A DATABASE OF RUSSIAN VERBAL ASPECT

OLGA BORIK & MAARTEN JANSSEN
Universitat Autònoma de Barcelona & Universitat Pompeu Fabra

[1] INTRODUCTION

When dealing with verbal aspect in Russian, it is not uncommon to find generic statements such as: all perfective verbs in Russian are morphologically complex, and all prefixed verbs in Russian are perfective (e.g., Ramchand 2004; Filip & Rothstein 2006). Statements of this type are so common that they are perceived as general rules about the aspectual system of Russian with a very insignificant class of exceptions (Forsyth 1970). Moreover, these statements often serve as the basis for theories of aspect, especially those which are syntactically oriented and treat prefixes as (roughly) secondary predicates (Svenonius 2004; Zaucer 2009). More complex generalizations often state the limitations on secondary imperfectivization in morphological terms, e.g. secondary imperfectives are not derived from verbs with prefixes contributing predictable/quantificational type of meanings, like, for instance, nepe- in the distributional meaning (Babko-Malaya 1999; Schoorlemmer 1995).

However, for such statements to be made with certainty, it is necessary to either verify each and every verb in Russian by hand, or to have a full database of verbal aspectual forms at your disposition, which to our knowledge does not exist for Russian. This paper describes a pilot project which aims to provide a morphologically oriented database of aspectual forms in Russian, which should be of great practical help for the numerous theoretical studies of Russian verbal aspect. Although the database itself will remain as theory independent as possible, we will show how this database can serve to either support or refute theoretical generalizations about Russian aspect.

It has been a while since the study of Russian aspect was a topic restricted to the domain of interests of traditional grammarians and Slavisists. Given a great theoretical importance of the notion of aspect for linguistic theory in general (cf. Verkuyl (1993); Krifka (1998); Smith (1991) and many others), researches of various theoretical backgrounds become intrigued by aspectual puzzles in different languages, including Russian. Clearly, not everybody who is interested in

[1] Although there have been databases which provide derivational information for Russian, e.g. Russian Derivational Morphology Database (http://courses.washington.edu/unimorph) which was pointed out to us by an anonymous reviewer, they do not focus on the aspectual information like our database does. Moreover, we provide an open online access to our database, whereas the Russian Derivational Morphology Database seems to be no longer available online.
Russian speaks it and/or has intuitions about it. However, it is indispensable for researchers to have access to the full range of data in the language they study to conduct independent investigations, and our database aims at providing this access for Russian data, at least when it comes to aspectual morphology.

The practical aim of the project, which forms part of the Open Source Lexical Information Network (OSLIN) for Russian, is to create a database of all Russian verbs. All prefixed verbs in the database will be linked to their base form. Independently of this, all the verbs will be classified as either perfective, imperfective or biaspectual, and the members of a perfective/imperfective opposition with the same derivational base are linked to each other. The linked members can further be classified into aspectual (perfective/imperfective) pairs. To keep the database as theory independent as possible, we establish the relation between the members of an aspectual pair for the largest part semi-automatically, on the basis of the information provided in their dictionary definitions.

With all the information above (i.e. aspectual class, morphological derivations, pairing) taken into account, the database can be used to group verbs in a way reminiscent of Janda’s 2007 ‘aspectual clusters’. A crucial difference with the aspectual clusters in the theory of Janda, however, is that the verb clusters generated from our database are based on an exhaustive lexicon, and created theory independently. The clusters are not meant to be an explanation by themselves, but to provide empirical grounds for data generalizations that, in turn, can be used to support theoretical explanations.

With respect to theoretical linguistics, the database is meant to render the means to conduct or facilitate the data search for research questions. That is to say, it can and should be used to provide data support for theoretical hypotheses, but does not itself concern with theoretical issues. For some theoretical questions, more information will be needed than the database can supply. Section [4] illustrates this: the database provides a list of 3a- verbs with the relevant morphological information (i.e. base verb for prefixed verbs, aspectual value, etc.) and is successfully used for the case study we present, even though this study still needs to be supplemented with additional information about the inchoative vs. resultative meaning of 3a- for each particular verb. However, even in those cases where additional information is required, the database already provides a large part of information needed by not only extracting the words starting with 3a-, but also stating whether or not those are prefixed words, what their aspectual class is, etc.

The remainder of the paper is organized as follows. The next section provides a detailed explanation of the database itself, including a thorough description of its contents, and a characterization of the website where the database can be consulted. In section [3] and [4], we illustrate how the database can be used to check various theoretical hypotheses and section [5] presents the conclusions.
[2] DATABASE AND WEB SITE

The Russian aspectual database described in this paper is part of the Open Source Lexical Information Network (Janssen 2005). OSLIN uses a database architecture with a modular set-up: the core of the database consists of a list of words of a given language, typically taken from a dictionary (see Section [2.1]). Any number of tables can be linked to this basic word list, each providing a specific type of information about the lexical units: their inflection, derived words, pronunciation, syllabic structure, etc. The database has an online interface, which allows easy access to all the data in the database. More information about the OSLIN project can be found on the website of the project: http://www.oslin.org.

The Russian database is only a pilot project, and the Russian OSLIN has not been developed in full. The focus of the pilot is Russian aspect, and therefore, all data that have no direct relevance for aspect have not been developed in a systematic way. Furthermore, the data that are relevant for aspect are themselves still under development: errors in the database are still being corrected. We will explain possible sources of errors in this section, not to emphasize the problems in the database creation process, but to warn the readers about more error-prone areas.

However, the percentage of errors is already low enough for the database to serve as a tool for linguistic purposes, though some care should be taken with the data at this point. The Russian data can be accessed via: http://ru.oslin.org, a screenshot is shown in figure 1.

