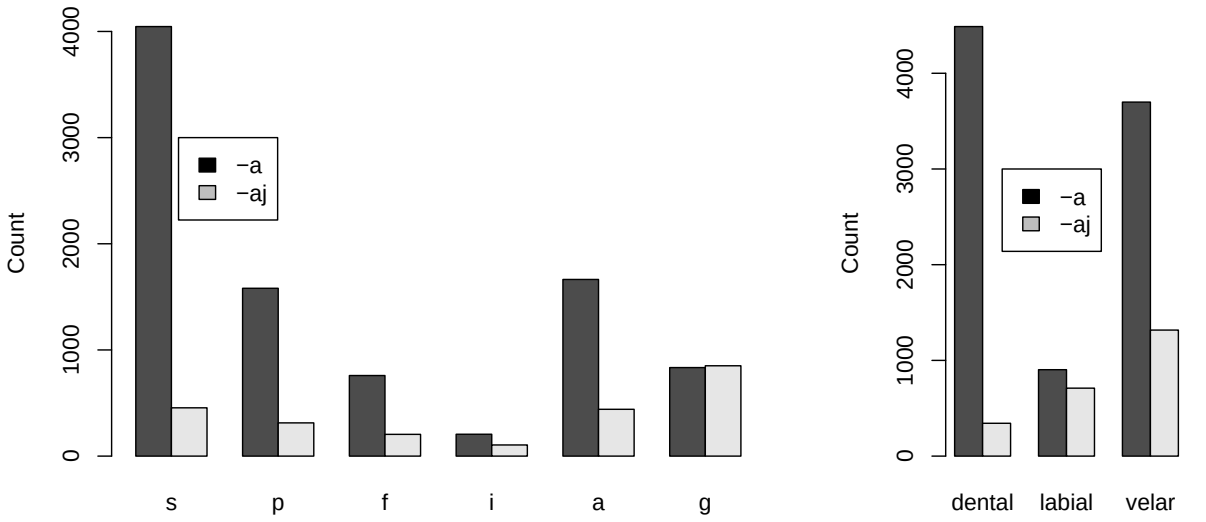

Figure 1: Counts of *-a* (black) and *-aj* (grey) realizations for six paradigm slots (left) and place of articulation of the final consonant of the root (right). a: active present participle, p: third person plural, s: third person sigular, f: first/second person (including both singular and plural), i: imperative, g: gerund.

This study examines only verbs for which both *-a* and *-aj* forms are known to exist. Thus some two dozen *-a* verbs that show no shift to *-aj*, as well as several thousand *-aj* verbs with no *-a* forms are excluded from the study. For an account of why the suffix shift is blocked in some *-a* verbs in Russian, see Nesset (2008).

A barplot of the counts for *-a* (black) and *-aj* (grey) shows that the extent to which *-a* is favored over *-aj* varies considerably across paradigm slots (see Figure 1). The third person singular favors *-a* most, while the gerund shows roughly equal counts for the two suffixes.

Although the skewed distribution of data in Figure 1 invites the use of a straightforward chi-squared test to evaluate whether there are significant differences in the use of the two suffixes across paradigm slots and place of articulation, the chi-squared test is not optimal. First, given the large differences visible in Figure 1, and given the large number of observations involved, 11460, a chi-squared test is certain to argue against the possibility that the likelihood of *-a* and *-aj* would not vary significantly for the different paradigm slots. Second, the chi-squared test is inappropriate as the observations underlying the counts are not independent. It is not the case that the 11460 observations summarized in Figure 1 represent 11460 different verbs sampled randomly from some population. To the contrary, there are only 37 verbs underlying our counts, with the number of observations for a given verb ranging from just 2 to no less than 1343. In this study, therefore, the verbs are the repeated units of analysis, each of which can be expected to have their own individual preferences for *-a* vs. *-aj* suffixed forms. To do justice to the verbs as a source of variation in the choice

between these two forms, we need to bring the verbs and their preferences into our model as a random-effect factor.

In what follows, we investigate these data with the help of logistic regression (see, e.g., Jaeger, 2008; Baayen, 2008; Bresnan et al., 2007). Logistic regression provides us with the means of estimating the likelihood of *-a* (and *-aj*), albeit indirectly, by transforming the counts of *-a* and *-aj* into a log odds ratio, the log of the ratio of 'successes' (*-a*) and 'failures' (*-aj*). We therefore begin with a graphical exploration of this log odds ratio as a function of paradigm slot, coding the log odds by hand (and backing off from zero by adding one to all counts before taking log odds in order to avoid dividing by zero, which yields a mathematically meaningless value). The log odds ratio (also known as 'logit') is thus calculated in this fashion: logit = log((number of *-a* forms + 1)/(number of *-aj* forms + 1)).

