PREDICTING RUSSIAN ASPECT BY FREQUENCY ACROSS GENRES

Hanne M. Eckhoff, University of Oxford, UK
Laura A. Janda, UiT The Arctic University of Norway
Olga Lyashevskaya, National Research University Higher School of Economics, Moscow

1. Introduction: How do we know which verbs are perfective and which ones are imperfective?

The verbal category of aspect is perhaps the most characteristic feature of Russian grammar, classifying the entire verbal lexicon into perfective and imperfective verbs, with a residue of biaspectual verbs. In Russian, aspect has become relatively independent of lexical semantics, such that the majority of verbs are "paired" for aspect (having the same lexical semantics and differing only in aspect), the numbers of perfective tantum and imperfective tantum verbs are very small (Janda "Aspectual clusters"), and perfectives can readily be formed even from verbs with atelic semantics (Dickey "Prototype account," "Varying role").

But how do language users know which verbs belong to which aspect? Is the aspectual behavior of verbs uniform and stable or does the broader context or genre play any role? And how might first language learners of Russian learn to sort out their verbs?

These are non-trivial questions that can be approached from various angles. Three possible approaches are morphological, syntagmatic, and paradigmatic. This article takes a paradigmatic approach. However, to put our study in perspective, we briefly review all three approaches first.

Morphological Approach

Russian is famous for its aspectual morphology, which provides important and systematic cues to the identification of verbs as perfective vs. imperfective. Indeed, the majority of verbs are overtly marked for aspect by means of prefixes and suffixes. Unfortunately, however, this system is not fully reliable,

Research by Olga Lyashevskaya has been funded by the Russian Academic Excellence Project «5-100».

largely because the morphology of aspect has been cobbled together over many centuries from various sources: prepositions have become perfectivizing prefixes, the imperfect desinence and imperfectivizing suffixes have evolved in close interaction (Wiemer and Seržants forthcoming; Kamphuis ch. 10), and the *–nǫ* suffix has developed into a semelfactive marker (Nesset), just to name a few. Inevitably, this long and complex process has left a considerable residue of exceptions. A simplex verb is usually imperfective (like *pisat'* 'write'), but some simplexes are biaspectual, like *ženit'sja* 'marry'. Some simplex verbs are perfective, and like *dat'* 'give', these verbs tend to be high-frequency (for the three samples described below, the frequency of perfective simplex verbs was 10% of all tokens for Fiction, 8% for Journalism, and 3% for Scientific-Technical prose). Prefixed verbs are largely perfective as long as they do not contain a suffix, as in *pere-pisat'* 'rewrite', but there are some cases where prefixation does not yield a perfective verb. For example, one can prefix the imperfective simplex *prjač'* 'harness' to get the perfective *soprjač'* 'conjoin'. However, if the same process is applied to *suščestvovat'* 'exist', the result is *sosuščestvovat'* 'coexist', which is an imperfective verb. Prefixed indeterminate verbs of motion are commonly imperfective, like *pere-xodit'* 'walk across'. But some prefixed indeterminate verbs of motion can be both perfective and imperfective, like *s-xodit'* 'go someplace once' as a perfective vs. 'go down' as an imperfective. Many verbs suffixed in *–nu* are perfective, like *krik-nu-t'* 'yell once'. But there are also imperfective verbs that end in *–nu-t'*, like *sox-nu-t'* 'dry'. Perhaps the most dependable prediction is that a verb with both a prefix and an imperfectivizing suffix is imperfective, as in *pere-pis-yva-t'* 'rewrite'. And of course there are more complex combinations of markers involving prefix stacking (Tatevosov). The inadequacy of morphological markers as cues to the aspectual identity of verbs has been flagged as a problem from the perspective of first language acquisition (Stoll). It is clear that while learners can use morphological cues in most cases, they must also supplement morphology with other information in order to fully sort out the aspect of Russian verbs.

Syntagmatic Approach

Could context be the answer? We know that there are certain "triggers" external to the verb that are usually associated with only one aspect, so perhaps syntagmatic contexts are enough to sort out the aspects of verbs. Language teachers drill into students of Russian a whole series of "rules" about when to use perfective vs. imperfective. For example: perfective verbs appear after adverbials like *nakonec* 'finally', *vnezapno* 'suddenly', *srazu* 'immediately', *čut' ne* 'nearly', *vdrug* 'suddenly', *uže* 'already', *neožidanno* 'unexpectedly', *sovsem* 'completely', *za tri časa* 'in three hours' and as complements of verbs like *zabyt'* 'forget', *ostat'sja* 'remain', *rešit'* 'decide', *udat'sja* 'succeed', *uspet'* 'succeed', *spešit'* 'hurry'; imperfective verbs appear with adverbials like *vsegda* 'always', *často* 'often', *inogda* 'sometimes', *poka* 'while', *postojanno* 'continually', *obyčno* 'usually', *dolgo* 'for a long time', *každyj den'* 'every day', *vse vremja* 'all the time', *tri časa* 'for three hours', under categorical negation as with *ne nado* 'should not', *ne stoit* 'not worth', *ne razrešaetsja* 'not allowed', as the complements of phasal verbs like *stat'* 'start', *načat'/načinat'* 'begin', *prodolžit'/prodolžat'* 'continue', *končit'/ končat'* 'stop', verbs of motion like *pojti* 'go', and a few others like *učit'sja* 'learn', *umet'* 'know how', *ljubit'* 'love'.[1] Reynolds checked the validity of these syntagmatic "rules" in corpus data and found that while they are quite reliable when they apply (giving 96% correct predictions of aspect), the contexts in which they apply occur with only 2% of verb tokens. In other words, the rules are not very useful because the syntagmatic "triggers" (adverbials and verbs that predict aspect) are not frequent enough.[2] In practice, of course, syntagmatic context may be augmented by extralinguistic cues and shared knowledge between speaker and hearer.[3] It may also be the case that there are more contextual cues that have not yet come to the attention of researchers but are utilized by speakers.

Furthermore, of course, many contexts are neutral with respect to aspect. In other words, there are contexts in which both a perfective and an imperfective verb can appear, and the choice is dependent upon what the speaker intends to convey. For example, one can ask both *Vy pročitali Vojnu i mir?*

[1] Only a sample of aspectual triggers are listed here. A longer list was aggregated from numerous reference grammars and textbooks, among them: Andrews et al., Bogojavlensky, Borras and Christian, Brown, Kagan et al., Lekic et al., Lubensky et al., Murphy, Offord, Pulkina et al., Rassudova, Robin et al., Rifkin, Timberlake, Wade.

[2] This may seem counter-intuitive, since some adverbs like *uže* 'already' and *vsegda* 'always' are of relatively high frequency. But a much higher frequency would be needed to provide every verb form, or even a majority of them, with a trigger word. Verbs are just much more frequent overall. In the manually disambiguated Morphological Standard of the Russian National Corpus containing 5,944,156 tokens, 1,007,526 of those (16.9%) are verb forms, but only 375,740 of them (6.3%) are adverbs (see Table 1 in Endresen et al.). In other words, there are nearly three times as many verb forms as adverbs. Adverbs have a variety of functions in Russian, so not all of these adverbs are modifying verbs. That leaves the majority of verbs without any adverb. And many adverbs are not specific to one aspect or the other.

[3] Tense could potentially help to disambiguate, but it is verb-internal (thus not strictly speaking a syntagmatic cue from the context), and also of limited use. If interlocutors share the knowledge that an event is present or future, this might make it possible to predict imperfective vs. perfective for nonpast forms. However, in many contexts both imperfective and perfective nonpast forms can be used to refer to future events, as in *Zavtra ja idu/pojdu na lekciju* 'I'm going to the lecture tomorrow'. Another context that might provide clues is the use of negation with imperatives (combining a syntagmatic marker of negation with verb-internal marking of imperative), as in *Ne snimajte očki!* 'Don't take off your glasses!'. However, imperative forms constitute only about 3% of verb forms in the Russian National Corpus, and even fewer of those are negated. And negation of an imperative is also possible for perfective verbs; in fact it is required for warnings about specific threats, as in *Ne ušibites' golovoj!* 'Don't hit your head!' (said when entering a place with a low clearance).

'Have you read War and Peace?' (using a perfective if the speaker has some reason to believe that the hearer was supposed to have read the book) and *Vy čitali Vojnu i mir?* 'Have you read War and Peace?' (using an imperfective if the speaker just wants to know whether the hearer is familiar with the book). In these cases syntagmatic context provides no cues to the aspect of the verb. Thus, in sum, syntagmatic context is of only limited help, and in contexts where aspect is variable, it is of no help whatsoever.

Paradigmatic Approach

Tomasello (ch. 4–5) noticed that an L1 learner of English can show strong preferences for the inflectional forms used with a given verb and that this is related to the type of verb. For example, achievement and accomplishment verbs often appear in past tense and past participle forms like *made*, *gone*, and *broken*, whereas verbs that name activities often appear as progressives like *sweeping* and *cooking*. Shirai and Anderson proposed a "Distributional Bias Hypothesis" for the acquisition of tense-aspect morphology in English, according to which children pick up on differences in the distribution of verb forms. Past tense is used predominantly with achievement verbs like *fell down* and progressive is used predominantly with activity verbs like *dancing*. In other words, aspectually-marked verb forms are distributed differently in the speech of adults, and children mimic this distribution. If so, the acquisition of aspect is aided by children's overall sensitivity to statistical tendencies in their languages (Goldberg). Aksu-Koç confirmed this hypothesis in a study of the acquisition of aspect in Turkish, finding that children not only reflect the distributions of verb forms of adults, but are also more conservative. This finding was corroborated for Russian by Stoll and Gries, who looked at the association between aspect and tense in child language vs. child-directed speech. They found that the correlation is stronger for children than for adults: whereas in child-directed speech perfective verbs have a tendency to appear in the past tense and imperfectives in the present tense, children distilled this tendency into something closer to a complementary distribution.

It is probably the case that L1 learners use all types of cues in building up their competence in Russian aspect. Languages are typically highly redundant systems in which the factors are multiply collinear (Dąbrowska), a fact that likely both supports language acquisition and facilitates communication. Morphological, syntagmatic, and paradigmatic cues probably supplement each other, each making up for what the others lack in terms of reliability. This makes it possible to build up a robust system of perfective vs. imperfective verbs that is more reliable than any one set of cues could provide.

Our study is not directly aimed at language acquisition, but we do ask what information is potentially available to L1 learners. We examine one of the sets of cues that has not received much attention thus far, namely paradigmatic cues. We ask: how reliably could one sort Russian verbs according to aspect based only on the distribution of their forms? How does this compare with the sorting of verbs facilitated by aspectual morphology? How might paradigmatic cues supplement morphology and other cues?