Typically, the most important table linked to the wordlist in OSLIN is the table with inflectional information for each word in that language. However, since inflection is not of direct interest for the aspectual data, only a part of the word list is inflected at this time. For the current project there are three relevant tables, each providing a separate type of information.

The first table indicates the aspectual class of all the verbs in the word list: it simply specifies for each verb whether it is perfective, imperfective, or biaspec- tual (see section [2.2]).

The second table characterizes the morphological (de)composition of words, which currently contains primarily the compositional structure for prefixed verb. For instance, for the verb написать (to write.Perf), this table specifies that it is composed of the prefix на- and the base verbписать (to write, Imperf., see section [2.3]).

The third table lists aspectual pairs: it links perfective verbs to their imperfective counterparts. Furthermore, all the aspectual pairs contain information about the formation process by which the members of the pair were derived: prefixation, alternation, stress shift, secondary imperfectivization etc. (see section [2.4]).
When all the information presented in each of the tables is brought together, we get clusters of morphologically and aspectually related verbs (see section [2.5]). Morphological information does not have to coincide with the aspectual one: many morphologically related verbs are not related aspectually (i.e. do not constitute pairs). Let us now explain the database setup in more detail.

[2.1] **Basic Word List**

The basic word list in the database consists of the combination of the vocabulary of two major Russian dictionaries: the Толковый словарь русского языка (Explanatory dictionary of the Russian language) by Ozhegov & Shvedova (1992), and the Большой толковый словарь современного русского языка (Large explanatory dictionary of the Russian language) by Ushakov (1935-1940). Both dictionaries were taken from the available online sources, and these online version in turn were generated from scanned versions of the respective dictionaries. For this pilot project, no other words were included.

Creating a list of words on the basis of those dictionaries is by itself not a trivial task: the problem is that dictionaries often combine various words in a single entry. Consider, for instance, the entry for the word **муслить** (to (be)slobber) shown in figure 2, taken from Ozhegov dictionary, which not only lists the word itself,
but also a more common spelling variant (мусолить), several perfective counterparts (замусолить, намусолить, ...), and several reflexive forms (мусолиться, ...).

When we fully dissect this dictionary entry, we find that there are 10 individual words compressed into a single entry: замусли́ть, намусоли́ть, муслиться, замуслиться, замусоли́ть, мусоли́ть, замусоли́ть, намуслить, мусоли́ться, and муслить. In the OSLIN database, all of these words are listed separately. The total number of words included in the OSLIN database from these two dictionaries is 81,852, amongst which there are 28,766 verbs.

The online versions of both of the dictionaries have errors in the optical character recognition of the scanned text. For instance, the original database lists the word гризну́ть instead of the word гризнуть (‘to sink’) in the orginal text. Although we attempted to correct as many errors as possible, there are still some remaining errors in the word list due to these mistakes.

The inflections included in the database at this moment are taken from the STARLING database by Starostin: http://starling.rinet.ru/morpho.php. Since the starling database does not include all the words in the dictionaries we are working with, not all the words have been inflected. Furthermore, given that inflection was not a primary objective of this pilot study, we only included those inflections that could be derived automatically from the database. This means that, for instance, in the case of homographs with different inflectional paradigms the inflectional information is not always provided.

For a substantial amount of words, the syllabic structure of the word is also indicated. For instance for the verb переговаривать́сь (to talk to each other, exchange words), the syllabic structure is listed as пе́рего-ва-ри-вать-ся, with seven syllables and the stress on the forth one. The main motivation for including this information in this pilot project is that in many cases, homographs with a different aspectual behaviour can be distinguished by means of the position of the stress, as, for instance, perfective уре́зать and imperfective уре́зать (to cut
down/off). Only for cases where stress was significant has the syllabic structure been developed in detail, in other cases it was only added when the position of the stress could be derived from the starling database.

[2.2] **Grammatical Classification**

All verbs in the database have received a grammatical classification, that is to say, an indication whether the verb in question is perfective, imperfective, or biaspectual in nature. In principle, the information was adopted from the source dictionaries, either mentioned explicitly (сов. – perfective, несов. – imperfective, сов-несов. – biaspectual), or implicitly as in the case of однокр. (semelfactive), which is perfective by default, and многокр. (multiple action), which is always imperfective. In cases where the aspectual information was missing (mostly in the case of run-ons) or different in the two sources, we by default adopted the information from the starling database.

[2.3] **Morphological Composition**

The morphological composition table provides morphological structure for complex verbs in the database. This part of the database was originally intended as a list of prefixed verbs, hence the largest part of the database consists of prefixed verbs, however, several suffixed and infixed verbs are also marked. For each prefixed verb a prefix and a base verb are indicated. In total, 9,838 verbs are listed as being morphologically prefixed, or about \( \frac{1}{3} \) of the verbs.

The focus in this pilot project with respect to prefixed verbs was largely driven by the following theoretical question: what exactly is the role of prefixation in aspectual mechanisms? In other words, we are interested in looking into the relation between prefixation as a morphological process and the aspectual semantic properties of verbs. In practice, the relation between the two processes is far from being simple and straightforward. This issue is partially addressed in section [3].

The morphological composition is intended to model the synchronic status of words: whether the word does consist of composing parts or not. It is not meant to indicate the etymology of the verb, nor does it represent the full morphological analysis, but only the outermost compositional elements. For instance, for the verb разноситься (to be spread, distributed out), the database only models that it consists of the verb разносить (to distribute, carry around), together with the reflexive suffix -ся. The fact that разносить itself is a prefixed verb is modelled only for the verb разносить, and not for the verb разноситься.