Figure 2 presents a trellis dotplot that summarizes the log odds for each of the six paradigm slots. This figure contains 37 plots, one for each verb, with six dots corresponding to the paradigm slots. The dots range across the vertical dimension, which is centered at zero. Thus a dot that is above zero indicates predominance of *-a* forms, whereas a dot that is below zero indicates predominance of *-aj* forms. For example, the plot for the verb *alkat'* 'hunger' is in the lower left corner of the trellis. This verb has predominantly *-a* forms for the 3sg (= 's'), 3pl (= 'p'), 1&2 person (= 'f') and active participle (= 'a'), but predominantly *-aj* forms for the imperative (= 'i') and gerund (= 'g'). For more information about the use of trellis graphics, we refer the reader to Sarkar (2008), see also Baayen (2008: 37-42).

Two things about this graph are noteworthy. First, some verbs show substantial variability in the extent to which they favor *-a* over *-aj* (e.g., *žaždať*, 'thirst') while for others (e.g., *bryzgať*, 'spatter') this variation is much reduced. For some verbs (e.g., *krapať*, 'sprinkle'), it seems as if there is no variation at all, but this is due to the presence of zero counts (for *krapať*, 'sprinkle', nonzero counts are available only for the third person singular). As a consequence, the log odds defaults to log(1) = 0 (recall that we add one to all counts before taking the log odds).

Second, Figure 2 also clarifies that the verbs differ substantially in their overall preference for *-a*. The verb *pryskať*, 'spray', clearly favors *-aj*, whereas a verb such as *dremať*, 'doze', favors *-a*. When we model the probability of *-a* and *-aj*, we will therefore have to take into account that verbs that have different individual overall preferences, as well as individual specific preferences depending upon which paradigm slot is considered.

Within the framework of mixed-effects modeling, we take these two verb-specific preferences into account by means of random intercepts for verbs combined with by-verb random contrasts for paradigm slot. The random intercepts allow us to model the verb's overall preferences as adjustments with respect to the population preferences, by making the intercept precise for each individual verb. The random contrasts provide the opportunity for fine-tuning the contrast coefficients for paradigm slot. Recall that the contrast coefficients for paradigm slot estimate the differences between a given paradigm slot and a reference paradigm slot, in our analysis 'a' (the active participle, selected because `R` picks the reference term alphabetically, unless instructed otherwise; which level is to be selected as reference level is essentially arbitrary, and the choice does not affect our conclusions).The contrasts that we estimate at the level of the fixed-factor 'paradigm slot' represent the average con-

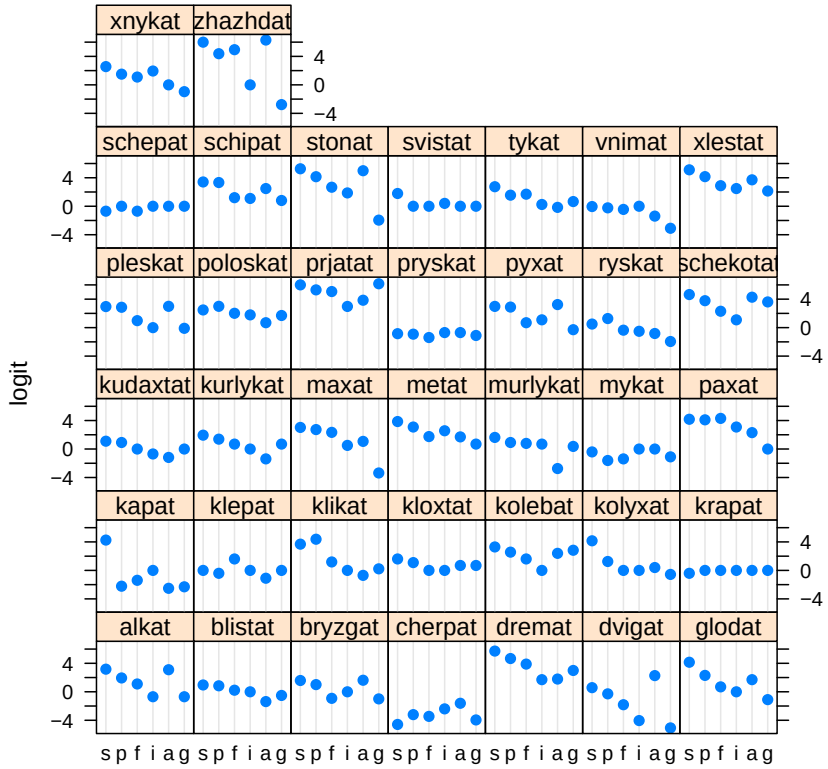

Figure 2: The log odds (of *-a* versus *-aj*) for each of the six paradigm slots ('s': third person singular, 'p': third person plural, 'f': first and second person, 'i': imperative, 'a': active participle, 'g': gerund). Log odds were calculated after backing off from zero by adding 1 to all counts. A log odds greater than zero indicates a preference for *-a*, a log odds smaller than zero a preference for *-aj*.

