

## **Constructing a learner-friendly corpus-based dictionary of Serbian verbal aspect**

Tanja Samardžić and Maja Miličević

(University of Belgrade)

*Primenjena lingvistika* 14 (2013). 77-89.

This is a pre-print version. The paper is under copyright; for permission to re-use or reprint the material in any form please contact the editors (<http://dpls-alas.rs/primenjenaling.html>)

## CONSTRUCTING A LEARNER-FRIENDLY CORPUS-BASED DICTIONARY OF SERBIAN VERBAL ASPECT<sup>†</sup>

*This paper describes ongoing work on the construction of a dictionary of Serbian verbal aspect based on data from corpora and specially adapted for the needs of L2 learners, who face the difficult tasks of choosing the appropriate aspect and deriving the correct morphological form of the verb. We explain the process of automatic extraction of information about the aspectual behaviour of verbs from a corpus. We show that a quantitative notion of productivity of verb forms based on the extracted corpus data is useful for a systematic and rich representation of the knowledge about Serbian verb aspect.*

### 1. Introduction

Aspectual properties of verbs consist of a set of semantic features that describe the dynamics and the duration of the eventuality described by the verb. Although aspectual features can be rather diverse and language specific, the notion of *temporal boundedness* seems to play a role in many distinctions made across languages. This notion is particularly important when it comes to the aspectual characterisation of Serbian verbs, as illustrated by the English examples in (1a) and (2a) and their Serbian translations (1b) and (2b).

- (1) a. On 29th November 1984 at 7:55, Winston *turned* the switch and the whole building disappeared.  
b. 29. novembra 1984. u 7:55, Vinston je *okrenuo* prekidač i cela zgrada je nestala.  
c. \*29. novembra 1984. u 7:55, Vinston je *okretao* prekidač i cela zgrada je nestala.
- (2) a. On 29th November 1984 at 7:55, Winston *turned* the switch for a while.  
b. 29. novembra 1984. u 7:55, Vinston je neko vreme *okretao* prekidač.  
c. \*29. novembra 1984. u 7:55, Vinston je neko vreme *okrenuo* prekidač.

The sentences in (1a-b) describe an eventuality that is temporally bounded. This means that there is, *implicit to the meaning of the verb*, a point in time when the action described by the verb ends. Temporal boundedness also implies that there is a result state which is a consequence of the action described by the verb; in (1a-b), the result state is the state of the switch that causes the building to disappear. Contrary to this, the sentences in (2a-b) describe an eventuality that is temporally unbounded. Even though the time span over which the sentences in (2a-b) are true is defined by temporal adverbials (*for a while* and *neko vreme*, respectively), the verb itself does not imply an end point or a result state. In other words, based on (2a-b), we do not know whether the turning of the switch is completed or what the final position of the switch is.

The distinction between (1) and (2) is based on fine semantic characterisation of the eventualities described by the verbs used, which can be perceived as rather abstract and difficult to

---

· Corresponding author: Tanja Samardžić, Faculty of Philology, Department of General Linguistics, Studentski trg 3, 11000 Belgrade, Serbia; samardzic@fil.bg.ac.rs.

<sup>†</sup> This paper has been written as part of the project No. 178004, *The Standard Serbian Language: Syntactic, Semantic and Pragmatic Explorations (Standardni srpski jezik: sintaksička, semantička i pragmatička istraživanja)*, funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia, participated in by Maja Miličević.

pin down. Native speakers of English, like the speakers of many other European languages, are not forced to make this distinction when selecting the verbs to use; note that the same verb in the same form (*turned*) appears in both (1a) and (2a). Speakers of Serbian (and other Slavic languages), however, need to make the aspectual distinction in order to select the appropriate verb. In other words, Serbian verb inventory contains two different forms, one that is suited for describing temporally bounded (*okrenuo* in (1b)), and the other for temporally unbounded eventualities (*okretao* in (2b)). Such verbs constitute pairs that are in fully complementary distribution: exchanging the forms between the unbounded and bounded contexts makes the sentences ungrammatical, as shown in (1c) and (2c). The forms that are used in temporally unbounded contexts are labelled as *imperfective*, and those used in temporally bounded contexts as *perfective*.