The morphological composition was established semi-automatically. As a first step, a list of all productive prefixes in Russian was created. This list was then used to see which verbs in the database can be split into a productive prefix and an existing verb, that is to say, which words have the correct orthography to be prefixed verbs. As an example: since за- is a productive prefix, we automatically
extracted all and only those verbs that start with за- and have a listed base without за-.

This procedure yields all the prefixed verbs, but also a small percentage of false candidates. The false prefixed candidates were sorted out manually. As an example of a false candidate, both words зачесть (to counterbalance, take into account) and честь (honor) exist in the database, hence зачесть was automatically classified as a verb with a prefix за- derived from честь. However, the verb зачесть and the noun честь are not related, so the potential morphological relation was deleted during the manual check-up.

Many databases that model morphological composition mark the composition in terms of a string of letters. However, this leads to problems in cases of homographs: analysing prefixation in terms of a prefix and the citation form of the verb can lead to ambiguous and partially incorrect data. To take another case of verbs starting with за-: the verb за·пах·нуть (to lap) is not a prefixed verb, but rather a verb derived from the noun за·пах (a lap of a garment). Therefore, it is not a prefixed form of the verb пах·нуть (to smell). However, the verb за·пах·нуть (to begin to smell) both has a prefix and is derivationally related to пах·нуть (to smell). If morphological composition were modeled only as за- + пахнуть, it would be impossible to tell these cases apart. Therefore, the morphological composition is modelled over an identifier that points to a specific verb.

In this database, prefixes are always listed in their orthographic form. That is to say, the prefixes рас- and раз- are listed as separate prefixes, despite the fact that they are commonly considered allomorphs. The same applies to prefixes о-/об-, с-/со-, из-/ис-, etc. This is not intended as a claim that there is a difference between these prefixes, but merely to keep the morphological structure transparent.

**Marginal cases**

Let us now demonstrate some cases where the analysis had to be done manually. For most words, it is clear whether a candidate for morphological composition is indeed correct or not. However, there are quite a few verbs for which it is not really obvious whether they need to be classified as prefixed or not. This cases are also valuable from the theoretical perspective, since they indicate the most non-trivial areas in a morphological derivation process.

The first class of manually treated verbs are the reflexives. Although for many verbs, such as пробовать (to try out.refl), it is clear that it is not a prefixed version of бовать, but rather a true reflexive form of пробовать (to try out.Imp), there are other verbs for which this is not so clear. If we consider за·сы·пать·ся (to be filled up.Pf), it can be analysed both as a prefixed form of сы·пать·ся (to pour, fill up.Imp) and a reflexive form of за·сы·пать (to pour, fill up.Pf). In those cases, we always opted to consider -ся to be the outermost compositional element, unless there were clear counter-indications.
A similar issue arises with aspectual pairs. The verbs разрезать (to cut.Imp) and разрезывать (to cut.Imp) are secondary imperfectives to their perfective counterpart разрезать (to cut.Pf), which in turn is a prefixed verb, composed of раз+резать. Thus, although both secondary imperfectives are related to a prefixed verb, they themselves are not. In the case of разрезать, the stress is a good indication: the verb *резать does not exist. But in cases in which the imperfective verb without prefix itself also exists, morphological composition becomes a complicated procedure. Consider, for instance, the verb до-делывать (to finish making). Should it be analyzed as a secondary imperfective of до-делать (to finish making.Pf), or a prefixed version of делывать (make.Imp, multiple action)? In those cases, for practical purposes we always linked the secondary imperfective to a prefixed perfective verb without explicitly indicating the prefix of a secondary imperfective. This was largely done to keep a parallel between cases like разрезывать and доделывать.

Problems also arise when the root verb itself has a marginal status. Consider the verb взимать (to collect, levy). Even though Ushakov dictionary lists its base verb, имать, as a separate entry, this verb is barely used in the modern language, which makes the morphological composition of взимать as вз+имать questionable.

Another good example of a marginal case is the verb замыкать, which can be perfective and imperfective. As a perfective entry, за-мы-кать means to make someone suffer and is most probably related to the imperfective entry мы-кать (to endure), which in modern Russian has a highly restricted use, as, for instance, in the expression горе мыкать (to bear/undergo grief, hardship) or in dialects in the meaning ‘to hatch’. Thus, if we set apart a dialectal use, there seems to be a morphological relation of prefixation between the two entries and it is reasonable for this case to give a complex morphological structure for the verb за-мы-кать (pf) involving the prefix за-. However, there is a homographic imperfective entry, за-мы-кать (to close, round out, shut), which is paired with the perfective замкнуть and does not seem to show any sign of a complex morphological structure involving a prefix за-. The problem is solved by the existence of two separate entries, за-мы-кать and за-мы-кать, and manually controlling that all the morphological and aspectual relations are linked to the correct entry.

[2.4] Aspectual Pairs

Pure aspectual pairs are those verbs which have an imperfective form and a perfective counterpart, and for which the perfective form expresses nothing more than the imperfective form made perfective. For instance, in the case of делать: сделать (to make.Imp:Pf) the perfective сделать is nothing but the action of делать brought to its logical conclusion. In such cases, the perfective verb is almost an inflected form of the imperfective verb, where the choice of which form to use
is determined by the context, much more than by the meaning the speaker wishes to convey. The database of aspectual pairs in meant to include such pairs of verbs.