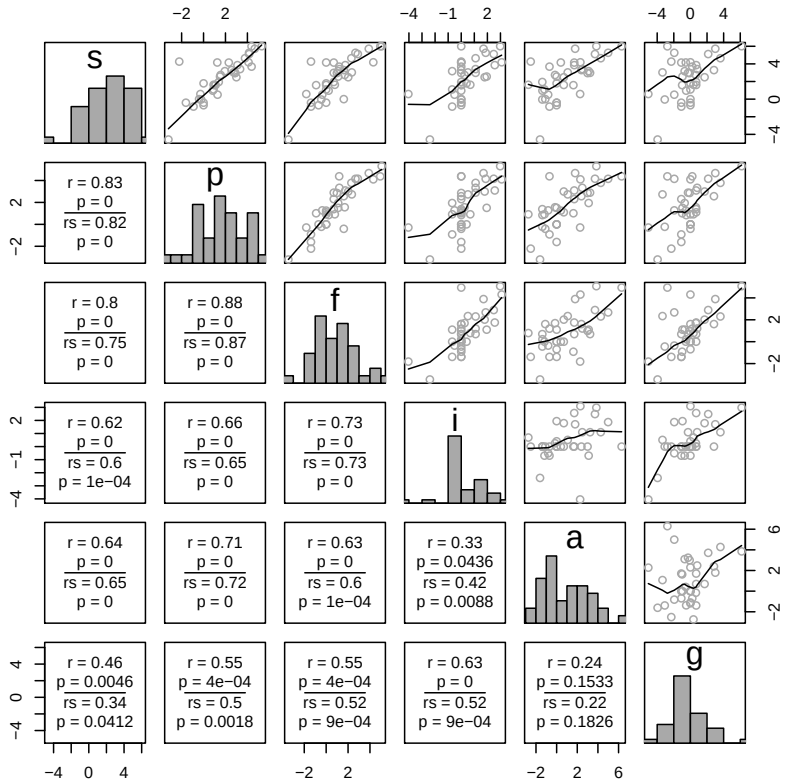

Figure 3: Pairwise correlations for the log odds for the six paradigm slots. Dots represent verbs. The lower half summarizes Pearson (above the line) and Spearman (below the line) correlation coefficients and the associated p-values.

trasts expected for some unseen, new verb, and will not be precise for most of the individual verbs in our study. To make these general contrasts precise for the 37 individual verbs in our sample, we need verb-specific adjustments for each of the contrasts for paradigm slot. Adding these adjustments to the fixed-effect contrasts results in 'random contrasts'.

Before we fit a mixed model to the data, we should consider whether we need a parameter in our model that captures potential correlational structure involving the random intercepts and the random contrasts.

For each verb, we have one intercept and five random contrasts for 'a' = active participle, 'f' = 1&2 person, 'g' = gerund, 'i' = imperative, 'p' = 3pl, and 's' = 3sg. The five contrasts are 'f' versus 'a', 'g' versus 'a', 'i' versus 'a', 'p' versus 'a' and 's' versus 'a'. Since all these adjustments are measured on the same verb, they might be correlated. As a next step, we therefore graphically examine our data for the presence of such correlational structure by means of a pairs plot. Figure 3 plots the pairwise correlations for the log odds across each of the six paradigm slots. Dots represent verbs. With only one exception ('a' and 'g'), the log odds in one paradigm slot enter into strong correlations with the log odds in other paradigm slots. This indicates that we will need a model with a non-trivial random effects structure.

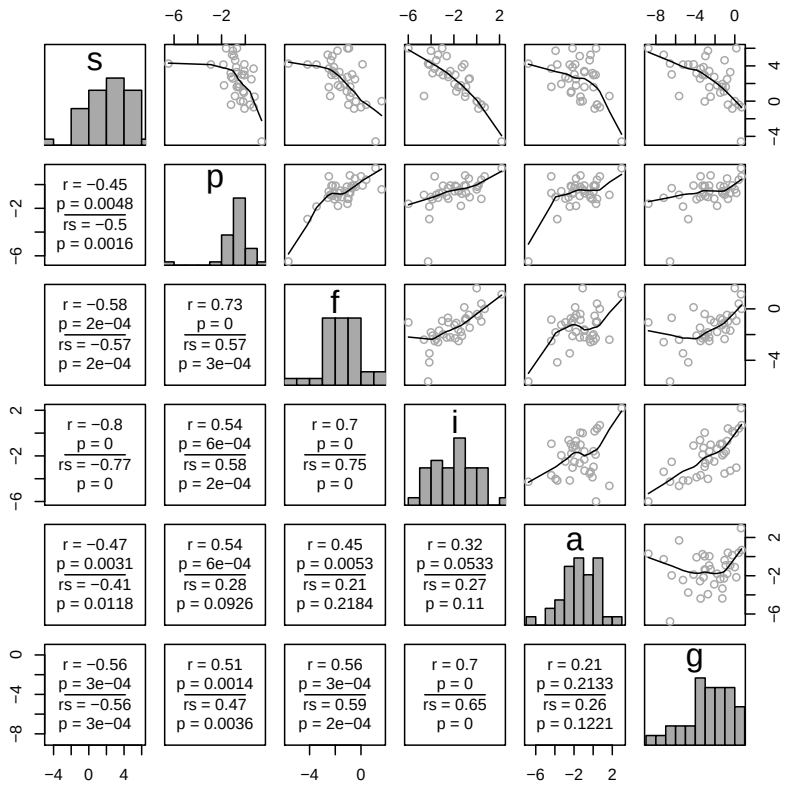

Figure 4: Pairwise correlations for the log odds for the reference level ('a', active participle) and the contrasts with the five remaining levels of paradigm slot. Dots represent verbs.