For a better understanding of these two categories, we can relate them to the widely used four-way taxonomy of aspectual classes proposed by Vendler (1967). As shown in Figure 1, Serbian perfective verbs can generally be seen as including *Achievements* (instantaneous events with an endpoint) and *Accomplishments* (incremental events with an endpoint), while Serbian imperfectives roughly correspond to *States* (static events with no endpoint) and *Activities* (dynamic events with no endpoint).<sup>1</sup>

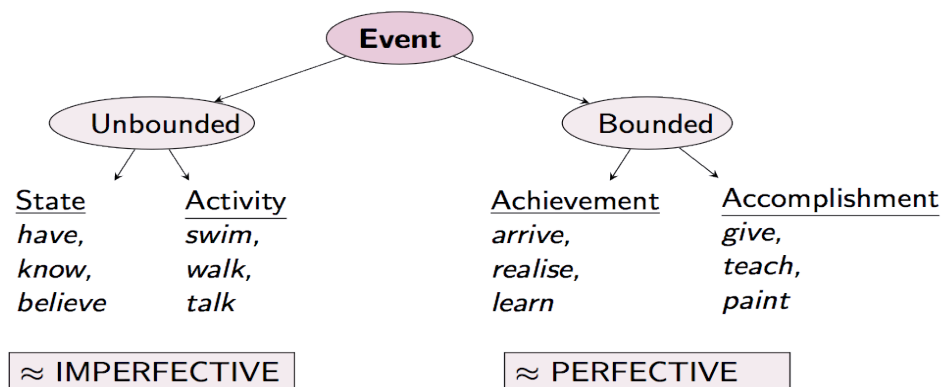


Figure 1. Verbal aspect in Serbian compared to Vendler's aspectual classes

Mastering the perfective/imperfective distinction has consistently been recognised as one of the most difficult tasks facing second language (L2) learners of Slavic languages in general (see Slabakova 2005 and Mikhaylova 2011 for Russian, Schmiedtová 2003 for Czech, Kozłowska-Macgregor 2002 for Polish), and Serbian in particular (Klajn 2006, Babić 2011b; see also Mønnesland 2003, Jelaska and Opačić 2005, Cvikić and Jelaska 2007 for Croatian).<sup>2</sup> The task is a difficult one for two reasons. The first difficulty lies in identifying the clues for deciding whether a perfective or an imperfective form is required, especially given that the clues are often provided only by the broader context of a particular utterance. The second difficulty is a consequence of the fact that aspect in Serbian, like in other Slavic languages, is encoded in the lexicon, which means that there are no general grammatical rules for deriving the required aspectual forms. The verbs in (1b) and (2b) are obviously morphologically related; however, as we will see in more detail in the next section, the *-ta-* to *-nu-* alternation is only one of the numerous possibilities that can be used in aspectual derivations.

As has already been pointed out in relation to L2 Croatian by the above authors, in addition to the intrinsic complexity of the phenomenon, the problem also seems to lie in the teaching and learning materials, which do not treat verbal aspect in a learner-friendly manner. Since general rules that would capture the aspectual nature of Serbian verbs are hard to define, we believe that learners

<sup>1</sup> There are some exceptions to the described correspondence. For example, the Serbian equivalent of the sentence *Winston stayed in the shop for two hours*, *Vinston je ostao u prodavnici dva sata*, contains a perfective verb describing a state. However, in the majority of the cases, the mapping to Vendler's classes does apply.

<sup>2</sup> A number of morphological differences exist between individual verbs in Serbian and Croatian, but overall it can be said that their aspectual systems are alike, posing the same kinds of problems to L2 learners.

need to have available to them specialised learning resources, rich in data and examples of use, that would support example-based learning; this need is far from being met by the currently available textbooks, workbooks, grammars and dictionaries.

In this paper, we propose a corpus-based method for compiling a novel and richer resource for the L2 acquisition of Serbian verbal aspect. We automatically extract and store information about the aspectual behaviour of a large number of verbs, and we perform a simple statistical analysis of the corpus data to identify characteristic trends which we then use to identify potential generalisations, and also to list as many idiosyncratic cases as possible. We argue that the quantitatively defined concept of derivational *productivity* plays an important role in making useful generalisations about aspectual derivations in Serbian. We use this concept in constructing a detailed yet compact representation of the key features of aspectual formation that will provide good training material for learners.

In the next section, we give a brief outline of aspectual derivations in Serbian, after which we provide an overview of the currently available learning materials and their approaches to verbal aspect (Section 3). We then describe our own approach, based on automatic extraction of relevant corpus data, a statistical analysis of the trends in the data, and a compilation of aspectual information to be included in each verb's dictionary entry (Section 4). Finally, we sum up the results and point to some directions for future work in Section 5.

## 2. Aspectual derivations in Serbian

As we will see in more detail in Section 3, the learning materials available for L2 Serbian largely base their treatment of aspect on verb pairs. However, as shown in Table 1, rather than pairs, Serbian verbs are organised into aspectual *sequences*. Multiple affixes can be added to the same base verb, modifying its meaning and its aspect in different ways. The forms in the first column are the result of prefixation, which is in many ways similar to the attachment of particles to English verbs, as the translations of the prefixed forms suggest (cf. Milićević 2004). The prefix specifies the meaning of the base verb by introducing an additional resultative predication into the verb's lexical representation (Arsenijević 2007), and the aspectual change in this derivation is a consequence of the fact that the result state introduced by the prefix makes the event temporally bounded.