However, there are many examples of imperfective/perfective pairs for which the difference between the two verbs also carries a semantic load, and is not merely aspectual. For instance, the verb валить (to bring down; fell, lay, low) has various morphological derivations such as завалить (to pile up, overwhelm), привалить (to lean against), подвалить (to pile more, to come up to someone (sl.)), развалить (to destroy), etc, which all mean different things. In these cases, although the perfective verb expresses some perfective action related to the imperfective verb, it is more than just the aspectual counterpart: they are different verbs that are most often related, but carry an additional semantic load. Also, there are many verbs for which there is more than one perfective form linked to a given imperfective verb as a pair. For instance, the same verb валить has two perfective counterparts listed as a pair in Ushakov dictionary: свалить (to dump, fell) and повалить (to lay, fell), depending on the meaning of the imperfective verb.

The database of aspectual pairs is meant to include all those and only those pairs of imperfective and perfective verbs that form an aspectual pair, and for this pilot project only those that form an aspectual pair according to the dictionary. In other words, we use the dictionaries as a source to decide whether or not two verbs should be listed as a pair, and with very few exceptions only list those pairs that are attested in either of the two dictionaries. For instance, the perfective verb свьежиться (to cringe, roll up.Pf) is linked to свьеживаться (to cringe, roll up.Imp) by Ushakov, but to ежиться (to cringe.Imp) in Ozhegov, hence, both are listed as aspectual pairs in the database. On the other hand, all dictionaryed pairs are included, except in cases where there was a clear error. Amongst other things, this means that in many cases, reflexive pairs are included alongside their non-reflexive counterparts: not only the pair ягнить/оягнить (to lamb), but also the pair ягниться/оягниться is listed in the database of aspectual pairs.

The total number of aspectual pairs is 11.455. Since some verbs have more than one pair, the number of imperfective verbs that has at least one aspectual pair is lower: 8.644. Given that there are just over 14.500 imperfective verbs in the database, this means that 59% of the listed imperfective verbs have an aspectual pair according to the dictionary. Perfective verbs have a pair even more often: 8.714 of the 12.810 perfective verbs (or 68%) has an aspectual pair.

The aspectual pairing is not unrelated to the morphological database: many of the prefixed verbs form an aspectual pair with their morphological base. The pairs in the aspectual database are selected for their functional relation independently of morphological structure of the members of this relation. This means that the relation between an imperfective verb and its perfective prefixed pair,
should it exist, is essentially the same as that between a perfective verb and its secondary imperfective pair. In other words, the aspectual pairing table does not make a distinction between the prefixed pairs and those made by secondary imperfectivization, focusing not so much on the morphological structure of the paired members, as on the functional relation between them.

Although functionally secondary imperfectives are no different from “simple” imperfectives, they are different from a morphological point of view: “simple” imperfectives are the morphological base of their perfective pair, whereas the secondary imperfectives are the derivates. Therefore, the database explicitly keeps track of the way in which the perfective was morphologically formed from the imperfective, or vice-versa.

The aspectual pairing table, especially in combination with the morphological decomposition table provides a rich set of data for the study of verbal aspect in Russian, and helps answering many theoretically relevant questions. We will get back to this issue in sections [3] and [4].

[2.5] Verb Clusters

On the website, the information from the tables described in the previous three subsections is combined and presented as a single resource, which provides all the relevant aspectual information about all the verbs in the database. For each verb, the site provides a page which lists all verbs that are morphologically or aspectually related to it.

Furthermore, it provides the option to automatically generate so-called verbal clusters. A verbal cluster is defined by a verb together with all verbs linked to it by either aspectual pairing or morphological derivation or both, plus the verbs related to those verbs. Such verbal cluster can become rather large and complex. As an example, consider the (imperfective) verb брать (to take), which has a perfective form взять, formed by suppletion. However, it also has several prefixed forms with their own lexicalized meaning, such as забрать (to collect), убрать (to clean, put away), and прибрать (to tidy), избрать (to elect), etc. As a lexicalized perfective, избрать has a secondary imperfective form: избирать. But избрать, in turn, can be further prefixed: переизбирать (to reelect), which has imperfective form переизбирать. In total, there are 47 verbs in the брать cluster, as demonstrated in figure 3. The image in figure 3 is a screenshot of a page of the database website (http://ru.oslin.org/?action=aspect&act=explode&id=3544).

As has already been mentioned, the aspectual database is meant as a convenient and necessary tool which can either support or refute various theoretical generalizations and hypothesis concerning Russian aspect. The next two sections provide some examples of empirical studies that can be carried out using the database. First we will demonstrate how various ‘common wisdom rules’ concerning the morphology of aspect can be empirically checked. The second section
**Verbal Cluster**

Cluster base: брать

<table>
<thead>
<tr>
<th>lookup verb</th>
<th>брать</th>
<th>(imperfective)</th>
</tr>
</thead>
<tbody>
<tr>
<td>perfective of брать</td>
<td>взять</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>забрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>вобрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>убрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>собрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>разобрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>пробрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>прибрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>подобрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>подобрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>перебрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>отобрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>обрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>обобрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>набрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>изобрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of брать</td>
<td>выбрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>imperfective of забрать</td>
<td>забирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>imperfective of вобрать</td>
<td>вбирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>imperfective of убрать</td>
<td>убирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>derivation of убрать</td>
<td>изубрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of убрать</td>
<td>приубрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of убрать</td>
<td>разубрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>imperfective of собрать</td>
<td>сбирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>derivation of собрать</td>
<td>собирать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>imperfective of разобрать</td>
<td>разбирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>imperfective of пробрать</td>
<td>пробирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>imperfective of приобрать</td>
<td>прибирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>imperfective of подобрать</td>
<td>подбирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>imperfective of перебрать</td>
<td>перебирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>imperfective of отобрать</td>
<td>отбирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>imperfective of обрать</td>
<td>обирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>derivation of набрать</td>
<td>нанабрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>imperfective of изобрать</td>
<td>изобрать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>derivation of изобрать</td>
<td>перезобирать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>imperfective of выбирать</td>
<td>выбирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>derivation of выбирать</td>
<td>переводить</td>
<td>(perfective)</td>
</tr>
<tr>
<td>imperfective of убирать</td>
<td>разбирать</td>
<td>(imperfective)</td>
</tr>
<tr>
<td>derivation of сбирать</td>
<td>насобирать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of собирать</td>
<td>насобирать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of собирать</td>
<td>присобрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of набирать</td>
<td>понабрать</td>
<td>(perfective)</td>
</tr>
<tr>
<td>derivation of избирать</td>
<td>переизбирать</td>
<td>(imperfective)</td>
</tr>
</tbody>
</table>