Our study takes the paradigmatic line of research in several directions. Rather than looking only at tense, we take into account the entire range of verb forms in Russian. Language acquisition data is by nature relatively sparse, making it impossible to gather enough data to make generalizations about individual verbs. Given the current state of the art, there are no samples of child-directed speech large enough to support investigation of paradigmatic cues. We use corpus data as a proxy for child-directed speech, and show that it is possible to find patterns at the level of the individual verb. While corpus data does not accurately reflect the input of acquisition, it can be stratified for genre. If our findings turn out to be robust across genres, they might be similar also in the "genre" of child-directed speech. Thus it is possible that our study also has implications for acquisition, although our main focus is on predicting the aspect of individual verbs and comparisons across genres.

Our study takes the perspective of a usage-based model of language (cf. Langacker, Janda "Cognitive Linguistics 2015"), according to which linguistic generalizations are expected to emerge from the language data that members of the speech community are exposed to, and linguistic categories have the same properties as other cognitive categories, namely that they have a radial category structure with prototypical vs. more peripheral members. For the purposes of our study, this means that learners and speakers are sensitive to the statistical tendencies that characterize the distribution of perfective vs. imperfective verbs. They can use these tendencies, along with other cues, to make generalizations about the categories of perfective vs. imperfective verbs. A prototypical perfective verb will have all of the characteristics associated with perfective verbs: a) a distribution of forms that resembles that of perfective verbs overall, b) unambiguous morphology marking it as perfective, and c) an association with contexts typical of perfective verbs. Some perfective verbs will deviate from this ideal to a greater or lesser degree. A prototypical imperfective verb will likewise show a strong alignment of morphological, syntagmatic, and paradigmatic cues, while individual verbs may deviate. Our purpose is to examine the specific contribution of paradigmatic cues to the emergence of the category of Russian aspect.

In Section 2 we present the paradigmatic methodology of grammatical profiling as applied to Russian and Old Church Slavonic, and also specify our research questions. Section 3 describes our data extracted from the Russian National Corpus (ruscorpora.ru) and stratified across the registers of Journalistic prose, Fiction, and Scientific-Technical prose, which we will call "genres" for the sake of brevity.[4] The analysis of each of these samples is presented in turn

[4] In the Russian National Corpus and Lyashevskaya and Sharoff's frequency dictionary of Russian these registers are called "functional domains."

in Sections 4-6, with comparisons across the genres in Section 7. In Section 8 we present calculations of the accuracy of morphological cues and compare those with the results for paradigmatic cues. We offer conclusions in Section 9.

2. Grammatical Profiling in Studies of Aspect in Russian and OCS

A grammatical profile is the frequency distribution of paradigm forms for a given word in a corpus. For example, in the sample of Journalistic prose described in Section 4 below, we find sixty-two tokens of the verb *čitat′* 'read' distributed as shown in Table 1: there are two nonpast[5] gerund forms (*čitaja*), eight imperatives (*čitaj*, *čitajte*), one indicative future (*budet čitat′*), fifteen indicative nonpast forms (*čitaju*, *čitaeš′* etc.), fourteen indicative past forms (*čital*, *čitala* etc.), twenty-two infinitive forms, and no attestations of participles or a past gerund. This array of numbers is the grammatical profile of the verb *čitat′* 'read' in our sample of Journalistic prose. We can of course extract the grammatical profiles of the other verbs in our sample. In order to put all the profiles on the same scale we can represent them in terms of percentages, as in the bottom row of Table 1. This facilitates comparison across the grammatical profiles of verbs. In a large corpus we would expect the grammatical profile of each verb to be unique, but we would also expect to find some patterns.

Table 1: The grammatical profile of *čitat′* 'read' in our sample of Journalistic prose

nonpast gerund	past gerund	imperative	indicative future	indicative nonpast	indicative past	infinitive	nonpast participle	past participle
2	0	8	1	15	14	22	0	0
3.2%	0%	12.9%	1.6%	24.2%	22.6%	35.5%	0%	0%

Janda and Lyashevskaya pioneered the use of grammatical profiles to analyze Russian verbs, with an aggregated study of nearly six million forms of verbs extracted from the Russian National Corpus. In that study, it was established that the overall grammatical profile of perfective verbs was indeed different from that of imperfective verbs, as shown in Figure 1. Of the four subparadigms examined, three are more characteristic of perfective verbs, namely past tense (to a very strong degree), infinitive, and imperative, whereas the nonpast tense is by far more characteristic of imperfectives. This result is statistically significant (chi-squared = 947756, df = 3, p-value $< 2.2e\text{-}16$) with a medium-large effect size (Cramer's V = 0.399).

[5] We use the term "nonpast" because the aim of our study is to show to what extent observable features of the input can be used to deduce the aspect of verbs. From the perspective of observable features, *kuplju* 'I will buy' (a nonpast form of a perfective verb, which usually expresses future tense) and *smotrju* 'I look' (a nonpast form of an imperfective verb, which usually expresses present tense) are inflectionally identical, representing the same paradigmatic form. We group these together in order to reflect only the observable facts in the data.

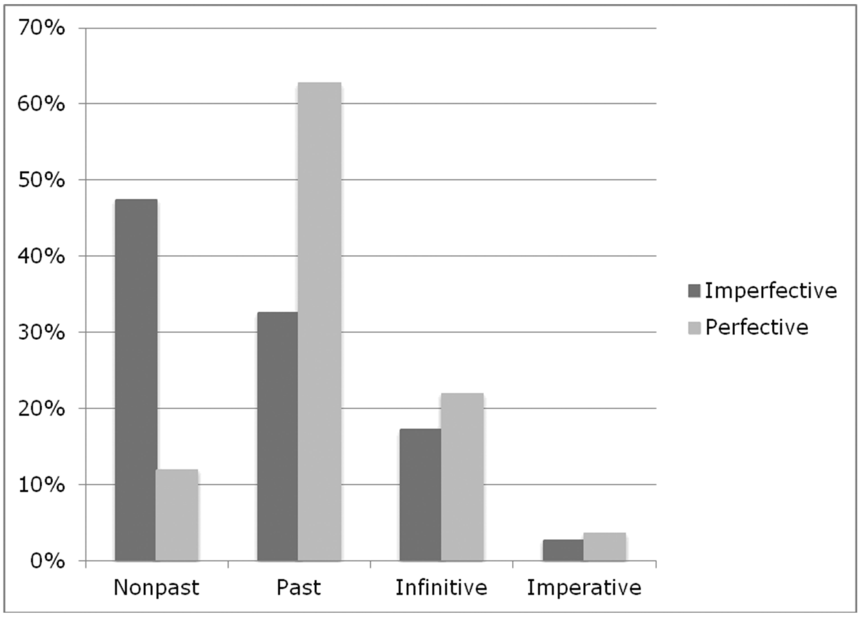

Figure 1: Aggregated Grammatical Profiles of Perfective and Imperfective verbs (reproduced from Janda and Lyashevskaya)

The Janda and Lyashevskaya study had, however, a number of limitations, the most important of which are that a) it labeled all verbs as perfective vs. imperfective from the outset, b) it compared the grammatical profiles in aggregate rather than individually, c) it was limited only to aspectually paired verbs formed either by prefixation (as in *na-pisat'* / *pisat'* ‘write’) or by suffixation (as in *pere-pisat'* / *pere-pis-yv-at'* ‘rewrite’),[6] d) and it specifically excluded grammatical forms that are restricted by aspect in modern Russian, namely gerunds and participles. This study did not show whether it is possible to use grammatical profiles to sort individual verbs of all types according to their aspect (nor was that the intention). We could characterize the Janda and Lyashevskaya study as a “top-down” approach, since the researchers first sorted the verbs into perfective vs. imperfective aspect and then looked at the grammatical profiles of each aspect in aggregate.

Eckhoff and Janda applied a “bottom-up” approach to the grammatical pro-

[6] Note that points b) and c) are true only of the first study presented by Janda and Lyashevskaya, where the grammatical profiles of perfective and imperfective verbs are compared. In their second study in the same article, all verbs are included and they are examined at an individual level, but that is in relation to specific combinations of tense, aspect, and mood features, not in relation to full grammatical profiles. Note also that Janda and Lyashevskaya did not present the imperfective periphrastic future (e.g., *budu* + infinitive) as a separate category.

filing of verbs in Old Church Slavonic in order to address the age-old controversy over whether there were perfective vs. imperfective verbs in the earliest attestations of Slavic. In that study, correspondence analysis was used to sort the verbs according to their grammatical profiles. In the input data for the correspondence analysis, each verb was represented as a row of figures citing the number of tokens of various forms attested for that verb. For example, *tvoriti* 'make' was found to have 0 aorist forms, 14 imperatives, 12 imperfect forms, 23 infinitive or supine forms, 99 present tense forms, 0 past participles, and 26 present participles (note the parallel to the grammatical profile of *čitat′* 'read' in Table 1 above). The input data had 129 rows representing the verbs. The correspondence analysis used this data to create two distance matrices, one for the rows (verbs) and one for the columns (subparadigms such as aorist, imperative, etc.). In effect, correspondence analysis measured the distances between the verbs by creating a multidimensional space defined by what the model calls "Factors"[7] that are mathematically constructed based on the data. The Factors are ordered according to their capacity to account for the variance in the data, with the strongest Factor numbered 1, followed by 2, etc. The results of correspondence analysis can be represented in a graph with rows (verbs) and columns (subparadigms) plotted according to their Factor 1 (x-axis) and Factor 2 (y-axis) values (see Figures 2–4 and more details in Sections 4–6 below; see also Eckhoff and Janda for more description of correspondence analysis).

In layman's terms, we could say that correspondence analysis takes a matrix of numbers and computes the distances between items in the matrix, dividing them into groups. In our matrix, rows represent verbs and columns represent the numbers of paradigm forms for each verb (as in Table 1). Since our items are verbs and their forms, correspondence analysis computes the distances between them, each time creating a mathematical "Factor" that splits the items into two groups. The job of the computer program is to sort out the items the way that makes most sense computationally. The farther an item is placed from the origin (the point where Factor 1 = 0 and Factor 2 = 0), the more its behavior deviates from the mean in rows and columns.

Eckhoff and Janda discovered that the strongest Factor (numbered 1) has a distribution that resembles aspect: it divides the Old Church Slavic verbs into two groups, one with negative values that correspond to imperfective verbs, and one with positive values that correspond to perfective verbs, and this result concurs 96% with Dostál's aspectual classification of verbs. Eckhoff and Janda's study of Old Church Slavic verbs is further corroborated by Kam-

[7] "Factor" is the standard term for correspondence analysis so we retain it here, although it may seem misleading. It is important to remember that these Factors result from the computation of the model and as such, are not set in advance and have no prior interpretation either. It is up to the researcher to interpret the meaning of each Factor that results from correspondence analysis.

phuis (ch. 7), who examined groups of verbs aggregated by morphological and aspectual features.

The present study is a logical next step in the implementation of grammatical profiling. The goal is to approach aspect in modern Russian "bottom-up," asking the following questions:

- Can the aspect of individual verbs be determined purely on the basis of their grammatical profiles?
- Does aspect in Russian interact with genre?