$pref(x) = P$ 'complete a specified $x$ '	$suff(pref(x)) = I$ 'do $pref(x)$ continuously or repeatedly'	$pref(suff(pref(x))) = P$ 'complete multiple $pref(x)$ '
skuvati 'cook <sub>P</sub> '	—	—
prokuvati 'boil <sub>P</sub> briefly'	prokuvavati	isprokuvavati
iskuvati 'cook <sub>P</sub> well'	iskuvavati	iziskuvavati
otkuvati 'clean <sub>P</sub> by boiling'	otkuvavati	izotkuvavati
zakuvati 'add <sub>P</sub> something into boiling liquid'	zakuvavati	izakuvavati

Table 1. Aspectual derivations of intransitive base verbs:  $x = kuvati$  (I) 'cook'

In some cases the derived verbs can be further modified, as shown in the second column of Table 1. By attaching a suffix, the verb becomes a *secondary imperfective*, obtaining a new imperfective interpretation, ambiguous between a progressive and an iterative meaning. Lastly, the prefixed forms in the third column, derived from secondary imperfectives, can be regarded as

describing multiple bounded events.<sup>3</sup>

It should also be noted that not all aspectual affixes that can be attached to Serbian verbs have the same grammatical status. The prefixes in the first column are associated with lexical content, which they contribute in the derivation of the prefixed verbs. Possible combinations of verbs with these prefixes are determined by their lexical compatibility (or lexical preferences); for instance, fewer prefixes can be combined with the verb *kuvati* from Table 1, than with the verb *pisati* ‘write’. In contrast, the progressive/iterative suffix in the second column is a functional morpheme with no independent lexical content. Even though the form of the suffix is not the same for all verbs (cf. *-va-* for *kuvati* vs. *-ja-* for *piti* ‘drink’, e.g. in *napijati* ‘get drunk’), this variation is not influenced by the verbs’ lexical content. Functional status can also be assigned to the prefix *iz-* in the third column; this prefix does not vary among verbs, except for the strictly phonetic adaptation (regressive assimilation).<sup>4</sup>

Lexical and aspectual derivations are also allowed with verbs whose basic form is perfective; this paradigm is illustrated in Table 2. In this case prefixation does not lead to a change in aspect and the verbs in the first column remain perfective. The remainder of derivations proceed in the same way as for the imperfective base forms.<sup>5</sup>

<i>suff(x) = I</i> ‘do <i>x</i> continuously or repeatedly’ → <i>bacati</i>		
<i>pref(x) = P</i> ‘complete a specified <i>x</i> ’	<i>suff(pref(x)) = I</i> ‘do <i>pref(x)</i> continuously or repeatedly’	<i>pref(suff(pref(x))) = P</i> ‘complete multiple <i>pref(x)</i> ’
<b>prebaciti</b> ‘transfer <sub>P</sub> ’	<b>prebacivati</b>	<b>isprebacivati</b>
<b>izbaciti</b> ‘throw <sub>P</sub> out’	<b>izbacivati</b>	<b>izizbacivati</b>
<b>ubaciti</b> ‘throw <sub>P</sub> in’	<b>ubacivati</b>	<b>izubacivati</b>
<b>odbaciti</b> ‘reject <sub>P</sub> ’	<b>odbacivati</b>	<b>izodbacivati</b>

Table 2. Aspectual derivations of transitive base verbs: *x = baciti* (P) ‘throw’

There are several other patterns of lexical expression of aspectual classes in Serbian: some verbs do not have a base form and are always prefixed (e.g. *premostiti* ‘move’ - \**mestiti*); some perfective verbs have no imperfective counterparts (such as *klonuti<sub>P</sub>* ‘succumb’, which is derived by attaching a perfective suffix directly to the basic form) and vice versa (*jadikovati<sub>I</sub>* ‘moan’); some verbs form suppletive pairs (*reći<sub>P</sub>* - *govoriti<sub>I</sub>* ‘say’), etc. However, it is possible to identify a core set of fairly regular aspectual sequences where knowing that the base form can be further transformed allows a rule-based derivation of a number of related forms. A summary of these regular possibilities is given in Figure 2.

<sup>3</sup> Another common way of forming perfective verbs is via the *-nu-* affix, as in *kucati* > *kucnuti* ‘knock’. These perfectives, however, cannot form secondary imperfectives.

<sup>4</sup> Other prefixes that can be used in a similar way, and with similar meanings, are *na-* and *po-* (cf. *naizbacivati*, *poizbacivati*).

<sup>5</sup> As illustrated in the top row of Table 2 for *baciti*, perfective base verbs can have imperfective forms, derived through a change in a stem vowel (*baciti* > *bacati*), a vowel and a consonant (*skočiti* > *skakati* ‘jump’), or an accent (*pogledati* > *poglédati* ‘look’).