Figure 4 is similar to Figure 3, but instead of considering the log odds for all six levels of paradigm slot, we consider the reference factor level, 'a', and the contrasts between the other factor levels and this reference level, as we will be using contrast coding for the handling of our factorial predictors. Because we are now dealing with differences with respect to 'a', the technical consequence is that the correlations in the top row of Figure 4 change sign. This is the form in which the tight correlational structure in our data will be captured by our mixed-effects model.

In our analysis, we also include as a covariate the log-transformed frequency of the verb, taken from Table 2. As the change from *-a* to *-aj* appears to take place in the less prototypical parts of the paradigm, we may expect that it also affects lower-frequency verbs more than higher-frequency verbs. Frequency is also important as a control variable, ensuring that paradigmatic effects are not simply frequency effects in disguise.

We now proceed with fitting a mixed model to the data with the log odds modeled as a function of paradigm slot, place of articulation of the root final consonant, and frequency, and with correlated by-verb random intercepts and random contrasts for paradigm. In `R`, using the `lmer()` function in the `lme4` package (Bates and Maechler, 2009), we proceed as follows, assuming that the data are available in a data frame named `dat` that is in long format (with

a given row in the data frame specifying the counts for *-a* (`a`) and *-aj* (`aj`) for each unique combination of `Verb`, `Paradigm` slot, and `Place` of articulation of the root final consonant):[1]

```
dat.lmer = lmer(cbind(a, aj) ~ Paradigm + Place + Frequency +
            (1+Paradigm|Verb), data=dat, family="binomial")
```

The algorithm takes care of backing off from zero, all we need to do is provide it with the raw counts for each verb, supplied to `lmer()` as paired counts (`cbind()` binds vectors column-wise). The random intercepts in our model (the '1' in `(1+Paradigm|Verb)`) take the verb-specific preferences for *-a* as compared to the population average into account. The random contrasts for verb (specified by `Paradigm` in `(1+Paradigm|Verb)` model the verb-specific preferences for *-a* across paradigm slots. Correlation parameters (specified in `(1+Paradigm|Verb)` by specifying both intercept and `Paradigm` before `|Verb`) are essential to do justice to the substantial non-independence that we observed for the verb-specific intercepts and contrasts (as shown in Figure 4).[2]

Table 3: Coefficients of the mixed-effects model and associated Z-statistics.

	Estimate	Standard Error	z-value	p-value
a, dental (intercept)	-0.370	1.921	-0.192	0.847
f - a (contrast)	-0.120	0.475	-0.253	0.800
g - a (contrast)	-2.537	0.712	-3.561	0.000
i - a (contrast)	-1.086	0.692	-1.570	0.117
p - a (contrast)	1.067	0.361	2.953	0.003
s - a (contrast)	1.533	0.424	3.616	0.000
labial - dental (contrast)	-2.972	1.125	-2.642	0.008
velar - dental (contrast)	-2.405	0.926	-2.597	0.009
frequency	0.710	0.296	2.401	0.016

Table 3 lists the estimates of the coefficients, together with their Z-statistics. Of the five contrasts pitting paradigm slots against the reference level of the active participle, three are significant, namely *g-a*, *p-a* and *s-a* (see the column labeled p-value). The two contrasts comparing labial and velar place of articulation with dental place of articulation also reach significance. Finally, the (log-transformed) frequency of the verb is also predictive: the greater the frequency of the verb, the greater the probability of the form with *-a*. In other words, higher-frequency, well-entrenched verbs are more resistant to the language change favoring *-aj* over *-a*.

Figure 5 presents the estimated probabilities of *-a* for each of the levels of paradigm slot and place of articulation, as well as the functional relation between frequency and probability of *-a*.We see that within the set of different paradigm slots, the gerund reveals an exceptional preference for *-aj*. Comparing the right with the left panel, we observe that the differences in the probabilities of *-a* vary more substantially with place of articulation than with paradigm

Table 4: The random effects structure: The column labeled Standard Deviation lists the standard deviations of the by-verb adjustments to the intercept and the contrast coefficients for paradigm slot. The correlation matrix to its right summarizes the pairwise correlations between all six sets of adjustments.