Unlike Janda and Lyashevskaya, this study takes as input all forms[8] of all types of verbs, without pre-sorting the verbs according to aspect. In addition, this study stratifies the data according to genre. This step is motivated by two observations. The first is that there are differences in the aspectual interpretation of Russian verbs across genres (narration vs. conversation, see Padučeva "Režim interpretacii"). The second is that there are genre-related differences in the behavior of verbs (frequency of perfective vs. imperfective and in the most frequent lemmas for each aspect) in Czech (Bartoň et al. 166–68). These facts lead us to suspect that aspect might behave differently across genres in Russian.

3. Our Data Stratified Across Genres

There are a number of reasons for examining the grammatical profiles of verbs across genres. As mentioned above, it has been claimed that aspect can interact with genre, yet no corpus-based study of this possible interaction has been undertaken for Russian previously. Thus this study fills a gap in our knowledge about aspect. A second reason is that the vocabulary of different genres, while they of course overlap, can be largely different, making it possible to validate our findings across independent and clearly distinct datasets. If we make similar findings across multiple genres, we can, in effect, validate our findings. Such validation exemplifies the scientific method, according to which the most valuable findings are those that can be replicated. By showing that we get very similar results across three independent samples, we show that our model is valid. And finally, while it is presently impossible to

[8] For comparison, we also ran the analysis for the three genres using only the four subparadigms that were used in Janda and Lyashevskaya: nonpast, past, infinitive, and imperative. This removed all paradigmatic forms where the aspect is unambiguous or very strongly skewed in one direction, namely the gerunds, participles, and periphrastic future. The results on this more restricted dataset were essentially the same with the difference that Factor 1 could account for a much larger portion of the variance: for example, 59.6% for Journalistic prose as opposed to 39.1% when all verb forms are taken into account. However, the four subparadigms (nonpast, past, infinitive, imperative) alone give somewhat worse results: overall correct separation of perfective vs. imperfective verbs is achieved for only 82.5% of verbs, and the breakdown according to genre is as follows: Journalistic prose 88.5% correctly sorted, Fiction 81.7% correctly sorted, and Scientific-Technical prose 77.2% correctly sorted. Compare these results with those presented in Sections 4–6.

access sufficiently large samples of child-directed speech, if our findings are valid across three genres, then they are likely to be valid also for other genres, such as child-directed speech.

The present study is based on data from the manually disambiguated Morphological Standard of the Russian National Corpus, comprising approximately six million words from texts produced in the period of 1991–2012. Because this corpus is manually disambiguated, the data is relatively free of misidentifications of verb forms that could be caused by homonymy with other parts of speech such as noun (cf. *dulo* as Nom/Accsg of 'muzzle' vs. neuter sg past tense of *dut′* 'blow'; *stali* as the Gen/Dat/Locsg of *stal′* 'steel' vs. plural past tense of *stat′* 'become, begin'; *vypej* as Gen/Accpl of *vyp′* 'bittern' vs. imperative of *vypit′* 'drink'), pronoun (cf. *ves′* as the Nom/Accsg masculine 'all' vs. imperative of *vesit′* 'weigh'), preposition (cf. preposition *pri* 'at' vs. imperative of *peret′* 'move'), and numeral (cf. *tri* as 'three' vs. imperative of *teret′* 'rub').

We stratify the data across genres by using the meta-text annotation available in the Morphological Standard. Journalistic prose (*publicistika*) is the largest genre represented, with 1.2 million words, followed by Fiction (*xudožestvennaja literatura*), with 0.7 million words. By combining texts tagged as *učebno-naučnaja literatura* and *proizvodstvo-texničeskaja literatura*, we obtain a sample of 0.4 million words that we designate as Scientific-Technical prose. To make the samples as uniform as possible, we take equal amounts of data from each genre, namely 0.4 million words for each.

Note that if we extract the grammatical profiles of all verbs in a corpus, we get a lot of verb lemmas that have very few tokens. This is due to the relationship between frequency and rank called the "Zipf-Mandelbrot Law," according to which approximately half of the unique lemmas will be hapaxes (words attested only once) in any corpus of about 30,000 words or more (Zipf). This applies to verbs as well, so the majority of grammatical profiles of verbs that we can extract from a corpus will consist of only five or fewer tokens. For example, if we look up the low-frequency verb *ispolosovyvat′* 'be in the process of flogging to pieces or do so repeatedly' in the Russian National Corpus (ruscorpora.ru), we find three attestations: one of the infinitive, one of the indicative past, and one of a nonpast gerund. But is this enough data to characterize this verb as having a grammatical profile that is 33% infinitive, 33% indicative past, 33% nonpast gerund and 0% for all other subparadigms? Probably not. This verb is so rare that chance plays too big a role in the tokens that we find. For this reason, it makes sense to set a frequency threshold and examine only the grammatical profiles for verbs at or above that threshold.[9] For this study we set the threshold at fifty, meaning that we included

[9] For details on the setting of frequency thresholds for corpus data, see Berdičevskis and Eckhoff.

only the verb lemmas for which we have a sample of fifty or more tokens. Table 2 displays the distribution of verb tokens and lemmas across our three samples. For example, in the sample of Journalistic prose we found 52,716 verb forms representing 5,940 unique verbs (lemmas). However, only 185 of these verbs were represented by fifty or more tokens. Our study focuses only on those 185 verbs in the Journalistic sample, 225 verbs in the Fiction sample, and 172 verbs in the Scientific-Technical sample that crossed our threshold for inclusion.[10]

Table 2: Verb tokens and lemmas across the three samples stratified for genre

Genre	# Verb Tokens	# Verb Lemmas	Frequency ≥50 # Verb Lemmas
Journalistic	52,716	5,940	185
Fiction	78,084	8,665	225
Scientific-Technical	43,528	4,494	172

Russian verb forms can express values for a large number of grammatical categories, including person, number, gender, and even case (for participles). However, some of these categories are less likely to be relevant to the behavior of aspect, and a proliferation of categories can cause data sparseness and obscure important patterns. For this reason, we restrict this study to the verbal categories of tense and mood which are known to interact with aspect (see Janda and Lyashevskaya and references therein).[11] The subparadigms and their abbreviations (used in Figures 2–4 below) are presented in Table 3.

The indicative nonpast forms typically express present tense for imperfective verbs (as in *čitaju* 'I read'), but future for perfective verbs (as in *pročitaju* 'I will read'). Voice is not distinguished, so active and passive participles are grouped together in each of the subparadigms for participles. Some of the paradigms have strong aspectual tendencies or even absolute restrictions: indicative future, nonpast gerund and nonpast participle are all restricted to imperfective verbs, whereas past gerund and past participle are predominantly represented among perfective verbs. This means that aspectual competition is most relevant in four of the nine subparadigms: indicative nonpast, indicative

[10] There are actually 173 verbs that cross the frequency threshold in the Scientific-Technical prose sample, however, one of those verbs behaves so deviantly that it skews the entire distribution. This verb is *smotret'* 'look', which is attested 199 times in that sample. 165 of the attestations of this verb are imperative forms; however, 160 of those are merely the abbreviation *sm*., as in *sm. grafik/tablicu* X 'see Figure/Table X'. Given that these parenthetic abbreviations do not represent the use of verbs in prose, *smotret'* 'look' was excluded from the Scientific-Technical prose sample.

[11] Janda and Lyashevskaya begin their study by taking into account all categories associated with verbal forms (person, number, gender, etc. in addition to tense and mood), but then removed the categories that were not found relevant to aspect.

Table 3: Subparadigms of verbs in this study

Subparadigm name	**Abbreviation**	**Illustrative examples *čitat'/pročitat'* 'read'**
nonpast gerund	gernonpast	*čitaja*/NA
past gerund	gerpast	*čitav/pročitav*
imperative	imper	*čitaj*, *čitajte/pročitaj*, *pročitajte*
indicative future	indicfut	*budu čitat'*, *budeš' čitat'* etc./NA
indicative nonpast	indicnonpast	*čitaju*, *čitaješ'* etc./*pročitaju*, *pročitaješ'* etc.
indicative past	indicpast	*čital*, *čitala* etc./*pročital*, *pročitala* etc.
infinitive	inf	*čitat'/pročitat'*
nonpast participle	partcpnonpast	*čitajuščij*, *čitaemyj*/NA
past participle	partcppast	*čitavšij*, *čitannyj/pročitavšij*, *pročitannyj*

past, infinitive, and imperative. From Janda and Lyashevskaya we know that approximately 85% of verb tokens in the Russian National Corpus come from precisely these four subparadigms where both aspects are represented (722).

Following the example of Eckhoff and Janda, our data is analyzed "bottom-up" by means of correspondence analysis, with each verb represented as a row of figures (as in Table 1), and the subparadigms as our columns. As described above in Section 2, in calculating the distances between the arrays of figures, correspondence analysis mathematically constructs "Factors" and orders them according to their importance (measured in terms of the percentage of the variance in the data that they account for). It is important to note that each Factor is constructed around a zero point that uniquely balances the data. The values of each Factor are comparable to correlation values, such that positive values and negative values reflect opposite tendencies in the data (cf. Greenacre 45). Therefore it makes sense to interpret the result of correspondence analysis as a sorting of the data into two groups along Factor 1: data points with negative values vs. data points with positive values.[12] The specific orientation of Factor 1 is not essential since it is a mathematical construct that has meaning only when interpreted in relation to the data. In other words, Factor 1 tells us that the data can be divided into two groups, but the identity of what those two groups are is a matter of interpretation.[13]

Again, in layman's terms this means that we are giving the computer program a matrix of verbs (rows) and the frequencies of their forms in the various subparadigms (columns). The program's task is to find mathematically plausible ways to divide the rows and the columns into two groups. The most plausible way to do this is calculated and labelled "Factor 1". And it happens

[12] Kamphuis (78, 152) claims that the zero point on Factor 1 is "arbitrary" and that the zero point "does not have a clear meaning," but this is at odds with the descriptions of statistical experts who have designed such models, such as Greenacre.

[13] In Eckhoff and Janda, Factor 2 was interpretable as tense. In the current study Factor 2 does not have such a straightforward interpretation. However, our main focus in this study is on Factor 1, which is clearly interpretable as aspect.

that the division made by Factor 1 is a very close approximation to perfective vs. imperfective aspect.

The input to our correspondence analysis is merely the grammatical profiles of the verbs in the three genres that are represented by 50 or more tokens in our samples. In other words, the computer does not know which verbs belong to which aspect. The only information that is fed into the analysis is the grammatical profiles (the distributions of forms). Only after running the correspondence analysis do we introduce aspectual information in the second step. The second step involves labeling the verbs as "p" for perfective, "i" for imperfective, and "b" for biaspectual and superimposing these labels over the datapoints after the correspondence analysis has done its job. This use of overlaid aspect labels to replace points representing verbs facilitates visualization of the data and interpretation of the results. In addition, we inspect the Factor 1 values assigned to all the verbs.