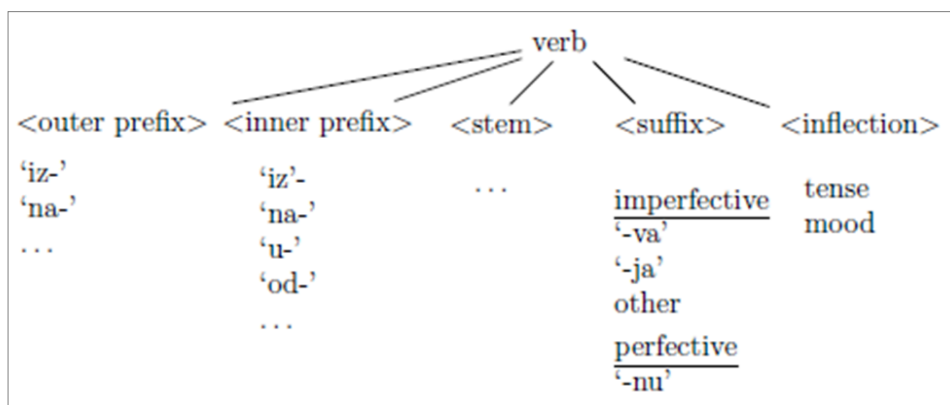


Figure 2. Summary of Serbian aspectual derivations

Having this picture in mind, perhaps the most important fact to note in the description of aspectual derivations in Serbian is that even in this ‘regular’ domain not all prefixed verbs can be further modified. The verb *skuvati* in Table 1, for example, does not have the forms that would belong to the second and the third column. The fact that further derivations are blocked in some cases is not only subject to much theoretical debate (see Svenonius 2004, Arsenijević 2006, Žaucer 2010), but is also crucial for L2 acquisition of verbal aspect, as learners need to exclude those derivations that are not acceptable. This task can make verbal aspect difficult to acquire even for speakers of other Slavic languages, as the options allowed by one language do not always coincide with those permitted by others. For instance, Mønnesland (2003: 23) points out that while Croatian (like Serbian) disallows the derivation *pisati* > *napisati* > \**napisivati* ‘write’, in Bulgarian it is possible to have *piša* > *napiša* > *napisvam*.

Contrary to existing theoretical proposals, which argue for structural differences between *superlexical* (for example *s-* in *skuvati*) and *lexical* prefixes (all the other prefixes in the first column of Table 1) (Svenonius 2004), we believe that a better insight into this problem can be gained through a quantitative analysis of verb uses in a corpus based on the notion of prefix productivity, and we argue that generalisations from such an analysis can help L2 learners in the difficult task of mastering verbal aspect.

### 3. Verbal aspect in L2 Serbian teaching and learning materials

Despite being recognised as a difficult area for learners, verbal aspect does not receive much attention in L2 Serbian teaching and learning materials. The approach adopted in textbooks is mostly inductive, and both perfective and imperfective verbs are used from the beginning, while explanations of their formation and use are postponed. This can in some respects be seen as being in line with the communicative teaching method; however, there appears to be a lack of systematicity in the presentation of aspectual phenomena: formation and use are typically dealt with simultaneously, and aspectual morphology, regardless of whether it is presented through rules or verb lists, tends to be covered in a way that makes it difficult for learners to see which forms are productive and which are not.

In *Serbian for Foreigners* (Ćorić 2008) the use of aspectual pairs is avoided until the point where verbal aspect is explained in lessons 15 and 16 (out of 20); a list of about 100 verb pairs is provided at this point, split into subsections based on the morphological mechanisms applied in the derivation. In the set *Let’s learn Serbian*, the first book (Bjelaković and Vojnović 2006) introduces several examples of aspectual pairs in lesson 4 (e.g. *piti* ‘drink’ - *popiti* ‘drink up’), while the second book (Alanović et al. 2007, lesson 7) gives the key usage and formation rules, as well as a larger number of examples, including those of aspectual “families” such as *napisati* ‘write’ - *zapisati* ‘write down’ - *dopisati* ‘add in writing’ - *prepisati* ‘copy’, and similar. The teacher’s manual *Lektorske vežbe* (2010) comprises a section on verbal aspect, with a theoretical recap and

sample exercises for different proficiency levels, again mostly based on aspectual pairs. Lastly, in the collection of exam practice tests by Krajišnik and Marinković (2009) verbal aspect is included in the descriptors for CEFR A2 level; at this stage, learners are expected to know the aspect of very frequent verbs such as *sesti/sedeti* ‘sit’ or *napisati/pisati* ‘write’. The knowledge of verbal aspect is tested in more detail at level B2, where some of the tasks involve choosing the aspectual form appropriate for the given context and conjugating the verb.