	Standard Deviation	Correlations				
		Intercept	f	g	i	p
Intercept	3.0762					
f	2.1937	-0.634				
g	3.5654	-0.303	0.762			
i	3.2843	-0.702	0.863	0.604		
p	1.6117	-0.386	0.911	0.893	0.743	
s	2.0086	-0.584	0.833	0.516	0.954	0.659

slot. Finally, we note that the effect of frequency, which is linear in the log odds, emerges as non-linear in the probability of *-a*. The likelihood of the innovative form is progressively larger as frequency decreases.

The random effects structure of our model is summarized in Table 4. To capture the correlational structure in the paradigm we need no less than 21 parameters (6 standard deviations and 15 correlations). Figure 6 visualizes this correlational structure by plotting the estimated by-verb adjustments to the population intercepts and contrasts (the so-called best linear unbiased predictors, BLUPs). A comparison of Figure 6 with Figure 4 shows that the model captures succesfully the interdependencies between the counts of *-a* and *-aj* across the different paradigm slots.

The question that we have to address at this point is whether the large number of parameters for the random effects structure is justified. We therefore consider Akaike's information criterion (AIC), a measure of goodness of fit. When comparing models, the smaller the AIC, the better the fit is. For a logistic model without any random effects structure, i.e., a model ignoring the verb altogether, the AIC equals 4789. When we bring into the model random intercepts for verb, the AIC reduces to 1395. Further inclusion of random contrasts and correlation parameters for paradigm slot results in the smallest AIC, 522. A likelihood ratio test provides further confirmation that the complex random effects structure of our model is justified compared to a model with only by-verb random intercepts ($X^2_{(20)} = 912.94, p < 0.0001$).[3] It is noteworthy that the `lmer()` function does not allow a model to be fit to the data in which the correlation parameters of the random effects structure are set to zero. Once by-verb adjustments for paradigm slot are taken into account, the correlation parameters must be taken into account as well.

Figure 7 graphs the log odds ratios for the different paradigm slots as estimated by our model against the corresponding log odds ratios in the data, aggregated over verbs (compare Figure 1 for the corresponding barplot of observed counts). This figure shows us how well the model fits the data. If there were a perfect fit, the estimated log odds ratios and the log odds ratios of the aggregated data would fall exactly on the diagonal. Since the points

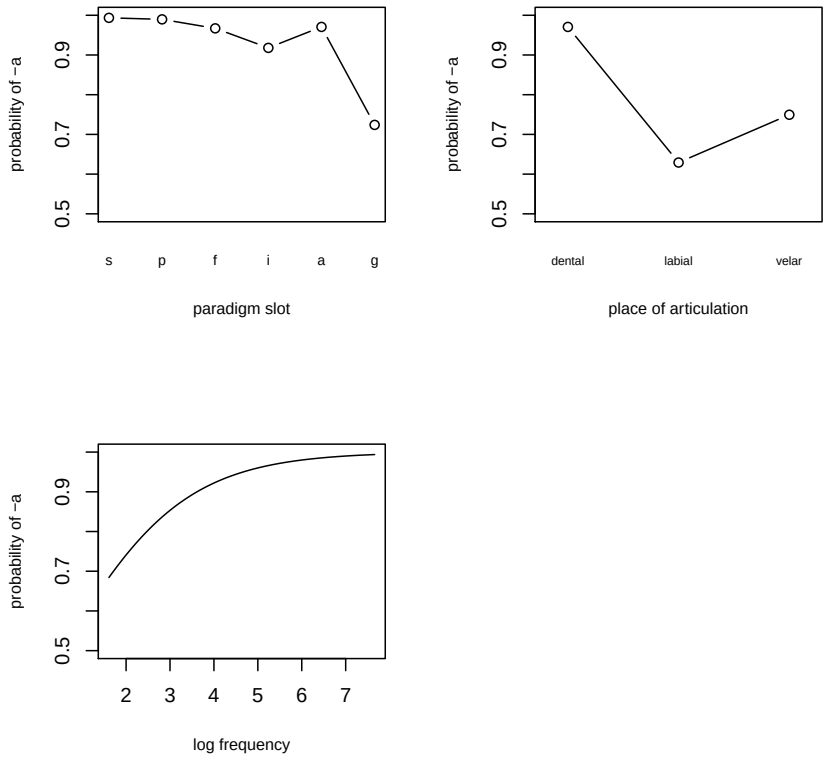

Figure 5: Probabilities of -*a* for paradigm slots (upper left), place of articulation of the final consonant of the root (upper right), and log-transformed verb frequency (lower left) as predicted by a mixed-effects logistic model on the basis of 11,460 observations. The probabilities shown in the upper left panel are adjusted to dental place of articulation. The probabilities in the upper right panel are adjusted for the active participle. The curve for frequency is adjusted for both dentals and the active participle. Key: 'a': active present participle, 'p': third person plural, 's': third person singular, 'f': first/second person (including both singular and plural), 'i': imperative, 'g': gerund.