In all three genres, we find that Factor 1 (the strongest factor) accounts for more than 30% of the variance (39.1% for the Journalistic sample, 31.4% for Fiction, and 35.1% for Scientific-Technical prose), and that Factor 1 is clearly interpretable as aspect. On average across the genres, Factor 1 is 93% accurate in sorting verbs such that those on one side of zero are mainly perfective verbs, and those on the other side of zero are mainly imperfectives. The specific orientation of Factor 1 happens to be different for Fiction than for the other two genres, but that is not significant. In the following three sections, we examine the results for each genre in turn, focusing particularly on verbs that lie at the extremes, verbs that lie in the middle, and verbs that land on the "wrong" side of the zero line (i.e., perfectives that have been grouped with imperfectives and imperfectives that have been grouped with perfectives). These misclassified verbs will be henceforth referred to as "deviations." Deviations are sorted according to how far they lie from zero on Factor 1, with weak deviations lying less than 0.1 units away, and strong deviations lying 0.1 or more units away. The specific grammatical profiles of strong deviations are analyzed in further detail.

All of the data, R code used in the analysis, and full lists of verbs and their Factor 1 values can be accessed from TROLLing (the Tromsø Repository of Language and Linguistics) at doi:10.18710/BIIGT6.

4. Journalistic Prose

Figure 2 presents the correspondence analysis of the grammatical profiles of verbs in the sample of Journalistic prose. The perfective verbs on the whole fall to the left of zero on the x-axis (Factor 1), whereas the majority of imperfective verbs are to the right.

Figure 2 likewise plots the subparadigms, showing their correlation with Factor 1 on the x-axis. The four subparadigms that were also represented in the Janda and Lyashevskaya study behave the same here: indicative past, infinitive

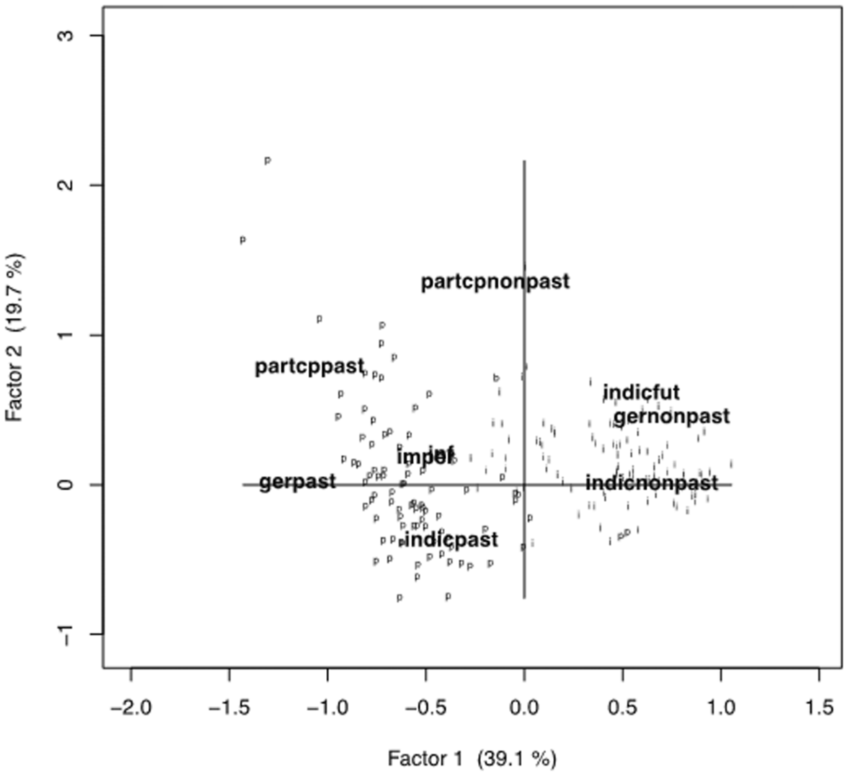

Figure 2: Correspondence analysis of Journalistic prose with aspect labels (p = perfective, i = imperfective, b = biaspectual) overlaid

and imperative (the latter two superimposed on Figure 2) are all correlated with perfective verbs (negative values for Factor 1), while the indicative nonpast is correlated with imperfective verbs (positive values for Factor 1). The remaining subparadigms land approximately where we would expect them, with past participles and past gerunds with the perfectives, indicative future and nonpast gerunds with the imperfectives. The nonpast participles land in the middle of the plot. The subparadigms that land furthest to the left and right of the diagram are those that are most decisive in sorting the verbs on Factor 1: these are the past participles and past gerunds that are the strongest indicators of negative Factor 1 values, as opposed to the nonpast gerund, indicative nonpast, and indicative future that are the strongest indicators of positive Factor 1 values. From the perspective of the model, these are the subparadigms that are doing the most “work” in terms of sorting the verbs.

Table 4 summarizes the results in terms of how well Factor 1 sorts the verbs according to aspect. Of the 185 verbs in this sample, 87 are perfectives and

96 are imperfectives. There are in addition two biaspectual verbs, both of which fall among the perfectives: *ispol'zovat'* 'use' (Factor 1 value −0.142) and *obeščat'* 'promise' (Factor 1 value −0.030). There are altogether sixteen deviations in this sample, three involving perfective verbs that have landed with imperfectives, and thirteen of imperfective verbs that have landed with perfectives. The bottom row of Table 4 presents the accuracy of Factor 1 in predicting aspect. In the All Verbs column, the two biaspectual verbs have been removed from the calculation of accuracy. So for Journalistic prose All verbs = 185 − 2 = 183 and then the accuracy was calculated as (183-16)/183 to get 91.3% correctly predicted by Factor 1.

Table 4: Overview of sorting of verbs in Journalistic prose[14]

	All Verbs	Perfective Verbs	Imperfective Verbs	Biaspectual Verbs
Total # Verbs	185	87	96	2
# Deviations	16	3	13	NA
Factor 1 Correctly Predicts Aspect	91.3%	96.6%	86.5%	NA

Table 5 gives a sample of the data at the two extremes and immediately on either side of zero in the middle of the Factor 1 distribution. The top left quadrant of the table gives the ten perfective verbs that are furthest to the left in Figure 2, while the imperfective verbs at the other extreme of the Factor 1 continuum are in the top right quadrant. The perfective verbs in the top of Table 5 are resultative verbs referring to specific events; many of them also usually express achievements like *otkryt'* 'open' and *obnaružit'* 'reveal'. The imperfective verbs at the top of Table 5 are nearly all statives like *javljat'sja* 'be' and *zaviset'* 'depend on'.

At the bottom half of Table 5 we find the perfective and imperfective verbs that are closest to zero on Factor 1. Several of the perfectives are rather generic verbs that express simply that something happened (*pojavit'sja* 'appear', *sostojat'sa* 'take place', *polučit'sja* 'turn out'), and several express social transactions (*napomnit'* 'remind', *pozvolit'* 'allow', *pomoč'* 'help'). The imperfective verbs close to zero express both states and activities, and several are relatively non-specific all-purpose verbs: *zanimat'sja* 'be occupied', *delat'* 'do', *byt'* 'be', *dejstvovat'* 'act'.

Table 6 presents all the deviations (misclassified verbs) that are found in the Journalistic prose sample, with the three perfectives on the left and the 13 imperfectives on the right. The data in Table 6 is sorted so that deviations with the largest absolute values for Factor 1 are at the top, with values decreasing downward. Deviations with absolute values of 0.1 or more are shaded to in-

[14] The difference between 96.6% for perfective verbs vs. 86.5% for imperfective verbs is borderline statistically significant: X-squared = 3.8298, df = 1, p-value = 0.05035, Cramer's V = 0.27.

Table 5: Most and least extreme perfective and imperfective verbs in Journalistic prose

	Perfective Verbs		Imperfective Verbs	
	Verb	Factor 1	Verb	Factor 1
10 Most Extreme Verbs	*posvjatit'* 'dedicate'	−1.431	*javljat'sja* 'be'	1.054
	vjazat' 'tie'	−1.304	*polagat'* 'suppose'	0.942
	naznačit' 'appoint'	−1.041	*nravit'sja* 'please'	0.931
	otkryt' 'open'	−0.945	*kasat'sja* 'concern'	0.914
	sobrat' 'gather'	−0.932	*pomnit'* 'remember'	0.903
	obnaružit' 'reveal'	−0.917	*stoit'* 'be worth'	0.890
	ostavit' 'leave'	−0.867	*zaviset'* 'depend on'	0.885
	peredat' 'transfer'	−0.843	*stanovit'sja* 'become'	0.871
	postavit' 'place'	−0.822	*byvat'* 'be'	0.868
	snjat' 'remove'	−0.812	*polučat'sja* 'turn out'	0.854
10 Least Extreme Verbs	*posmotret'* 'look'	−0.356	*igrat'* 'play'	0.125
	pojavit'sja 'appear'	−0.319	*brat'* 'take'	0.111
	privesti 'bring'	−0.293	*sozdavat'* 'create'	0.098
	vystupit' 'adress'	−0.275	*exat'* 'ride'	0.097
	sostojat'sja 'take place'	−0.198	*žit'* 'live'	0.091
	polučit'sja 'turn out'	−0.172	*zanimat'sja* 'be occupied'	0.079
	napomnit' 'remind'	−0.111	*delat'* 'do'	0.062
	pozvolit' 'allow'	−0.047	*byt'* 'be'	0.042
	pomoč' 'help'	−0.044	*dejstvovat'* 'act'	0.010
	pojti 'go'	−0.005	*nazyvat'* 'name'	0.005

Table 6: Verbs in Journalistic prose misclassified by Factor 1

Perfective Deviations		Imperfective Deviations	
Verb	Factor 1	Verb	Factor 1
prijtis' 'be necessary'	0.524	*smotret'* 'look'	−0.273
smoč' 'be able'	0.491	*učastvovat'* 'participate'	−0.239
obojtis' 'do without'	0.029	*čitat'* 'read'	−0.196
		platit' 'pay'	−0.164
		borot'sja 'struggle'	−0.158
		ožidat' 'expect'	−0.126
		provodit' 'carry out'	−0.114
		rassčityvat' 'reckon'	−0.097
		iskat' 'seek'	−0.094
		prinimat' 'accept'	−0.078
		xodit' 'walk'	−0.040
		rešat' 'solve'	−0.011
		ezdit' 'ride'	−0.003

dicate that these are the strongest deviations, in contrast to weaker deviations that are very near to zero (absolute Factor 1 values <0.1).

The two strong perfective deviations in the Journalistic prose sample are also modal verbs: *prijtis'* 'be necessary', *smoč'* 'be able'. These two deviations are pushed across the zero line by their strong affinity to the indicative nonpast, which is otherwise associated with imperfectives. Both verbs are found almost exclusively in the indicative nonpast and past forms (with the exception of one attestation of an infinitive for *prijtis'* 'be necessary'). The relative distribution for these verbs is: *prijtis'* 'be necessary' 64.9% indicative nonpast vs. 34.5% indicative past, *smoč'* 'be able' 63% indicative nonpast vs. 37% indicative past.