More theoretical information is provided in some of the other resources, most notably two grammars aimed at L2 learners of Serbian. Klajn (2006) gives numerous examples of aspectual pairs derived via different mechanisms; he highlights, however, that morphophonological properties of the two types of verbs and a complete set of rules for their derivation are impossible to define in a principled way, which makes it necessary for verbal aspect to be learned from contextualised use. Mrazović and Vukadinović (2009) also focus on aspectual pairs, explaining in detail the (ir)regularities in their derivation and providing an elaborate list of about 30 verb pairs formed through different morphological processes, followed by examples of use. This grammar also mentions that some derived prefixed verbs (*napisati* ‘write’, *raširiti* ‘spread’, *iscediti* ‘squeeze, drain’, among others) lack secondary imperfectives, but without discussing this issue in more depth.

In the domain of dictionaries, we single out Babić (2011a), a collection of verbs used in the *Let’s learn Serbian* textbooks. Each verb’s entry in this dictionary contains information on whether the verb is perfective or imperfective (or unspecified). In addition, each entry includes a section that provides the verb’s aspectual pair(s); crucially, multiple pairs are given for some verbs: the base and the secondary imperfective for derived perfectives (e.g. *izabrati - birati* and *izabirati* ‘choose’), and several prefixed verbs for basic imperfectives (e.g. *trčati* ‘run’ - *potrčati* ‘start running’, *dotrčati* ‘arrive by running’, *istrčati* ‘run out’).<sup>6</sup> However, the criteria that the choice of pairs is based on are not always clear (e.g. some frequent prefixed forms like *pretrčati* ‘run across’ are not listed), and the above principles are not consistently obeyed (*trčati* is listed for *pretrčati*, but not the other way round). Moreover, due to the focus on pairs, not all related forms are presented together.

As can be seen from this overview, L2 Serbian resources tend to focus on aspectual pairs, largely disregarding the derivational sequences described in the previous section; secondary derivations are rarely referred to as such and prefixed verbs are not shown in complete batches. To some extent, this is understandable, as textbooks need to cover a lot of material, and Babić’s dictionary also provides other kinds of grammatical and lexical information about Serbian verbs. The main problem, however, lies in the failure of the existing resources to point to the regularities and patterns that can be identified in this complex domain: verb pairs that are introduced together belong to different morphological types, some derived lexically and some grammatically, some permitting secondary derivations and some disallowing them; furthermore, the focus is on frequent verbs, while the less frequent ones are neglected despite being productive and thus easier to learn.

As a (partial) solution to these problems, in the next section we present our proposal for the construction of a specialised learner-friendly dictionary of Serbian aspectual forms. We argue that such a dictionary should be corpus-based, in order to reflect actual language use, and to enable making an informed distinction between productive and unproductive forms.

## 4. Corpus data for L2 acquisition of verbal aspect in Serbian

### 4.1 Rationale, materials and method

Our approach to compiling a resource that will faithfully represent the complex knowledge about verb aspect formation in Serbian is based on automatic extraction of relevant data from language corpora. The main idea of this approach is to look at the different types of aspectual derivations in a corpus, which we consider to be a sample of spontaneous language use, and draw generalisations on the basis of the statistical tendencies observed.

---

<sup>6</sup> The verbs listed in the aspectual pairs section are not necessarily used in the textbooks (Babić 2011b: 188).

The corpus that we use for this particular study is the manually annotated Serbian translation of G. Orwell's "1984" (Multext-East project; Krstev et al. 2004, Erjavec 2010); it contains 108.805 tokens, with 8.392 annotated lemmas. The corpus data are extracted automatically using an analyser that implements the derivational rules described in Section 2. Even though such an analyser does not identify all the possible forms, it can provide a good approximation of the morphological structure of a large set of verbs. We extract and analyse all occurrences of verbs derived with nine different prefixes: *iz-*, *od-*, *na-*, *u-*, *o-*, *po-*, *pre-*, *raz-*, and *za-*, which is roughly one half of the total number of verb prefixes in Serbian. The information about the verbs' aspect is obtained from the morphosyntactic annotation.

Our goal is to provide learners with a representation of aspectual forms such that it groups together multiple derivations that apply to single basic verbs; clearly, these derivations must first be identified in the corpus. Identifying which derivations apply to which verb consists of two tasks: 1) isolating the set of prefixes and suffixes that can be attached to a given basic form, and 2) distinguishing between the prefixed forms that can be further imperfectivised and those that cannot.

To address these tasks, for each verb we identify the set of co-occurring prefixes and suffixes and we estimate the co-occurrence likelihood based on corpus data. We calculate the type frequency of each prefix, i.e. we count the number of different verbs it forms relative to its overall frequency; the obtained frequencies represent the distribution of prefixes across verb types. If a prefix can be associated with many different verbs, it can be considered as productive, and thus likely to be attached to any verb. This further implies that it can be expected to co-occur even with verbs it was not found with in a particular corpus, and it should thus be listed in a dictionary by default. In contrast, if a prefix is frequently found with a smaller number of verbs, and only occasionally with others, this points to an association pattern, and such a prefix should be listed only with a limited set of verbs (possibly only those it was found with in the corpus).