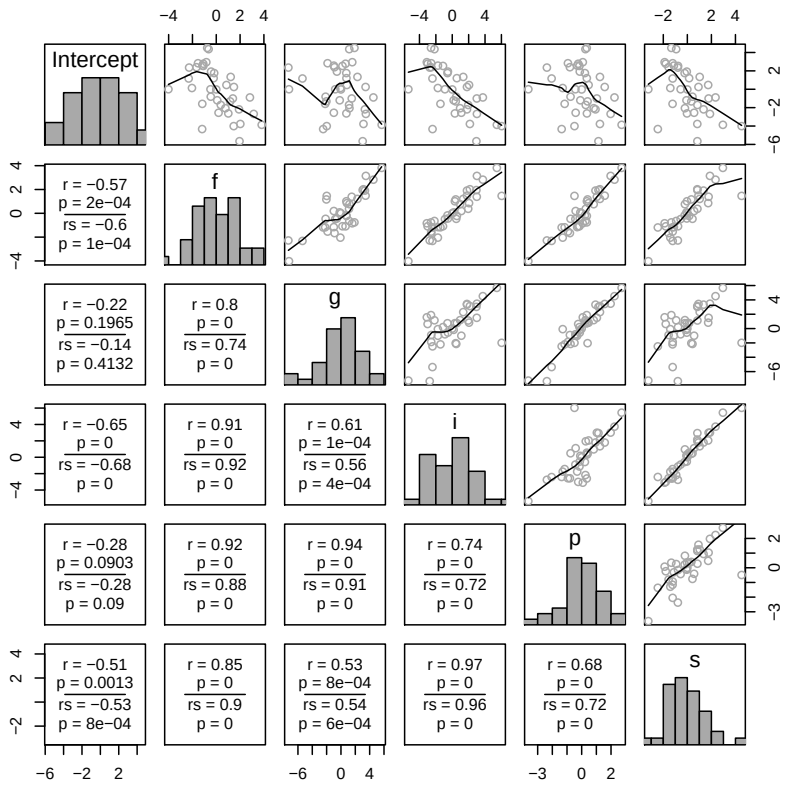

Figure 6: Pairs plot for the by-verb adjustments (BLUPs) to the intercept and contrast coefficients for paradigm slot as estimated by the logistic mixed-effects model.

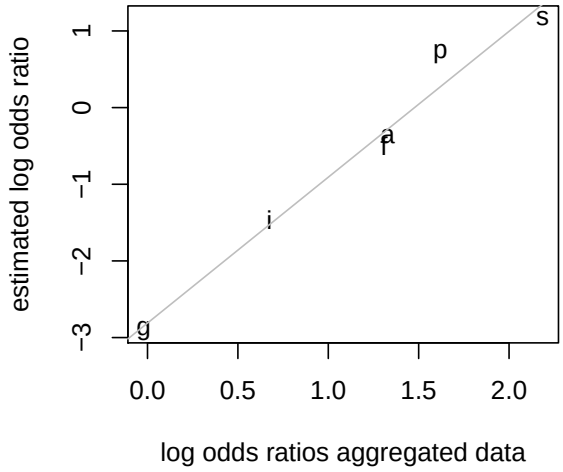

Figure 7: The log odds ratios for the data aggregated by paradigm slot (compare Figure 1) compared to the corresponding log odds ratios as estimated by the mixed model.

are very close to the diagonal, we clearly have a good fit. The small departures present for the individual data points are due in part to imprecision in the fit, and in part to the mixed model taking into account the differences between verbs, both with respect to the number of observations the verbs contribute (ranging from 2 to 1343), as well as with respect to their idiosyncratic preferences for *-a* versus *-aj*. Furthermore, the model predictions on the Y-axis are calibrated for dentals (the reference level of place of articulation, the active participles, and verbs with a log frequency of zero (an absolute frequency of 1). The log odds for the aggregated data, on the other hand, are calculated across dentals, labials and velars, and across frequencies. This also explains why the fitted log odds ratios in Figure 7 are smaller than the observed logits, as can be seen by inspecting the units of the two axes, which range from -3 to 1 for the Y-axis but form 0 to 2 for the X-axis.

While the corrected estimates are very similar to the group averages, as shown in Figure 7, the estimates of the standard errors for the estimates and the corresponding probabilities are very different. This can be seen in Table 5, which lists standard errors and p-values for logistic models with and without random effects structure for the verbs. The estimated standard errors for the model with random effects are sometimes more than 10 times those estimated by the model without random effects structure. Unsurprisingly, the p-values for the latter model are consistently substantially reduced compared to the model that includes the random effects structure. This illustrates that by ignoring the random effects structure in the data, we run the risk of obtaining highly anti-conservative p-values.