47 verbs in cluster

**FIGURE 3: The verb cluster for the word valjat’**
is dedicated to a particular case study that has been conducted on the basis of collected data. We look at the behaviour of prefix 3a- which, as is well-known, has (at least) two meanings: an ingressive one (i.e. to begin to do something), and a resultative one, which (loosely) refers to a complete action meaning. A natural question that arises is whether there is any rule(s) governing the distribution of the meanings of the prefix, depending on the type of verb the prefix is attached to. We present the data to discuss one possible factor that might play a role in the distribution of ingressive/resultative 3a-: the presence/absence of a direct object.


The database provides a wide range of information that can be directly used to verify tendencies and claims about Russian by statistical means. We demonstrate that in this section by verifying some claims that have been made repetitively in the literature. Even though the statistical data that we provide in this paper are not completely fixed yet due to the possible mistakes in the database that have not been eliminated, they nevertheless give a very good estimate of the general tendencies.

[3.1] Prefixes and perfectivization

It is generally assumed that there is a very tight connection between perfectivization and prefixation in Russian. Various authors differ in their precise definition of the nature of this connection, but most of them agree on the following basic fact: the output of the morphological operation of prefixation is aspectually perfective. In the formulation of Forsyth (1970, 18), “the addition of a prefix to a simple verb makes it perfective (in addition to possibly altering its meaning in other ways)...”. Taking this view to the extreme, some researches have proposed to treat prefixes as perfectivizers and nothing else (Pinón 1994; Zucchi 1999; Borer 2005). The traditional view on Russian prefixation, however, is that prefixes are not aspectual perfectivizers, but derivational morphemes with their own function (Isachenko 1960).

The question of whether prefixes perform other functions apart from forming perfective verbs falls outside the scope of this paper. However, what we can verify is a part of the claim by Forsyth, namely, that the addition of a prefix to a verb always makes it perfective. Interestingly, Forsyth himself, after giving this strict rule about prefixation, starts listing counterexamples to it. However, these counterexamples have been ignored in much of the subsequent literature on aspect, and many authors assume the perfectivizing effect of prefixes to be an absolute rule. In our database, there are 9,838 verbs marked as formed by prefixation. If prefixes always had perfectivizing effect, one would expect that all prefixed

---

[2] By using the term ‘resultative’, we do not assume that any completed action necessarily lead to a certain resultant state. We use the term in the most theoretically neutral way.
verbs are perfective. Ideally, prefixes would also attach to an imperfective stem, to have an obvious aspectual effect. However, the reality is rather more complex: although the majority of verbs does indeed follow this pattern, about 8% of the prefixed verbs are in fact formed on the basis of a perfective or biaspectual verb, and about 10% of the prefixed verbs remain imperfective themselves. A complete breakdown of these numbers is shown in Table 1.

We can use the data in table 1 to verify the claim that all verbs become perfective after prefixation. For the majority of prefixed verbs (about 80%), it is indeed true that they are perfective verbs created from imperfective bases. However, there is a substantial amount of verbs (about 10%) where the resulting verb is still imperfective. To give some examples of these:

выявляться = вы + являться
разобраться = разобрать + и
забегать = за + бегать
забирать = за + брать
соотносить = со + относить
справляться = спрашивать + и
как бы = как бы

Concerning the statement that all perfective verbs are prefixed, the database also shows that for most verbs, that is indeed true: of the just over 12,500 perfective verbs in our database, about 8,700 (or about 70%) are prefixed. However, that still leaves about 30% of the perfective verbs that are not in fact prefixed. This is hardly in line with the common view that there is only a small and insignificant set of exceptions to the rule that ‘all perfective verbs are prefixed’. The non-prefixed verbs are not a homogeneous class and are formed by different means, among

<table>
<thead>
<tr>
<th>Aspect of base verb</th>
<th>Aspect of prefixed verb</th>
<th># examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>biaspectual</td>
<td>biaspectual</td>
<td>7</td>
</tr>
<tr>
<td>biaspectual</td>
<td>perfective</td>
<td>88</td>
</tr>
<tr>
<td>perfective</td>
<td>perfective</td>
<td>783</td>
</tr>
<tr>
<td>imperfective</td>
<td>biaspectual</td>
<td>16</td>
</tr>
<tr>
<td><strong>imperfective</strong></td>
<td><strong>perfective</strong></td>
<td><strong>7942</strong></td>
</tr>
<tr>
<td><strong>imperfective</strong></td>
<td><strong>imperfective</strong></td>
<td><strong>1000</strong></td>
</tr>
<tr>
<td>perfective</td>
<td>biaspectual</td>
<td>2</td>
</tr>
</tbody>
</table>

**Table 1:** Distribution of prefixed verbs by aspectual classes
which a well-known suffix -ну́, but also cases of suppletion, various types of alternation and so on. Each of these classes might be small and maybe insignificant, but taken together, they present a serious class of counter-examples to the claim that all perfective verbs are prefixed.