In order to identify the prefixed forms that can be further imperfectivised, for each prefix we count the number of secondary imperfectives formed by attaching the suffix *-va-*, which can be considered a regular derivation. We then calculate the ratio between the number of these forms and the overall number of imperfectives; the greater the value of the ratio, the more regular the attachment of the secondary imperfective suffix.

#### 4.2 Results and discussion

The results of our sample analysis are shown in the top row of Figure 3. We can see in the graph on the left-hand side that the prefixes we analysed differ in productivity. Some of them (*iz-*, *za-*, *po-*, *u-*) are combined with many different verbs, while others (*pre-*, *raz-*, *na-*) are combined with fewer; note that this difference cannot be caused by different overall frequencies of the prefixes, as the counts are relative to the overall frequency. As can be observed in the graph to the right, secondary imperfective derivations are also more frequent with some prefixes than with others.

What is particularly interesting about these results is the fact that the two distributions might be correlated: the similarity of their shapes suggests that the more productive prefixes tend to be the same ones that are typically found in secondary imperfectives. This implies that secondary imperfective derivations can be expected with verbs formed by prefixes that are generally more productive, while they are less likely with verbs formed by unproductive prefixes.



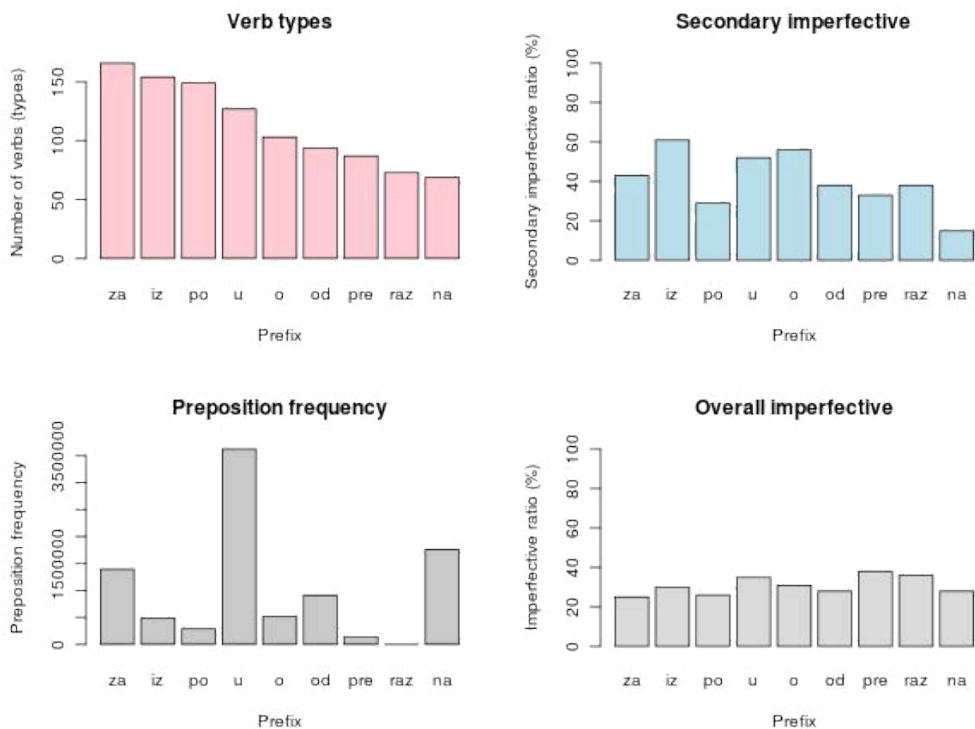


Figure 3. Frequency distributions of counted items across 9 studied prefixes: the number of different verb stems combined with each prefix, the proportion of secondary imperfectives out of all imperfective forms, the frequency of corresponding prepositions, and the overall proportion of imperfective forms

As we had not predicted this outcome, we did not set up the experiment in a way that would allow us to measure the strength of the correlation, so for the time being we are dealing with an approximate insight. We have, however, performed additional tests to make sure that this result is not due to some other factors. To check that the observed asymmetry is not related to the prefixes' frequencies when used as independent prepositions, we extracted the frequency data on the prepositions corresponding to the prefixes included in our study. The counts are shown on the left-hand side of the bottom row in Figure 3. Since morphological tags were not necessary in this case, we extracted these counts from the larger, non-annotated *Corpus of Contemporary Serbian Language* (113M words, <https://korpus.matf.bg.ac.rs>); the count equals to 0 for *raz-* because this prefix does not have a corresponding preposition (despite having a clear spatial meaning, 'apart'). To make sure that the proportion of secondary imperfectives is not influenced by the overall proportion of imperfective forms, we counted the total number of imperfectives for each prefix; this is shown in the graph in the bottom right corner of Figure 3.