All of the strong deviations among the imperfectives show a strong affinity for the infinitive, ranging from 53.7% for *borot'sja* 'struggle' to 27.4% for *smotret'* 'look'. Two verbs in this group also have unusually high proportions of imperative forms in their grammatical profiles: *smotret'* 'look' with 29.3%, and *čitat'* 'read' with 12.9%. In addition, *učastvovat'* 'participate' has many past tense forms (38.2%), and some of the weaker deviations are also often used in the past tense, for example the indeterminate motion verbs *xodit'* 'walk' and *ezdit'* 'ride' to express a round trip.

None of the strong deviations are attested in past gerund forms, but past participles are found for *borot'sja* 'struggle', *ožidat'* 'expect', *smotret'* 'look', and *učastvovat'* 'participate'.

5. Fiction

The correspondence analysis for our Fiction sample is represented in Figure 3, where Factor 1 sorts the perfective verbs on the right of zero from the imperfective verbs on the left of zero. Although this graph has the opposite orientation for the x-axis, the correspondences remain similar. The infinitive, indicative past, past participle and past gerund are all on the same side as the perfective verbs, and the indicative future, indicative nonpast, nonpast gerund and nonpast participle are all on the same side as the imperfective verbs. The imperative clings to the zero line, which indicates that it does not correlate closely with aspect in this genre.

Table 7 reveals that Factor 1's sorting of verbs according to aspect is almost exactly the same for Fiction as for Journalistic prose, with an overall accuracy of 91.5%. There is only one biaspectual verb in this sample, *obeščat'* 'promise'

Table 7: Overview of sorting of verbs in Fiction

	All Verbs	Perfective Verbs	Imperfective Verbs	Biaspectual Verbs
Total # Verbs	225	114	110	1
# Deviations	19	6	13	NA
Factor 1 Correctly Predicts Aspect	91.5%	94.7%	88.1%	NA

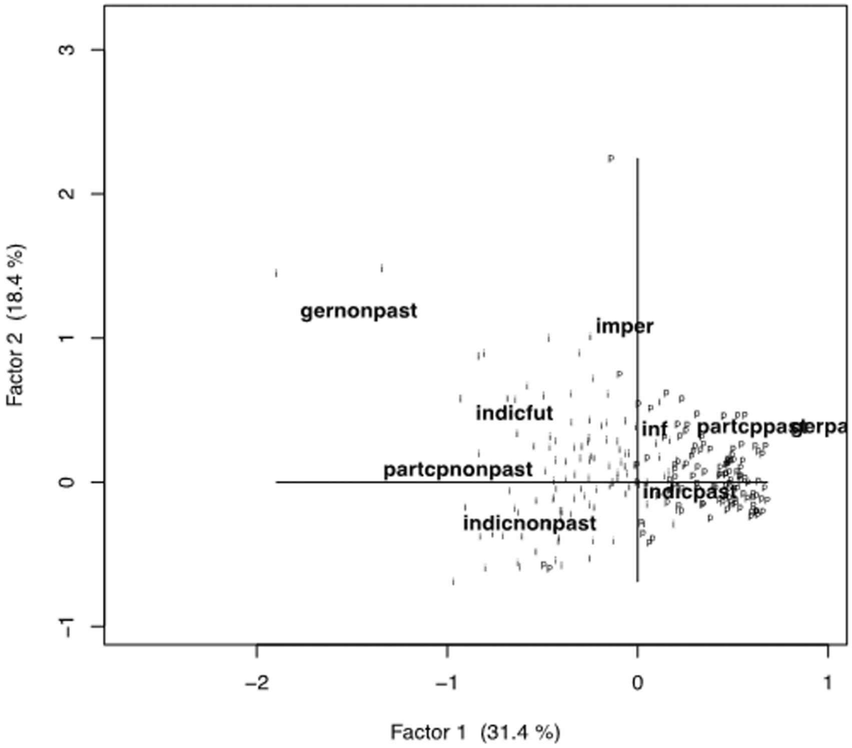

Figure 3: Correspondence analysis of Fiction with aspect labels (p = perfective, i = imperfective, b = biaspectual) overlaid

(Factor 1 value 0.212), which patterns with the perfective verbs as it does in the Journalism sample, though more strongly so in Fiction.

Table 8 (like Table 5) presents ten verbs in each quadrant, displaying the perfective and imperfective verbs at the far extremes of the distribution in the top half of the table, and the verbs closest to the zero line in the bottom half.

The overall impression these verbs make is very similar: once again we find clearly resultative verbs among the most extreme perfectives and stative verbs among the most extreme imperfectives. However, we also find four punctual verbs suffixed in *–nu-* among the extreme perfectives, and the verbs themselves are almost completely different. The three verbs shared across Tables 5 and 8 are *polučit'sja* 'turn out' (a non-extreme perfective), *pomnit'* 'remember' (an extreme imperfective), and *igrat'* 'play' (a non-extreme imperfective). These verbs behave similarly across Journalistic prose and Fiction.

Table 9 shows the deviations found in the Fiction data.

There are three strong perfective deviations: *xvatit'* 'be enough', *smoč'* 'be able' and *prostit'* 'forgive'. Recall that *smoč'* is also a strong deviation for Journalistic prose. Once again this deviation can be attributed to an unusual

Table 8: Most and least extreme perfective and imperfective verbs in Fiction

	Perfective Verbs		Imperfective Verbs	
	Verb	Factor 1	Verb	Factor 1
10 Most Extreme Verbs	*proiznesti* ‘pronounce’	0.681	*želat'* ‘desire’	−1.897
	prosnut'sja ‘wake up’	0.672	*gljadet'* ‘look’	−1.345
	privyknut' ‘get used to’	0.665	*značit'* ‘mean’	−0.966
	zakričat' ‘yell’	0.656	*ulybat'sja* ‘smile’	−0.930
	podnjat' ‘lift’	0.655	*ponimat'* ‘understand’	−0.904
	vzdoxnut' ‘sigh’	0.651	*vspominat'* ‘recall’	−0.837
	poterjat' ‘lose’	0.635	*pokazyvat'* ‘show’	−0.835
	uslyšat' ‘hear’	0.635	*pomnit'* ‘remember’	−0.829
	progovorit' ‘speak’	0.630	*starat'sja* ‘try’	−0.809
	kivnut' ‘nod’	0.624	*nazyvat'sja* ‘be called’	−0.802
10 Least Extreme Verbs	*dat'* ‘give’	0.151	*stavit'* ‘place’	−0.087
	rasskazat' ‘narrate’	0.141	*igrat'* ‘play’	−0.068
	ustroit' ‘organize’	0.108	*lezt'* ‘crawl’	−0.066
	končit'sja ‘end’	0.078	*učit'sja* ‘learn’	−0.063
	otdat' ‘submit’	0.067	*ostavat'sja* ‘remain’	−0.063
	prijtis' ‘be necessary’	0.062	*vesti* ‘lead’	−0.056
	ubit' ‘kill’	0.049	*razgovarivat'* ‘converse’	−0.049
	polučit'sja ‘turn out’	0.028	*prixodit'* ‘arrive’	−0.045
	uexat' ‘depart’	0.019	*est'* ‘eat’	−0.016
	pogovorit' ‘speak’	0.003	*rasskazyvat'* ‘narrate’	−0.011

Table 9: Verbs in Fiction misclassified by Factor 1

Perfective Deviations		Imperfective Deviations	
Verb	Factor 1	Verb	Factor 1
xvatit' ‘be enough’	−0.492	*bit'* ‘beat’	0.236
smoč' ‘be able’	−0.463	*byt'* ‘be’	0.206
prostit' ‘forgive’	−0.140	*kazat'sja* ‘seem’	0.189
pomoč' ‘help’	−0.096	*streljat'* ‘shoot’	0.166
poexat' ‘go’	−0.005	*prodolžat'* ‘continue’	0.163
napisat' ‘write’	−0.005	*ezdit'* ‘ride’	0.138
		provodit' ‘carry out’	0.115
		molčat' ‘be silent’	0.115
		zvonit' ‘ring’	0.094
		stojat' ‘stand’	0.048
		viset' ‘hang’	0.047
		vygljadet' ‘look’	0.035
		kurit' ‘smoke’	0.008

affinity for nonpast tense forms, which make up 56.4% of this verb's grammatical profile (the remaining 43.6% of its forms are past tense). *Xvatit'* is very similar, with 59.1% nonpast forms, while *prostit'* is deviant due to its strong preference for the imperative (69.1%).

For the Fiction genre, the imperfective deviations are largely accounted for by an attraction to the indicative past. Among the strong deviations, the proportion of indicative past ranges from 76.1% for *prodolžat'* 'continue' to 46.4% for *streljat'* 'shoot'. *Byt'* 'be' is among the least extreme imperfectives for Journalism, but appears as a deviation for Fiction, where 71.9% of its grammatical profile consists of indicative past forms. This finding supports Padučeva's ("O biaspektual'nosti") claim that Russian *byt'* 'be' is actually a biaspectual verb, with perfective interpretations when it expresses resultative bidirectional movements such as 'visit', 'go to a place'. *Provodit'* 'carry out' is among the strong deviations for both Journalism and Fiction, but whereas the indicative past is most prominent in Fiction (accounting for 48.1% of the grammatical profile), it is the infinitive form of this verb that makes the strongest showing in Journalistic prose (36.3% of the profile). *Ezdit'* 'ride' appears among the deviations in both genres as well, but is a strong deviation only in Fiction.

6. Scientific-Technical Prose

Figure 4, the correspondence analysis plot for Scientific-Technical prose, looks very similar to those for the other two genres, sorting the verbs according to aspect along Factor 1. The past participle, past gerund, imperative, indicative past, and infinitive (note that these last two are superimposed) are on the same side of zero as the perfective verbs, whereas the indicative nonpast, indicative future, nonpast gerund and nonpast participle are on the side of the imperfective verbs.

Table 10 summarizes the relationship between Factor 1 and aspect for Scientific-Technical prose, showing that the correspondence is best of all for this genre, with aspect correctly predicted over 95% of the time across the board.