While the larger types of pairs (rows) in table 1 on the preceding page should be self explanatory, the more marginal cases deserve some exemplification. There are six examples of biaspectual verbs derived on the basis of another biaspectual verb. To give an example, consider:

\[
\text{сонаследовать} = \text{со} + \text{наследовать}
\]

to co-inherit pref + inherit

The row with the smallest number of examples in table 1 on the previous page is for biaspectual verbs which are derived from perfectives:

\[
\text{переосвидетельствовать} = \text{пере} + \text{освидетельствовать}
\]

to re-examine pref + witness.Pf

\[
\text{исповедать} = \text{ис} + \text{поведать}
\]

to receive confession pref + narrate.Pf

The first example is a relatively clear case of a biaspectual/perfective pair. However, the second example might not really be an example of a biaspectual/perfective pair, since there is disagreement between various dictionaries concerning the aspectual marking: исповедать is classified as biaspectual by Ushakov, but as perfective by Ozhegov. More remarks on biaspectual verbs will be made in the next subsection.

### [3.2] Biaspectual verbs

As the last generalizing claim, let us consider biaspectual verbs. Biaspectuality is often presented as a marginal phenomenon which applies to a small class of borrowed verbs only. Since they are seen as a marginal phenomenon, very little have been said explicitly about this class of verbs. Let us point out that in the database, 956 of all verbs are classified as biaspectual. This is 3.5% of the total amount of verbs, which is not an overwhelming number, but not all that little either.

Biaspectual verbs present a difficult case at least for the statement that any verb in Russian is obligatorily classified as perfective or imperfective (Zaliznjak & Shmelev 2000). In the case of biaspectual verbs, there is a single lexical entry which can be used either with a grammatical value perfective or an imperfective

---

[3] Note, again, that not all -ну́ verbs are perfective either. Cf: пахнуть (to smell), бухнуть (to expand, swell), etc.
(Isachenko 1960), meaning that in context, all uses are either perfective or imperfective, but the verb itself has no (single) aspectual value. Unless we want to say that the aspectual class is not a property of a verb, but of a verb in use (which would immediately lead to problems in describing the effect of prefixation in relation to aspect), we still have to assigns such verb an aspectual class, and in our database they are simply classified as biaspectral. One possibility is that biaspectral verbs are in fact homonymous aspectual pairs: that is to say, that a biaspectral verb is in fact a pair of verbs, one imperfective, the other its perfective pair, and in use, of course, one of the two is used. For a complete treatment of the effect of prefixation, the exact status of biaspectral verbs is important. However, a full discussion of it is beyond the scope of this paper.

Concerning the claim that the majority of biaspectral verbs are borrowed (e.g. Zaliznjak 1977), it might be true for many cases, but, nevertheless, there are entries like велеть (to order, tell), бежать (to run, escape), даровать (to grant, bestow) that can hardly be classified as borrowings and still are biaspectral verbs.

As we have hopefully illustrated in this section, a lot of common wisdom statements concerning aspectual behaviour of verbs in Russian apply, at best, to a majority of verbs. However, the classes of exceptions do not amount to dozens, but rather to hundreds and thousands verbs and hence cannot simply be ignored, if the validity of theoretical observations is to be taken seriously.

[4] A CASE STUDY: PREFIX ЗА-

To illustrate possible practical applications of the database for various theoretical problems, we have conducted a case study on the basis of the verbs with prefix за-, one of the most studied prefixes in the literature (Boguslavski 1960; Braginsky 2008; Dickey 2000; Ferm 1990; Golovin 1964; Isachenko 1960; Janda 1986; Schooneveld 1978; Zemskaja 1955; Paillard 1991; Zaliznjak 1977, and many others).

We here first make a statistical analysis of some of the general claims about this prefix, directly using the data from the database. We then go on to a more detailed analysis of the formation process and meaning contribution of this prefix. The data necessary for a more detailed analysis are not all present in the database; for instance, the database says nothing about whether verbs are ingressive or not. The classification of meanings (ingressive vs. resultatives) was done manually. However, the database does directly provide the complete set of, for instance, all verbs that are prefixed by за-, together with their aspectual class, pairs, and derivate. This makes the type of analysis provided in [4.2] a lot faster and more reliable.

[4.1] Statistical data

Let us first look at some of the most common statement found in the literature that can be verified by means of the database. The first one concerns the per-
fectivity of prefixed verbs, the second one is about secondary imperfectivization. Verifying these claims is similar to the analysis in [3.1], but then specific for за-verbs.

The first statement has been most recently re-formulated in (Braginsky 2008, i), who says: “While it is correct that all ZA-prefixed verbs in the lexicon are perfective, the prefix ZA- has additional semantic effects on the unprefixed verbs it applies to.” Note that this statement is just an instantiation of a more common claim that all prefixed verbs are perfective, which does not really hold, as was discussed in the previous section. However, it could be the case that although in general, prefixes are not always perfectivizers, the prefix за- really does turn all verbs into perfectives.

If we look at the за- verbs in the database, we find that Braginsky’s claim holds better than the general claim about prefixed verbs, but still not entirely. There are 1019 verbs in our database with a prefix за- (as the outermost morphological element), and of those, 946 (or 93%) are perfective. This is a large percentage of the за- verbs, but nevertheless, about 7% of them are imperfective.

The group of exceptions consists mostly of prefixed motion verbs. Thus, the verb забегать (за+бегать) has two meanings: a) to start to run (around) and b) to pop by, and in the second meaning the verb remains imperfective, just like its base verb бегать (to run).