The distributions shown in the lower portion clearly do not follow the pattern that can be observed in the upper portion of Figure 3. On the basis of these tests, we can exclude the above two factors as potential sources of the observed asymmetry, which supports our claim that the productivity of a prefix and the likelihood of attaching a secondary imperfective suffix are correlated. However, further research is needed to definitely confirm this correlation. In addition to providing an exact measure of the correlation, it is also necessary to develop a more elaborate measure for the productivity of verb-prefix combinations, e.g. a measure based on hapax legomena, as proposed by Baayen (1992). We do not pursue this issue further in the current study; instead, we focus on the role that the distinction between productive and unproductive derivations plays in the representation of verbal aspect for learners. We demonstrate a possible application of our findings on a practical example.

### 4.3 A practical example

Table 3 shows the counts collected for the verb *pisati* ‘write’ using the approach outlined above. The solution that we propose for the problem of verbal aspect coverage in L2 Serbian materials consists in representing three kinds of information automatically acquired from a corpus. First, we find the basic form of the verb (*pisati* in Table 3) and its aspect (in this case, imperfective). Second, we identify the unproductive derivation (*napisati*) and its aspect (perfective); this form is derived using an unproductive prefix (note that *na-* is placed on the far right of the first graph in Figure 3). Third, we list a number of representative productive forms, together with their secondary imperfective derivations; there is no need to list the perfective forms that correspond to the third column of Tables 1 and 2, as these forms are very rare, and are formed in a fully regular manner, which means that they can be described by a rule.

Basic form:	
<i>pisati</i> (43)	Imperfective
Unproductive prefix:	
<i>napisati</i> (22)	Perfective
Productive forms:	
Perfective:	Imperfective:
<i>ispisati</i> (6)	—
<i>opisati</i> (6)	<i>opisivati</i> (1)
<i>potpisati</i> (3)	<i>potpisivati</i> (1)
<i>prepisati</i> (2)	<i>prepisivati</i> (1)
<i>prispisati</i> (1)	<i>prispisivati</i> (1)
—	<i>propisivati</i> (1)
<i>upisati</i> (1)	<i>upisivati</i> (2)
<i>zapisati</i> (3)	<i>zapisivati</i> (1)

Table 3. An example: *pisati* ‘write’

The counts shown in brackets represent the number of times each item was found in the corpus used for this study; these numbers can be taken as the basis for assessing the probability of each form. The probability itself does not have to be a part of the dictionary entry, but it is implicitly included in the representation we propose: the unproductive item (*napisati* in Table 3) is much more frequent than any of the productive forms in the lower part of the table. The empty slots in the productive forms section (the missing counterparts of *ispisati* and *propisivati*) represent cases that need to be included in the representation even though they are not found in the corpus, as discussed in Section 4.1.

In addition to the information about acceptable derivations, the knowledge that this representation makes available to the learners includes the important fact that the form *\*napisivati* is excluded. This is made obvious not only by omitting this form from the list, but also by the prominent position of its existing perfective counterpart, which is separated from all the other forms.

## 5. Conclusion

In this paper we discussed the problem of how to best present verbal aspect to L2 learners of Serbian. The existing materials do not seem to be based on a systematic approach, as they often introduce aspectual pairs of different types together, while neglecting sequences involving secondary derivations; they also focus almost exclusively on frequent verbs, failing to notice that the infrequent ones are more productive and easier to learn. We thus believe that a novel systematic

dictionary of aspectual forms, based on the concept of derivational productivity rather than frequency alone, could be of great help to learners. Our preliminary analysis of prefix productivity and its relation to the availability of secondary aspectual derivations indicates that this idea is on the right track. In further work we aim to explore additional productivity measures, complement the descriptive statistical analyses with inferential tests, and ultimately apply the developed methodology to a larger corpus, which will enable us to extract sufficient information on the forms to be included in the learner dictionary.

Key words: second language acquisition, Serbian, verbal aspect, morphological productivity, learner dictionary, corpus-based approach