Table 10: Overview of sorting of verbs in Scientific-Technical prose

	All Verbs	Perfective Verbs	Imperfective Verbs	Biaspectual Verbs
Total # Verbs	172	64	103	5
# Deviations	7	2	5	NA
Factor 1 Correctly Predicts Aspect	95.8%	97.0%	95.2%	NA

The five biaspectual verbs and their Factor 1 values in this sample are as follows: *realizovat'* 'realize' (−0.551) and *organizovat'* 'organize' (−0.316) pattern with the perfectives, while *ispol'zovat'* 'use' (−0.016) is very close to

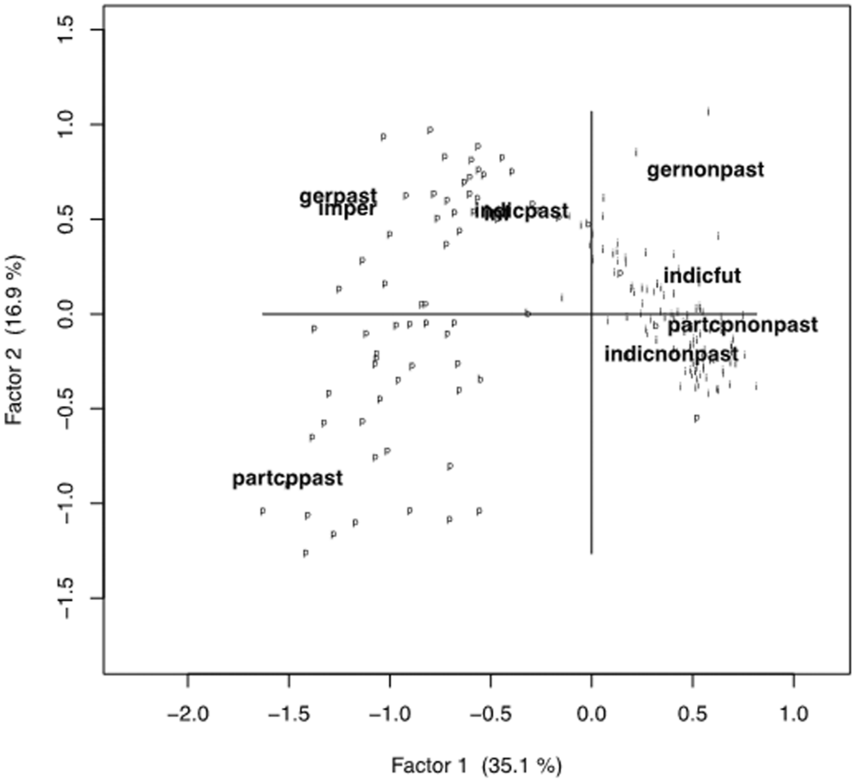

Figure 4: Correspondence analysis of Scientific-Technical prose with aspect labels (p = perfective, i = imperfective, b = biaspectual) overlaid

the zero line and and *issledovat'* 'examine' (0.171) and *ispol'zovat'sja* 'be used' (0.318) pattern with the imperfectives.

Table 11 has the same structure as Tables 5 and 8, showing the most and least extreme perfective and imperfective verbs in the Scientific-Technical prose sample. *Svjazat'* 'tie' turns up as an extreme perfective in both Journalistic and Scientific-Technical prose, while *obnaružit'* 'reveal' is among the ten most extreme perfectives in Journalistic prose, but among the least extreme perfectives in Scientific-Technical prose. *Pojti* 'go', and *pozvolit'* 'allow' are among the least extreme perfectives in both genres. The verb *dat'* 'give' is among the least extreme for both Fiction and Scientific-Technical prose.

On the imperfective side, *žit'* 'live' and *delat'* 'do' appear among the least extreme imperfectives in both Journalistic and Scientific-Technical prose. Likewise, *vesti* 'lead' is a non-extreme imperfective in both Fiction and Scientific-Technical prose. *Provodit'* 'carry out' is among the least extreme imperfectives in Scientific-Technical prose but a strong deviation in both of the other two genres.

Table 11: Most and least extreme perfective and imperfective verbs in Scientific-Technical prose

	Perfective Verbs		Imperfective Verbs	
	Verb	Factor 1	Verb	Factor 1
10 Most Extreme Verbs	*složit'sja* ‘form’	−1.629	*podležat'* ‘submit’	0.816
	ukazat' ‘indicate’	−1.507	*vxodit'* ‘enter (into)’	0.760
	svjazat' ‘tie’	−1.415	*opisyvat'* ‘describe’	0.751
	napravit' ‘direct’	−1.406	*imet'sja* ‘exist’	0.713
	vydelit' ‘separate’	−1.385	*predlagat'* ‘suggest’	0.699
	vybrat' ‘select’	−1.373	*obladat'* ‘possess’	0.693
	zaključit' ‘conclude’	−1.325	*otražat'* ‘reflect’	0.693
	polučit' ‘receive’	−1.301	*trebovat'* ‘require’	0.690
	rasprostranit' ‘spread’	−1.278	*sootvetstvovat'* ‘correspond’	0.687
	rassmotret' ‘examine’	−1.251	*pozvoljat'* ‘allow’	0.686
10 Least Extreme Verbs	*uvidet'* ‘see’	−0.563	*ožidat'* ‘expect’	0.131
	ocenit' ‘evaluate’	−0.561	*vesti* ‘lead’	0.125
	prednaznačit' ‘predetermine’	−0.556	*otvečat'* ‘answer’	0.110
	okazat'sja ‘turn out’	−0.535	*videt'* ‘see’	0.107
	dat' ‘give’	−0.464	*proxodit'* ‘go through’	0.079
	stat' ‘become’	−0.444	*govorit'* ‘speak’	0.059
	obnaružit' ‘reveal’	−0.396	*provodit'* ‘carry out’	0.056
	pojti ‘go’	−0.293	*delat'* ‘do’	0.053
	pozvolit' ‘allow’	−0.270	*žit'* ‘live’	0.005
	pokazat' ‘show’	−0.162	*izučat'* ‘study’	0.003

The Scientific-Technical prose sample yields eight deviations, as shown in Table 12.

Table 12: Verbs in Scientific-Technical prose misclassified by Factor 1

Perfective Deviations		Imperfective Deviations	
Verb	Factor 1	Verb	Factor 1
zamknut' ‘close, lock’	0.521	*pisat'* ‘write’	−0.223
smoč' ‘be able’	0.141	*provodit'sja* ‘be carried out’	−0.147
		byt' ‘be’	−0.111
		sčitat' ‘consider’	−0.054
		učastvovat' ‘participate’	−0.013

Once again we see that *smoč'* ‘be able’ is a perfective verb that patterns strongly with imperfectives, as in both Journalistic prose and Fiction. And once again the cause for misclassification of perfective verbs is their affinity for indicative nonpast forms. Indicative nonpast forms dominate the grammatical profiles of both *smoč'* ‘be able’ (58.8% indicative nonpast, the remaining 41.2% indicative past) and *zamknut′* 'close, lock' (96.5% indicative nonpast, the remaining 3.5% past participle). Among the misclassified imperfective verbs, both indicative past (values ranging from 48.8% for *pisat′* 'write' to 16.2% for *sčitat′* 'consider') and, to a lesser extent, infinitive forms (values ranging from 32.3% for *sčitat′* 'consider' to 13.1% for *pisat′* 'write') have dragged these verbs over toward the perfectives.

7. Comparisons Across Genres

Paradigmatic cues, as expressed by Factor 1 values, perform very well in sorting perfective verbs from imperfectives, with an overall accuracy of 92.7% = (582 verbs − 8 biaspectuals − 42 deviations) / (582 verbs − 8 biaspectuals) * 100. Accuracy is best of all for Scientific-Technical prose, and best in each genre for perfective verbs. When perfective verbs are misclassified, this is almost always due to an affinity for indicative nonpast forms. When imperfectives are misclassified, the main cause is different for each genre: an affinity for infinitive and imperative forms in Journalistic prose, for indicative past forms in Fiction, and a combination of infinitive and indicative past for Scientific-Technical prose. Two verbs are particularly prominent among the deviations: *smoč′* 'be able' and *byt′* 'be'. We examine each verb individually.

Smoč′ 'be able' is the only strong deviation that appears in all three genres. In every case, we see that this verb has more indicative nonpast forms than one would expect for a perfective verb: 63% in Journalistic prose, 56.4% in Fiction, and 58.8% in Scientific-Technical prose. The answer might be that this verb has been recruited to express imperfective indicative future in Russian. Choi argues that the corresponding imperfective *moč′* 'be able' is morphologically defective, lacking indicative future forms **budu moč′*, **budeš′ moč′*, etc. (50). As a result, the indicative nonpast forms of *smoč′* 'be able' fill this gap and, according to Choi, "the 'aspectual' meaning is neutralized." In other words, the indicative nonpast forms *smogu*, *smožeš′* etc. express future tense, but do not necessarily express perfective aspect.

Byt′ 'be' patterns with the least extreme imperfectives in Journalistic prose, but appears as a strong deviation in both Fiction and Scientific-Technical prose. Of course, *byt′* 'be' is unusual on many counts: it is the only verb with a zero form in the nonpast (which is not counted, of course, as our data include only observed forms), it is the only verb with a non-periphrastic indicative future, it is used far more often in the past tense than most imperfective verbs (indicative past makes up 71.9% of its grammatical profile for Fiction), and it has some perfective-like interpretations (Padučeva "O biaspektual'nosti").

While most of the subparadigms have a stable relationship to either perfective or imperfective aspect, three of them vary across genre. The infinitive and imperative are clearly associated with perfective in Journalistic and

Scientific-Technical prose, but in Fiction the infinitive is weakly associated with imperfective and the imperative lands in the middle of the plot. Non-past participles are clearly associated with imperfective in both Fiction and Scientific-Technical prose, but land in the middle of the plot for Journalistic prose.

In Section 1 we referenced the fact that morphology is not entirely reliable as a predictor of aspect in Russian. Our correspondence analysis facilitates a comparison of different morphological types of verbs, and shows that even when the morphological cues for aspect are unambiguous, the behavior of verbs can vary. Figure 5 presents the distributions of Factor 1 values for verbs in the Journalistic prose sample grouped according to relevant aspectual morphology.[15] The x-axis is Factor 1 values, and the morphological types are arranged along the y-axis. The labels of the groups are explained with examples in Table 13.

The verbs *smočʹ* 'be able' and *bytʹ* 'be' are represented individually due to their unusual behavior documented above. The distribution of verbs of each type is represented by a "box and whiskers" plot of their Factor 1 values, where the thick line indicates the median (the point at which 50% of the verbs fall above and 50% fall below that value), the edges of the box show the so-called "interquartile range" (the left edge shows the point at which 25% of the distribution below the median is included, and the right edge marks 75% of the distribution), the whiskers (dotted lines) show the location of data points that lie within 1.5 times the interquartile range, and circles indicate the position of outliers.

We see that there is a good correspondence between the Factor 1 value and aspect for most, but not all, morphological types of verbs. The perfectives form the clearest cluster. Prefixed perfectives, prefixed determinate motion verbs, and simplex perfective verbs all have their medians squarely on the perfective side, with values more than 0.5 from the zero line. However, note that the prefixed perfectives also have a very wide distribution, with outliers on both sides. The primary imperfectives, secondary imperfectives, and prefixed indeterminate verbs similarly have their medians all clearly on the imperfective side. Note, however, that the secondary imperfectives fall further to the imperfective side than the primary imperfectives, and for the Journalism sample this difference is significant (a t-test yields $t = -2.39$, $df = 82.27$, p-value $= 0.019$ and Cohen's $D = 0.52$, indicating a medium effect size). These data show three clusters of groups: i) the three perfective types all very close together and well removed from zero on the perfective side, ii) the simplex imperfectives and prefixed indeterminate motion verbs on the im-

[15] Plots for the Fiction and Scientific-Technical prose samples give results that are overall similar to Figure 5.