However, there is a much larger class of imperfective verbs with a за- prefixed which do not constitute a part of the 7% mentioned above. This is because in the 1019 prefixed verbs selection, we did not take into account secondary imperfective verbs formed on the bases of prefixed perfectives, such as, for instance, заделывать (to patch, seal), which has the following morphological derivational ‘history’:

<table>
<thead>
<tr>
<th>Imp.</th>
<th>Pf</th>
<th>Imp</th>
</tr>
</thead>
<tbody>
<tr>
<td>делать</td>
<td>➔</td>
<td>заделать</td>
</tr>
<tr>
<td>to make</td>
<td>to patch, seal</td>
<td>to patch, seal</td>
</tr>
</tbody>
</table>

The morphological structure of the last verb in the string, заделывать still clearly contains prefix за-, although the prefix here is not the outermost morphological element. While it is true in this case, that more aspectually relevant morphology is present which makes the verb imperfective, it still remains a verb that has a prefix за-, and its aspectual value is imperfective. For that reason, although almost all verbs do become perfective after adding a за- prefix, not all verbs that contain a за- prefix are perfective.

The second common statement about за-verbs has more broadly to do with the process of secondary imperfectivization. It is often claimed that secondary imperfectives do not form from perfective verbs with predictable prefix meanings (Isachenko 1960; Schoorlemmer 1995; Babko-Malaya 1999) or, in more recent
terminology, that perfective verbs with ‘superlexical’ prefixes do not have secondary imperfectives (Di Sciullo & Slabakova 2005; Romanova 2006). In the case of за-, the superlexical ‘version’ of it would contribute a ‘phasis’, or ingressive meaning. Hence, the statement with respect to the superlexical за- would be that we should not find secondary imperfective verbs with an ingressive meaning.

The database search shows that, indeed, this prediction is largely borne out, even though a list of exceptions to this rule also emerges. However, these exceptions are small in number, and only comprise the following 8 verbs:

- заболевать (from pf. заболевать) to become sick
- загнаиваться (from pf. загноиться) to begin to fester
- заговаривать (from pf. заговорить) to begin to talk
- загораться (from pf. загореться) to begin to burn
- замолкать (from pf. замолчать) to become silent
- запевать (from pf. запеть) to begin to sing
- зардевать (from pf. зардеть) to redden
- зацветать (from pf. зацвести) to begin to blossom

Thus, we can conclude that the database does show that the general claims for verbs with a за- prefix, namely, the statement that they are always perfective, and that they have no secondary imperfectives for ingressive meaning, do hold almost perfectly, with only a small number of exceptions.

[4.2] Detailed analysis
In this case study, we have focused on two meanings of за-: ingressive and resultative, exemplified below, with an indication of the total number of verbs with that type of meaning:

**Ingressive meaning:** петь ⇒ запеть to sing to start.singing 410 verbs

**Resultative meaning:** резать ⇒ зарезать to cut to stab (to death) 395 verbs

As can be seen, the distribution of these two types of meaning is about 50/50. Although these two meanings are not the only meaning of за- verbs, together they do account for most of verbs with a за- prefix. In this subsection, we try to determine whether factors as transitivity play any role at the distribution of the meanings of за- among the prefixed verbs.

It has often been suggested that the resulting meaning of за- can be predicted on the basis of the (in)transitivity of the base verb, i.e. the presence or absence of a direct object. One of the proposals to link ingressive за- verbs to transitivity was put forward in Isachenko (1960), who claimed that all the ingressive за- verbs are
intransitive. A similar claim is made by Schoorlemmer (1995), ad Babko-Malaya (1999), who claim that the resultative (lexical) meaning of за- arises only in the case when a prefix attaches to a verb with a direct object, and the ingressive за-verbs are formed on the basis of unergative unprefixed bases. A similar relation between direct objects and the resultative reading is either explicitly made or is expected in many syntactic approaches (Svenonius 2004; Ramchand 2004; Borer 2005). In this subsection, we will examine whether (in)transitivity is indeed a necessary and/or sufficient condition for an ingressive/resultative meaning.

First, let us examine the relation between the resultative meaning of за- and the presence of a direct object in the argument structure of the base verb. If the direct object is a necessary condition for the emergence of the resultative meaning of prefixed за- verbs, we expect that no resultative за- verbs can be formed from unergatives. However, this statement is not supported by the data. There are quite a few resultative за- verbs that have an unergative base, to name just a few:

<table>
<thead>
<tr>
<th>Unergative</th>
<th>Prefixation</th>
</tr>
</thead>
<tbody>
<tr>
<td>брызгать</td>
<td>забрызгать</td>
</tr>
<tr>
<td>мусорить</td>
<td>замусорить</td>
</tr>
<tr>
<td>ночевать</td>
<td>заночевать</td>
</tr>
<tr>
<td>прыгнуть</td>
<td>запрыгнуть</td>
</tr>
</tbody>
</table>

All the verbs on the left side are unergative. The prefixed verbs on the right side, all have a resultative, and not an ingressive meaning. Thus, замусорить really means “make a place dirty by littering” and not “to begin littering”, as would be expected by the general rule.