The distribution observed in Figure 5 corresponds neatly to the cline that we would predict on *a priori* grounds given what one would expect for the internal structure of the verbal paradigm. We can make the following assumptions, all justifiable on independent grounds (cf. Janda 1995 for relationships between markedness and prototypicality for such

Table 5: Standard errors and p-values as estimated in logistic models with (first two columns) and without (second two columns) random effects structure for the verbs.

	mixed model		model without random effects	
	Standard Error	p-value	Standard Error	p-value
a, dental (intercept)	1.9211	0.8474	0.2183	0.0000
f - a (contrast)	0.4747	0.7999	0.1043	0.0012
g - a (contrast)	0.7123	0.0004	0.0803	0.0000
i - a (contrast)	0.6920	0.1165	0.1417	0.0009
p - a (contrast)	0.3612	0.0031	0.0892	0.0000
s - a (contrast)	0.4241	0.0003	0.0796	0.0000
labial - dental (contrast)	1.1249	0.0082	0.0853	0.0000
velar - dental (contrast)	0.9260	0.0094	0.0703	0.0000
frequency	0.2958	0.0164	0.0303	0.0000

categories, and Bybee 1985: 50-52 for further discussion of 3sg as a prototypical verbal form):

- Finite forms are more prototypical than non-finite forms;
- indicative is more prototypical than imperative;
- singular is more prototypical than plural; and
- third person is more prototypical than first and second person.

This yields the following expected hierarchy, ranging from most prototypical (finite indicative, third person singular) to least prototypical (non-finite forms):

3sg	>	3pl	>	1&2person	>	imperative	>	gerund/active participle
's'		'p'		'f'		'i'		'g', 'a'
1		2		3		4		5, 6

The order in the hierarchy ('s':1, ..., 'g','a': 5,6) receives confirmation from the ranking of the mean log odds for each paradigm slot, estimated from the coefficients in Table 3 ('s': 1.16, 'p': 0.70, 'f': -1.46, 'i': -2.91, 'a': -0.49, 'g': -0.37). Spearman's ρ (the rank correlation coeffcient) is estimated at -0.8286, $p = 0.0292$, one-tailed test. We note that the expected hierarchy recieves this support in an analysis that does not just consider the simple group probabilities, but also takes individual variation attached to verbs and paradigm slots into account, and is therefore appropriately conservative.

This study provides further evidence that paradigms have structure and that this structure is reflected in the way in which language change unfolds. Furthermore, the structure of a paradigm is co-determined by considerations of prototypicality, even when the degree of entrenchment of the verb is taken into account.Thus the peripheral forms of a paradigm, such as the gerund, are more affected by language change than the prototypical forms, such as the 3sg, which is insulated from change. The linguistic arguments for this conclusion and

its implications for morphology are presented in a more comprehensive fashion in Nesset & Janda (in prep.).

The only paradigm slot in the hierarchy that does not follow the anticipated distribution is the active participle, which reveals a probability that is most similar to the 1&2 person finite indicative forms, instead of being most similar to the probability of the gerund (from which it differs substantially, see Table 3). Interestingly, a closer look at this participle reveals that it has close ties to indicative forms that have probably influenced its behavior. Although in terms of the abstract semantics of paradigm categories the participle is, of course, a non-finite form, it is formally closely related to the 3pl form: The suffix of the active participle always contains the same vowel as the 3pl form. Our data indicate that, apparently, this formal relationship interacts with the semantic hierarchy. Whereas the participle is pulled in one direction because of the semantic hierarchy, which would place it near the bottom with the gerund, at the same time it is pulled in the other direction because of its close formal relationship with the 3pl. The participle winds up between the two, closest to the 1&2 person forms. For a more detailed discussion of this cline the reader is referred to Nesset & Janda (in prep.).

3 Concluding remarks

A pervasive characteristic of language is that it is a system in which there are large numbers of interdependencies. When analyzing quantitative linguistic data, it is essential that these interdependencies are brought into the statistical model. Failure to do so may give rise to anti-conservative analyses and may cause the researcher to draw incorrect conclusions from the data. We have shown how mixed-effects modeling can serve as a tool that may help us to better model the complex interdependencies in language data. Our present example addressed dependencies in paradigm structure, but similar dependencies may arise whenever multiple data points are collected, for given linguistic units (verbs, constructions, multiword units, etc.) as well as for speakers and writers (in corpus linguistics) and informants (in sociolinguistics).

A direction for further research on the emergence of *-aj* in Russian is to also bring into the model the identity of the speakers/writers, as different language users may have have their own preferences or dispreferences for *-aj*. In addition, one might consider adding the text in which the word appears into the model as a third random effect factor, and perhaps even covariates for how often a verb has appeared before in that text. Additional issues that could be considered include stress patterns, morphological changes, and transitivity. The more such predictors are taken into consideration, the better the model will be able to take into account the many interdependencies that characterize natural language. But as we bring more such considerations into our analysis, the problem of data sparseness increases exponentially. The Russian National Corpus comprises over 52,000 sources, with on average some 3000 words. As even our most frequent verb (*maxat'*, with 1789 occurrences), has only an expected frequency of 0.04 in a text of average length, even the 140 million words of the Russian National Corpus provide too small a sample to properly address the full complexity of the choice between *-aj* and *-a*.

Notes

[1] The model was fitted with relativized maximum likelihood, the default of the `lmer` function, which is optimal for estimating and evaluating random effects.