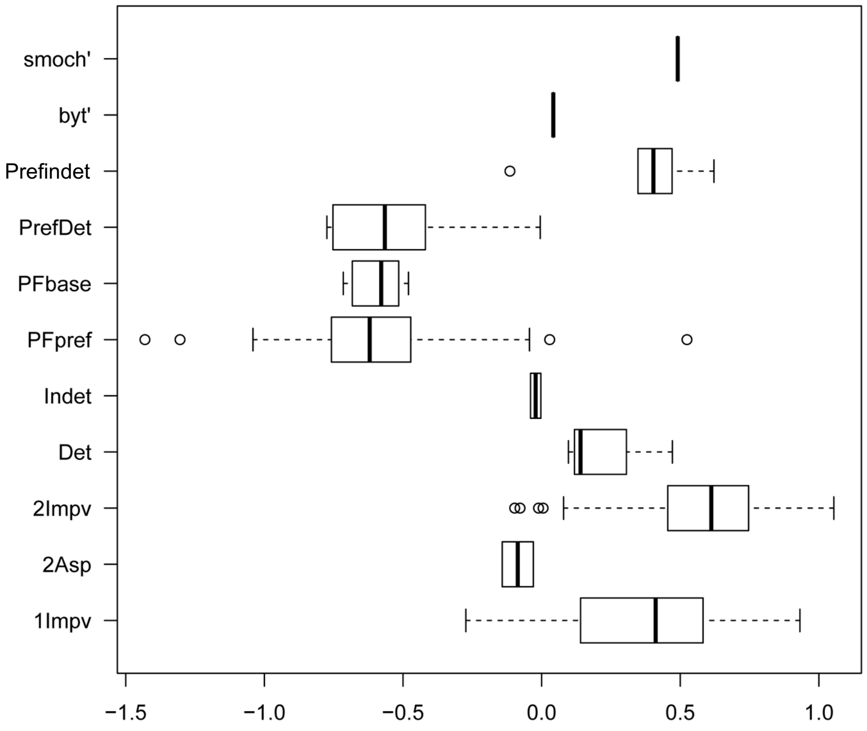

Figure 5: Aspectual morphology and Factor 1 values for Journalistic prose (see labels in Table 13)

Table 13: Labels in Figure 5 with examples

Abbreviation in Figure 5	Label Spelled Out	Example Verb
smoch'		*smoč'* 'be able'
byt'		*byt'* 'be'
PrefIndet	prefixed indeterminate motion verb	*provodit'* 'conduct'
PrefDet	prefixed determinate motion verb	*vyjti* 'exit'
PFbase	simplex perfective verb	*dat'* 'give'
PFpref	prefixed perfective verb	*napisat'* 'write'
Indet	indeterminate motion verb	*xodit'* 'walk'
Det	determinate motion verb	*idti* 'walk'
2Impv	secondary imperfective verb	*pokazyvat'* 'show'
2Asp	biaspectual verb	*obeščat'* 'promise'
1Impv	primary imperfective verb	*igrat'* 'play'

perfective side but less far removed from zero, and iii) the secondary imperfectives furthest removed from zero on the imperfective side.[16]

The biaspectual verbs are, unsurprisingly, very close to zero on the perfective side. The unprefixed motion verbs are interesting. While both of them are located near the middle of the continuum, the determinate verbs fall on the imperfective side, but the indeterminate verbs fall on the "wrong" side with the perfectives. This is possibly due to the frequent use of the verbs *xodit'* 'walk' and *ezdit'* 'ride' in the past tense to refer to a round trip, as mentioned above.

Paradigmatic cues reliably sort two types of verbs that lack overt aspectual morphology: the Factor 1 value distributions of primary imperfectives and perfective base verbs do not overlap. However, we see that many of the morphological types have rather wide distributions, and this is especially true of the types where the morphological aspectual marking is unambiguous, namely the prefixed perfectives and the secondary imperfectives. Both morphological types include verbs that cross the zero line on Factor 1, with individual verbs that behave more like the opposite aspect.

In other words, the paradigmatic cues sort out aspect particularly well precisely for the groups of verbs that cannot be sorted out reliably on the basis of aspectual morphology, namely simplex verbs that lack morphological markers. There we see a clear separation between perfective and imperfective base verbs that can be deduced solely from their grammatical profiles, which neither overlap nor show outliers. In the case of groups of verbs that have unambiguous morphological marking, namely the prefixed perfectives and suffixed secondary imperfectives, while the overall trends are similar, the distributions are more scattered and show outliers. This is particularly the case with the prefixed perfectives, which have the widest distribution of Factor 1 values.

8. Comparison of Accuracy of Paradigmatic vs. Morphological Cues

We have shown that by relying only on the paradigmatic cues of grammatical profiles of verbs, it is possible to predict the aspect of Russian verbs with 92.7% accuracy across the three genres of Journalism, Fiction and Scientific-Technical prose. Since Russian makes systematic use of overt morphological aspect markers, it is helpful to compute how well one could predict aspect on the basis of such markers and compare those results with the results we have obtained for paradigmatic cues. In order to make this comparison meaningful, it should of course be carried out on the same dataset. In other words, we

[16] Note that contrary to Janda and Lyashevskaya, where no difference was found between aspectual pairs overtly marked by prefixation vs. those overtly marked by suffixation, this study reveals that suffixation pushes the secondary imperfective verbs farther away from perfective verbs, yielding a larger difference than that observed between simplex imperfectives and prefixed perfectives.

will calculate the accuracy of predictions of aspect based on morphological cues for the same three samples. And once again, by using three independent samples from three different genres, we can validate our model and check for possible variance.

As in Sections 4–6 above, we compute accuracy as the number of verbs guessed correctly (the numerator in our formula) divided by the total number of verbs to be guessed (the denominator in our formula), and multiplied by 100, to obtain a percent. We must first remove the biaspectual verbs from the total number of verbs in our denominator, since they do not express aspect inherently (their aspect depends on context). So our denominator is the total number of verbs minus the biaspectual verbs (Total verbs − 2Asp). Thus the denominator is 183 verbs for Journalism, 224 verbs for Fiction, and 167 verbs for Scientific-Technical prose (the same numbers as in our previous calculations for paradigmatic cues). The numerator is the value of the denominator minus the number of verbs guessed incorrectly (or, alternatively, just the number of verbs guessed correctly).

In order to give a broader perspective, we will use two different calculations for morphological cues: a "simple" model based only on aspectual prefixes and suffixes, and a "complex" model that also includes the interaction of types of motion verbs with aspectual morphology. Here are the assumptions made by the simple model:

1. *simplex verbs* (verbs lacking overt aspectual markers) *are imperfective*
   [this assumption is mostly correct, but wrong for perfective simplex verbs]
2. *prefixed verbs are perfective*
   [this assumption is mostly correct, but wrong for prefixed indeterminate verbs]
3. *verbs suffixed in -y/iva*, *-va*, *and -(j)a are imperfective*
   [this assumption is always correct]

This simple model with three assumptions based on morphological cues can be applied to the Journalism and Scientific-Technical samples. But the model should be modified for Fiction, which is the only genre sample containing -*nu* suffixed semelfactives that crossed our frequency threshold, in fact five of them. However, the Fiction sample also has two -*nu* suffixed imperfective verbs: *paxnut'* 'smell' and *tjanut'* 'pull'. Therefore, we add this assumption, which is relevant only for Fiction:

4. *verbs suffixed in -nu are perfective*
   [this assumption is correct for 5 verbs, incorrect for 2 verbs]

These assumptions make it possible to calculate the numerator in our formula by subtracting the number of perfective simplex verbs and the number of prefixed indeterminate verbs from the number in the denominator. In the case of Fiction, we also need to subtract 2 to correct for the assumption that *-nu* marks perfective verbs. The results of these calculations based on a simple model for morphological cues are presented in the middle column of Table 14, alongside the results reported for paradigmatic cues in Sections 4–6.

In addition, we could calculate the accuracy of a more complex model which takes into account the behavior of motion verbs. In the complex model, assumptions 1 and 3–4 remain, but assumption 2 is revised as follows:

2. *prefixed determinate motion verbs are perfective*, *prefixed indeterminate motion verbs are imperfective*, *and all other prefixed verbs are perfective* [this assumption is always correct, at least for our data]

Thus, in order to obtain the numerator for the complex model, we subtract from the value of the denominator only the number of perfective simplex verbs (and also subtract 2 for Fiction, as above). These results are presented in the rightmost column in Table 14.

Table 14: Comparison of accuracy for paradigmatic vs. morphological cues

	Paradigmatic Cues Alone	Morphological Cues Alone (Simple Model)	Morphological Cues Plus Interaction with Motion Verbs (Complex Model)
Journalism	(167/183)*100 = 91.3%	(172/183)*100 = 94.0%	(177/183)*100 = 96.7%
Fiction	(205/224)*100 = 91.5%	(203/224)*100 = 90.6%	(209/224)*100 = 93.3%
Scientific-Technical	(160/167)*100 = 95.8%	(158/167)*100 = 94.6%	(163/167)*100 = 97.6%

The comparison in Table 14 shows us that, for our samples, paradigmatic cues and morphological cues are approximately equally reliable. While it may seem that the complex model for morphological cues outperforms the use of paradigmatic cues alone, chi-square tests reveal no significant differences for comparison between accuracy of paradigmatic cues and morphological cues.[17]

In other words, paradigmatic cues are just as reliable as morphological cues for predicting the aspect of verbs. And remarkably enough, as we saw in Sec-

[17] Chi-square tests were performed on 3×2 tables of the raw numbers of correctly guessed verbs (numerator values in Table 14) for paradigmatic cues vs. each model for morphological cues across the three genres. The test statistic values for the comparison of paradigmatic cues with the simple model for morphological cues yields a chi-square value of 0.095193, with 2 degrees of freedom, and p-value = 0.95352. The values for the comparison of paradigmatic cues with the complex model for morphological cues plus interaction of motion verbs yields a chi-square value of 0.089895, with 2 degrees of freedom, and p-value = 0.95605. In other words, there is over a 95% chance that one would get this much difference between the two samples represented by the two sets of cues if there was no difference at all in their accuracy.

tion 7, it seems that the paradigmatic cues are quite accurate in separating out the groups of verbs for which morphological marking is unreliable, namely the perfective simplex verbs and prefixed indeterminate motion verbs. It seems likely that the two types of cues can be used to supplement each other, making it possible to achieve nearly 100% accuracy.