If we look at the hypothesised generalization the other way around, i.e. if we expect the direct object to be a sufficient condition for resultative за-, then we should not be able to find any transitive verbs which have an ingressive meaning after за- has been added. This, again, is incorrect and we find numerous examples of ingressive за- verbs formed from the transitive bases:
Unaccusative verbs, i.e. the verbs whose only argument is an underlying direct object (Perlmutter 1978), also regularly produce ingressive за- derivatives, as shown below:

<table>
<thead>
<tr>
<th>verb (part.)</th>
<th>за-verb (part.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ворошить (to stir up, rake)</td>
<td>заворошить (to begin to stir up)</td>
</tr>
<tr>
<td>курить (to smoke, optionally)</td>
<td>закурить (to begin to smoke)</td>
</tr>
<tr>
<td>торопить (to hurry)</td>
<td>заторопить (to begin to hurry)</td>
</tr>
<tr>
<td>шатать (to shake)</td>
<td>зашатать (to begin to shake)</td>
</tr>
<tr>
<td>шептать (to whisper)</td>
<td>зашептать (to begin to whisper)</td>
</tr>
<tr>
<td>затошнить (to nauseate)</td>
<td>заразошнуть (to become nauseous)</td>
</tr>
<tr>
<td>затлеть (to smolder)</td>
<td>заразыпнуть (to begin to smolder)</td>
</tr>
<tr>
<td>запахнуть (to smell)</td>
<td>заапахнуть (to begin to smell)</td>
</tr>
</tbody>
</table>

Thus, this selection of data shows convincingly that no direct relation can be maintained between the syntactic direct object and the emergence of the resultative meaning of за-. That is to say, although typically resultative за- verbs have a transitive base verb, it is neither true that all transitive verbs less to a resultative за- verb, nor that all resultative за- verbs have a transitive basis.

Let us know verify if, on the other hand, there is a strict relation between the ingressive за- verbs and the absence of a direct object. Once again, we need to see if an absent direct object is both a necessary and a sufficient condition for the emergence of the ingressive за-, otherwise the generalization is either wrong, or incomplete at best.

Should the absence of a direct object be a necessary condition for the ingressive за-, we would expect to find no transitive or unaccusative verbs to form a basis for the ingressive за- prefixed verb. However, as we saw in the examples above, the cases of transitive verbs with ingressive за- prefix are many, including торопить, ворошить, пахнуть, etc. Here are some more examples:
As the reader might have guessed by now, the absence of a direct object cannot be held to be a sufficient condition for ingressive за- derivations either. If this were the case, we would not find any prefixed за- verbs with resultative meaning formed on the basis of unergative verbs. To add to the examples given above (i.e. брызгать, мусорить, etc.), consider the following pairs:

- воевать (to be at war, to fight.Imp) ⇒ завоевать (to conquer.Pf.)
- гадать (to tell.fortune.Imp) ⇒ загадать (to make.wish.Pf)
- служить (to serve.Imp) ⇒ заслужить (to deserve.Pf)

Thus, we can safely conclude this section with the statement that none of the possible correlations between the presence/absence of a direct object in the base verb and the meaning of prefix за- in the derived verb really hold. We cannot predict the meaning of a prefixed verb on the basis of transitivity or, more broadly, the syntactic property of having a direct object in the base verb.

This outcome urges us to explore other hypotheses. There can be many potential proposals to investigate, one of the latest trends being the relation between the semantic type of a verb (in the sense of Vendler 1967) and the resulting за- derivation (e.g. Braginsky 2008). This is not a simple relation to discuss, and even though we believe that the database we are describing here can help solving a lot of questions, a detailed discussion of the semantic issues will take too much space and, hence, falls outside the scope of this paper.

[5] CONCLUSION

As we have shown in this paper, a full lexical database of Russian aspect is a valuable, if not a crucially necessary resource for verifying general claims about the behaviour of aspect in Russian. The verbal aspect database now exists, even though as a pilot project, and is available online. In the course of time, incidental errors will be corrected, and more types of information will be incorporated into

---

BORIK & JANSSEN

петь ⇒ запеть
(to sing.Imp, optionally tr.) (to begin to sing.Pf, optionally tr.)
знобить ⇒ зазнобить
(to shiver.Imp, unacc.) (to begin to shiver.Pf, unacc.)
ругать ⇒ заругать
(to scold, reprimand.Imp, tr.) (to begin to scold.Pf, tr.)
the database. We have shown that several claims which have often been assumed to be universal and undisputed for Russian verbal aspect, in fact, do not concur with the reality of the Russian lexicon. To sum up the major claims defended in this paper:

1. Despite the fact that it is often said that all (or at least the majority) of perfective verbs are prefixed, there are many perfectives in Russian which are neither formed by prefixation, nor have a prefix in their morphological structure.

2. Contrary to a common assumption, there is a rather large class of biaspectual verbs that are used both perfectively and imperfectively. The class does not consist of borrowed verbs only.

3. There is no connection between the ingressive or resultative meaning of за- prefixed verbs, on the one hand, and the presence or absence of a direct object in the argument structure of the base verb, on the other.

We have furthermore shown that neither the aspectual classification of verbs, nor the decision whether verbs are in fact prefixed is not, or at least not in all cases, a straightforward matter. This complicates a precise analysis of the relation between prefixation and perfectivity.

Since the database is not yet error free, the data and the statistics presented in this paper have a small margin of error. Those mistakes which might have undermined some of our conclusions were eliminated. For instance, there seemed to be a handful of perfective verbs that become imperfective after prefixation. The list was checked manually and it turned out that all the supposed instances were, in fact, due to errors in the Ushakov dictionary. Because of these issues, in this paper we included only those cases where there is a substantial class of exceptions to a rule, and where the margin of error is very small. Thus, the percentages presented here will not be seriously affected by occasional errors in the database.

As a final note, we want to point out that the Russian database project is at this point a non-funded pilot project. We appreciate any help or feedback in the further development and improvement of the project. We hope that the the OSLIN database in general, and the database of Russian verbal aspect in particular, will prove a useful resource for data-driven linguistic research.

REFERENCES


AUTHOR CONTACT INFORMATION

Olga Borik
Departament de Filologia Catalana
Universitat Autònoma de Barcelona
08193 Bellaterra (Barcelona)
Spain
Olga.Borik@uab.cat

Maarten Janssen
IULA, Universitat Pompeu Fabre
Roc Boronat, 138
08018 Barcelona
Spain
Maarten.Janssen@upf.edu