[2] We also considered whether by-verb random contrasts for Place might be required. However, a likelihood ratio test comparing models with and without such random contrasts showed that the additional model parameters required for these contrasts do not contribute to a significant increase in goodness of fit. We therefore did not include random contrasts for Place.

[3] For the evaluation of goodness of fit of logistic models, see, e.g., Harrell (2001), and for some discussion of evaluating the goodness of fit of mixed models, and why measures such as R-squared are not recommended, see Baayen (2008).

References

Andersen, H. (1980). Russian conjugation: Acquisition and evolutive change. In Traugott, E. C., editor, *Papers from the 4th international conference on historical linguistics*, pages 285–301. John Benjamins, Amsterdam.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge University Press, Cambridge, U.K.

Bates, D. and Maechler, M. (2009). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-31.

Bresnan, J., Cueni, A., Nikitina, T., and Baayen, R. H. (2007). Predicting the dative alternation. In Bouma, G., Kraemer, I., and Zwarts, J., editors, *Cognitive Foundations of Interpretation*, pages 69–94. Royal Netherlands Academy of Arts and Sciences.

Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. Benjamins, Amsterdam.

Gagarina, N. (2003). The early verb development and demarcation of stages in three Russian-speaking children. In Bittner, D., Dressler, W. U., and Kilani-Schoch, M., editors, *Development of Verb Inflection in First Language Acquisition: A Cross-Linguistic Perspective*, pages 131–170. Mouton de Gruyter, Berlin – New York.

Gor, K. and Chernigovskaya, T. (2001). Rules in the processing of Russian verbal morphology. In Zybatow, G., Junghanns, U., Melhorn, G., and Szucsich, L., editors, *Current Issues in Formal Slavic Linguistics*, pages 528–536. Peter Lang, Frankfurt am Main.

Gor, K. and Chernigovskaya, T. (2003a). Formal instruction and the acquisition of verbal morphology. In Housen, A. and Pierrard, M., editors, *Current Issues in Instructed Second Language Learning*, pages 103–136. Mouton de Gruyter, Berlin and New York.

Gor, K. and Chernigovskaya, T. (2003b). Generation of complex verbal morphology in first and second language acquisition: Evidence from Russian. *Nordlyd*, 31(6):819–833.

Gor, K. and Chernigovskaya, T. (2003c). Mental lexicon structure in L1 and L2 acquisition: Russian evidence. *GLOSSOS*, 4:1–31.

Gries, S. T. (2009). *Quantitative corpus linguistics with R: a practical introduction*. Routledge, Taylor & Francis Group, London.

Harrell, F. (2001). *Regression modeling strategies*. Springer, Berlin.

Jaeger, F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, X:in press.

Janda, L. A. (1995). Unpacking markedness. In Casad, E., editor, *Linguistics in the Redwoods: The expansion of a new paradigm in Linguistics*, pages 207–233. Mouton de Gruyter, Berlin.

Kiebzak-Mandera, D., Smoczynska, M., and Protassova, E. (1997). Acquisition of Russian verb morphology: the early stages. In Dressler, W., editor, *Studies in Pre- and Protomorphology*, pages 101–114. Verlag der Österreichischen Akademie der Wissenschaften, Wien.

Kopotev, M. and Janda, L. (2006). Review of nacional'nyj korpus russkogo jazyka [russian national corpus] (`www.ruscorpora.ru`), by Plungjan, V. A., Raxilina, E. V. et al. *Voprosy jazykoznanija*, 5:149–155.

Krysin, L. P. (1974). *Russkij jazyk po dannym massovogo obsledovanija*. Nauka, Moscow.

Lyashevskaya, O. and Janda, L. A. (2009). What can the grammatical profile of a verb tell us about its semantics? *in preparation*, 1:1–2.

Lyashevskaya, O. and Sharoff, S. (2009). *Častotnyj slovar' sovremennogo russkogo jazyka*. Moscow.

Nesset, T. (2008). Ob"jasnenie togo, čto ne imelo mesto: Blokirovka suffiksal'nogo sdviga v russkix glagolax. *Voprosy jazykoznanija*, 6:35–48.

Nesset, T. and Janda, L. A. (in preparation). Paradigm structure: evidence from Russian suffix shift.

Sarkar, D. (2008). *Lattice. Multivariate Data Visualization with R*. Springer, New York.

Tkachenko, E. and Chernigovskaya, T. (2006). Focus on form in the acquisition of inflectional morphology by L2 learners: Evidence from Norwegian and Russian. paper presented at The Second Biennial Conference on Cognitive Science, St. Petersburg, June 9–13, 2006.

Švedova, N. J. (1980). *Russkaja Grammatika (vol. 1)*. Nauka, Moscow.

Zaliznjak, A. A. (1977). *Grammatičeskij slovar' russkogo jazyka*. Izdatel'stvo Russkij Jazyk, Moscow.

Zemskaja, E. A. (1983). *Russkaja razgovornaja reč'*. Nauka, Moscow.