9. Conclusions

The answer to both our research questions,

- Can the aspect of individual verbs be determined purely on the basis of their grammatical profiles?

  and

- Does aspect in Russian interact with genre?

is mostly “Yes.” There is indeed a distributional bias in the use of verb forms for perfective vs. imperfective verbs, and it is robust not only at the aggregate level, but also at the level of the individual verb. If you have fifty or more tokens of a given verb, you can be about 93% certain that you know what aspect it is, based only on the relative frequencies of the verb’s forms. Although different genres are characterized by different vocabularies of verbs, and sometimes the same verb will behave differently in different genres, the reliability of paradigmatic cues to aspect is quite stable across genres. This fact shows that our results are replicable, and thus scientifically valid. The validation across independent data samples justifies our suggestion that our data can also be used as a proxy for the “genre” of child-directed speech. However, there are certainly differences between corpus distributions and the input to L1 acquisition, so more research would be needed to confirm whether grammatical profiles play a role in acquisition of Russian aspect.

Our data support the usage-based model of language, showing that the category of aspect is in principle learnable from the input. Paradigmatic distributions are a very good cue on their own, and their accuracy in predicting aspect is statistically indistinguishable from the accuracy of morphological cues. While the exact status of syntagmatic cues will require further research, it is conceivable that children use the relative frequencies of verb forms to supplement morphological and syntagmatic cues in sorting out which verbs are perfective vs. imperfective.

Our findings also reveal prototype effects in the categories of perfective vs. imperfective verbs. The grammatical profiles of some verbs are more strongly patterned as perfective or imperfective as opposed to others that tend to show less extreme aspectual distinction. In some cases, individual verbs can deviate strongly from overall patterns. Alignment of paradigmatic and morphological cues is clearly demonstrated by the distributions visualized in Figure 5. The two major classes of verbs with unambiguous aspectual morphology, namely the prefixed perfectives (PFperf) and secondary imperfectives (2Impv), also have their medians maximally far from the zero line, so that the prefixed perfectives tend also to have grammatical profiles most characteristic of perfective verbs, while secondary imperfectives have grammatical profiles most characteristic of imperfective verbs. These two classes of verbs clearly behave as prototypes for the two aspects. However, both of these classes also allow considerable freedom in distributions of grammatical forms for individual verbs (a wide range of progressively less prototypical members of each category), which is reasonable since they are clearly marked morphologically. Grammatical profiles are most unambiguous (non-overlapping) precisely for the two classes of verbs which lack morphological marking: the base verbs that are either perfective or imperfective. Here the paradigmatic cues play a particularly important role, making it possible to distinguish perfective from imperfective verbs in the absence of morphological cues.

REFERENCES

Aksu-Koç A. 1998. The role of input vs. universal predispositions in the emergence of tense–aspect morphology: Evidence from Turkish. *First Language* 18, 255–80.

Andrews, Edna, Galina N. Aver'janova, Galina I. Pjadusova. 2001. *The Russian Verb: Form & Function*. Moskva: Russkij jazyk.

Bartoň, Tomáš, Václav Cvrček, František Čermák, Tomáš Jelínek, Vladimír Petkevič. 2009. *Statistiky češtiny*. Praha: Nakladatelství lidové noviny.

Berdičevskis, Aleksandrs & Hanne M. Eckhoff. 2014. Verbal constructional profiles: reliability, distinction power and practical applications. In Henrich, Verena, Erhard Hinrichs; Daniel de Kok; Petya Osenova; Adam Przepiórkowski (ed.), Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT13), 2–13. Tübingen: University of Tübingen.

Bogojavlensky, Marianna. 1982. *Russian Review Grammar*. Columbus, OH: Slavica.

Borras, Frank M. & Reginald F. Christian. 1959. *Russian Syntax: Aspects of Modern Russian Syntax and Vocabulary*. Oxford: Clarendon Press.

Brown, Nicholas J. 1996. *The New Penguin Russian Course*. London: Penguin Books.

Choi, Sung-Ho. 1999. Semantics and Syntax of мочь and смочь: Their "Aspectual" Relationship. *Russian Linguistics* 23:1, 41–66.

Dąbrowska, Ewa. 2016. Cognitive Linguistics' seven deadly sins. *Cognitive Linguistics* 27:4, 479–91.

Dickey, Stephen M. 2007. A Prototype Account of the Development of Delimitative PO- in Russian. *Cognitive Paths into the Slavic Domain*. Edited by Dagmar Divjak and Agata Kochanska. Berlin: Mouton de Gruyter, 326–71.

———. 2011. The Varying Role of PO- in the Grammaticalization of Slavic Aspectual Systems: Sequences of Events, Delimitatives, and German Language Contact. *Journal of Slavic Linguistics* 19:2, 175–230.

Dostál, Antonín. 1954. *Studie o vidovém systému v staroslovenštině*. Praha: Státní pedagogické nakladatelství.

Eckhoff, Hanne M., Laura A. Janda. 2014. Grammatical Profiles and Aspect in Old Church Slavonic. *Transactions of the Philological Society* 112: 2, 231–58. DOI: 10.1111/1467-968X.12012.

Endresen, Anna, Laura A. Janda, Robert Reynolds and Francis M. Tyers. 2016. Who needs particles? A challenge to the classification of particles as a part of speech in Russian. *Russian Linguistics* 40: 2, 103–32. DOI 10.1007/s11185-016-9160-2.

Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalizations in Language*. Oxford: Oxford UP.

Greenacre, Michael. 2007. *Correspondence Analysis in Practice*. Hoboken, NJ: Taylor & Francis.

Janda, Laura A. 2007. Aspectual clusters of Russian verbs. *Studies in Language* 31:3, 607–48.

———. 2015. Cognitive Linguistics in the Year 2015. *Cognitive Semantics* 1, 131–54.

Janda, Laura A., Olga Lyashevskaya. 2011. Grammatical profiles and the interaction of the lexicon with aspect, tense and mood in Russian. *Cognitive Linguistics* 22:4, 719–63.

Kagan, Olga, Frank Miller, & Ganna Kudyma. 2006. *V Puti*. *Russian Grammar in Context*. *2nd ed*. Upper Saddle River, NJ: Prentice Hall.

Kamphuis, Jaap. 2016. *Verbal Aspect in Old Church Slavonic*. Doctoral Dissertation, University of Leiden.

Langacker, Ronald W. 2013. *Essentials of Cognitive Grammar*. Oxford: Oxford UP.

Lekic, Maria D., Dan E. Davidson, Kira Gor. 2008. *Russian Stage One: Live From Russia: Vol 1 & 2*. Dubuque, IA: Kendall Hunt.

Lubensky, Sophia, Gerard L. Ervin, Larry McLellan, Donald K. Jarvis, et al. 2002. *Nachalo*. New York: McGraw Hill.

Lyashevskaya, Olga N. and Serge A. Sharoff. 2009. *Častotnyj slovar' sovremennogo russkogo jazyka (na materiale Nacional'nogo korpusa russkogo jazyka)*. Moskva: Azbukovnik.

Murphy, Arthur B. 1965. *Aspectival Usage in Russian*. Oxford: Pergamon Press.

Nesset, Tore. 2013. The History of the Russian Semelfactive: The Development of a Radial Category. *Journal of Slavic Linguistics* 21.1, 123–69.

Offord, Derek. 2005. *Using Russian: A guide to contemporary usage*. Cambridge: Cambridge UP.

Padučeva, Elena V. 2008. Režim interpretacii kak kontekst, snimajuščij neodnoznačnost'. *Komp'juternaja lingvistika i intellektual'nye texnologii*. *Vyp*. *7 (14)*. *Po materialam meždunarodnoj konferencii "Dialog 2008"*, 412–19.

———. 2015. O biaspektual'nosti russkogo glagola BYT'. *Aspectual'naja zona: tipologija sistem i scenarii diaxroničeskogo razvitija*. *Sbornik statej V Meždunarodnoj konferencii Komissii po aspektologii Mežduranodnogo komiteta slavistov*, 176–84. Kyoto: University of Kyoto.

Pul'kina, Il'za M., Ekaterina B. Zaxava-Nekrasova. 1978. *Učebnik russkogo jazyka*, *6th ed*. Moskva: Russkij jazyk.

Rassudova, Ol'ga P. 1968. *Upotreblenie vidov glagola v russkom jazyke*. Moskva: Izdatel'stvo Moskovskogo universiteta.

Reynolds, Robert J. 2016. *Russian natural language processing for computer-assisted language learning*. Doctoral Dissertation, UiT The Arctic University of Norway.

Rifkin, Benjamin A. 1995. *Grammatika v kontekste: Russian Grammar in Literary Contexts*. New York: McGraw Hill.

Robin, Richard M., Karen Evans-Romaine & Galina Shatalina. 2013. *Golosa*. New York: Pearson Education.

Shirai, Y. & Andersen, R.W. 1995. The acquisition of tense–aspect morphology: A prototype account. *Language* 71, 743–62.

Stoll, Sabine. 2001. *The Acquisition of Russian Aspect*. Doctoral Dissertation, University of California, Berkeley.

Stoll, Sabine & Stefan Th. Gries. 2009. How to measure development in corpora: an association strength approach. *Journal of Child Language* 36(5), 1075–1090.

Tatevosov, Sergei. 2007. Intermediate prefixes in Russian. In Antonenko A., Baylin J., Bethin

C. (ed.) *Formal approaches to Slavic Lingustics*. *The Stony Brook Meeting 2007*. Ann Arbor: Michigan Slavic Publications, 423–42.

Timberlake, Alan. 2004. *A Reference Grammar of Russian*. Cambridge: Cambridge UP.

Tomasello, Michael. 1992. *First Verbs*. Cambridge: Cambridge UP.

Wade, Terence. 1992. *A Comprehensive Russian Grammar*. Oxford: Blackwell.

Wiemer, Björn & Ilja Seržants. Forthcoming. Diachrony and Typology of Slavic Aspect: What does morphology tell us? In: Andrej Malchukov & Walter Bisang (ed.), *Unity and diversity in grammaticalization scenarios*. *[Studies in Diversity Linguistics]*. Berlin: Language Science Press.

Zipf, G. 1949. *Human Behavior and the Principle of Least Effort*. New York: Addison-Wesley.

Аннотация

Ханне М. Экхофф
Лора А. Янда
Ольга Ляшевская

Предсказание глагольного вида

В статье рассматривается вопрос, можно ли предсказать вид отдельных глаголов исходя из статистического распределения их грамматических форм и влияют ли жанровые особенности текста на такое распределение. Исследуются так называемые «грамматические профили» (распределения относительной частоты глагольных словоформ) на материале трех выборок из Национального корпуса русского языка для следующих жанров: публицистика, художественная литература, и научно-техническая проза. Мы приходим к выводу, что вид отдельных глаголов может быть предсказан исходя лишь из распределения его форм с точностью 92.7%. Исследование показывает, что нет статистически значимой разницы между предсказанием вида на основе распределения словоформ и предсказанием вида исходя из видовых словообразовательных моделей. Это может свидетельствовать в пользу того, что при усвоении русского глагольного вида дети воспринимают тенденции в дистрибуции словоформ наряду со словообразовательными, семантическими и синтаксическими особенностями глагола.