## References

- Alanović, M. et al. (2007). *Naučimo srpski 2/Let's learn Serbian 2*. Novi Sad: Dnevnik.
- Arsenijević, B. (2007). Slavic verb prefixes are resultative. *Cahiers Chronos* 17: 197-213.
- Babić, B. (2011a). *Naučimo srpski 1 i 2. Rečnik glagola*. Novi Sad: Filozofski fakultet.
- Babić, B. (2011b). Morfološko-sintaksički minimalni rečnik glagola srpskog jezika kao stranog. *Prilozi proučavanju jezika* 42: 165-196.
- Baayen, R. H. (1992). Quantitative aspects of morphological productivity. In *Yearbook of Morphology 1991* (G. E. Booij and J. van Marle, eds), Dordrecht: Kluwer Academic Publishers, 109-149.
- Bjelaković, I. and J. Vojnović (2006). *Naučimo srpski 1/Let's learn Serbian 1*. Novi Sad: Dnevnik.
- Cvikić, L. and Z. Jelaska (2007). Složenost ovladavanja glagolskim vidom u inojezičnome hrvatskome. *LAHOR: časopis za hrvatski kao materinski, drugi i strani jezik* 4: 190-216.
- Ćorić, B. (2008). *Srpski za strance/Serbian for Foreigners*. Beograd: Čigolja štampa.
- Erjavec, T. (2010). MULTEXT-East version 4: multilingual morphosyntactic specifications, lexicons and corpora. In *LREC 2010 Proceedings*, Valletta, 2544-2547.
- Jelaska, Z. and N. Opačić (2005). Glagolski vid i vidski parovi. In *Hrvatski kao drugi i strani jezik* (Z. Jelaska et al., eds), Zagreb: Hrvatska sveučilišna naklada, 152-170.
- Klajn, I. (2006). *Gramatika srpskog jezika za strance*. Beograd: Zavod za udžbenike.
- Kozłowska-Macgregor, M. (2002). *The state of near-native grammar: A study of aspect in L2 Polish*. Doctoral dissertation, Montreal: McGill University.
- Krajišnik, V. and N. Marinković (2009). *Testovi za polaganje srpskog kao stranog jezika*. Beograd: Filološki fakultet.
- Krstev, C., D. Vitas, and T. Erjavec (2004). MULTEXT-East resources for Serbian. In *Proceedings of 8th Informational Society - Language Technologies Conference, IS-LTC*, 108-114.
- Lektorske vežbe*. (2010). Beograd: International Centre for Slavic Studies.
- Mikhaylova, A. (2011). Interaction of Aspectual Morphology in L2 and Heritage Russian. In *Selected Proceedings of the 2010 Second Language Research Forum* (G. Granena et al., eds), Somerville, MA: Cascadilla Proceedings Project, 63-77.
- Milićević, N. (2004). The lexical and superlexical verbal prefix *iz-* and its role in the stacking of prefixes. *Nordlyd*, 32(2): 279-300.
- Mønnesland, S. (2003). Glagolski vid u hrvatskom jeziku. In *Zbornik Zagrebačke slavističke škole 2002*. (S. Botica, ed.), Zagreb: FF press, 21-31.
- Mrazović, P. and Z. Vukadinović (2009). *Gramatika srpskog jezika za strance*. Novi Sad: Izdavačka knjižarnica Zorana Stojanovića.
- Schmiedtová, B. (2003). The use of aspect in Czech L2. In *Acquisition of Aspect*, ZAS Papers in Linguistics 29 (D. Bittner and N. Gagarina, eds), Berlin: ZAS, 177-194.
- Slabakova, R. (2005) What is so difficult about telicity marking in L2 Russian? *Bilingualism: Language and Cognition* 8(1): 63-77.

- Svenonius, P. (2004). Slavic prefixes inside and outside VP. *Nordlyd* 32(2): 205-253.
- Vendler, Z. (1967). *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.
- Žaucer, R. (2010). The reflexive-introducing *na-* and the distinction between internal and external Slavic prefixes. In: *Formal Studies in Slavic Linguistics* (A. Smirnova et al., eds), Newcastle/Tyne: Cambridge Scholars Publishing, 54-102.

Tanja Samardžić, Maja Miličević

Rečnik glagolskog vida zasnovan na podacima iz korpusa za učenike srpskog jezika kao stranog

#### Sažetak

U radu se predlaže teorijski okvir kao i odgovarajuća metodologija za unapređenje opisa glagolskog vida u srpskom jeziku za potrebe učenja srpskog kao stranog. Uočavajući nedostatak ključnih informacija o izvođenju i upotrebi vidskih formi u postojećim učeničkim materijalima, predlažemo izradu specijalizovanog rečnika u kome bi učenici mogli da pronađu sistematizovane informacije bez kojih je teško dostići neophodnu kompetenciju u korišćenju glagolskog vida. U konstrukciji rečnika koristi se originalan pristup zasnovan na kvantitativnoj distinkciji između produktivnih i neproduktivnih oblika. Produktivni oblici su izvedeni prefiksima koji se kombinuju sa brojnim glagolima, dok su neproduktivni oblici izvedeni prefiksima čija je distribucija ograničena na relativno mali broj glagola. Istovremeno, produktivne forme su ređe u upotrebi, ali više podležu pravilima od neproduktivnih formi. Pomoću praktičnog primera obrade jedne odrednice pokazujemo na koji način izloženi koncepti pomažu u iscrpnoj, ali kompaktnoj prezentaciji glagolskog aspekta u srpskom.