

Designing a Croatian Aspectual Derivatives Dictionary: Preliminary Stages

Kristina Kocijan
Department of information
and communication sciences
Faculty of Humanities
and Social Sciences
University of Zagreb
Croatia
krkocijan@ffzg.hr

Krešimir Šojat
Department of linguistics
Faculty of Humanities
and Social Sciences
University of Zagreb
Croatia
ksojat@ffzg.hr

Dario Poljak
Department of information
and communication sciences
Faculty of Humanities
and Social Sciences
University of Zagreb
Croatia
dpoljak@ffzg.hr

Abstract

The paper focuses on derivationally connected verbs in Croatian, i.e. on verbs that share the same lexical morpheme and are derived from other verbs via prefixation, suffixation and/or stem alternations. As in other Slavic languages with rich derivational morphology, each verb is marked for aspect, either perfective or imperfective. Some verbs, mostly of foreign origin, are marked as bi-aspectual verbs. The main objective of this paper is to detect and to describe major derivational processes and affixes used in the derivation of aspectually connected verbs with NooJ. Annotated chains are exported into a format adequate for a web-based system and further used to enhance the aspectual and derivational information for each verb.

1 Introduction

In this paper we deal with the representation of derivational processes in Croatian, a South Slavic language with rich inflectional and derivational morphology. The paper focuses on derivationally connected verbs, i.e. on those verbs which share the same lexical morpheme and which are derived from other verbs mostly via prefixation and suffixation. As in other Slavic languages, each verb is always marked for aspect and classified as perfective, imperfective, or bi-aspectual. Generally, the imperfective aspect is used to describe actions, processes and states as unfinished or ongoing, whereas the perfective aspect refers to them as finished or completed, e.g.:

- 1a. *Pisao sam* (imperfective) pismo jedan sat. 1b. I *was writing* a letter for an hour.
- 2a. *Napisao* (perfective) *sam* pismo za jedan sat. 2b. I *wrote* a letter in an hour.

Verbs like *pisati* "to write + imperfective" and *napisati* "to write, to finish writing + perfective" are referred to as aspectual pairs. Verbs in aspectual pairs are closely related in meaning, except that one expresses imperfective and the other perfective aspect. Aspect in Croatian is morphologically marked in each verbal form and it affects inflectional properties of verbs (e.g. only perfectives can be used in aorist, and imperfectives in imperfective past tense; gerunds are commonly formed by imperfectives etc). Aspect in Croatian is regarded as a word-formation process and members of aspectual pairs are treated as separate lexical entries in dictionaries. In terms of derivation, perfectives are commonly derived from imperfectives by prefixation, while imperfectives are formed from perfectives by suffixation or stem alternation. The presence of a certain affix indicates whether a verb is a perfective or an imperfective. A relatively small group of bi-aspectual verbs, predominantly of foreign origin, can be used as perfectives and imperfectives in the same morphological form. Various factors can determine whether they will be used as a perfectives or imperfectives (e.g. a context, the type of time adverbial used in a sentence etc.). Although based on the opposition of only two aspects and overtly marked, numerous

studies in the area of second language acquisition indicate that aspect is one of the most complicated category for learners of Slavic languages.

In this paper we present preliminary stages in the construction of the database of Croatian aspectually and derivationally connected verbs, i.e. aspectual derivatives. Apart from its potential pedagogical use, the database of aspectual derivatives is one of the first attempts to systematically present this area of Croatian derivational morphology. The paper is structured as follows: In Section 2, we briefly describe major derivational processes in Croatian and focus on the derivation of verbs from other verbs and aspectual changes that take place. Sections 3 and 4 present the processing of aspectual derivatives in Croatian in NooJ and provide an overview of underlying principles. In Section 5, the design and the structure of the database is discussed. The paper concludes with an outline of future work.

2 Derivation of verbs

Major word-formation processes in Croatian are derivation and compounding. Further we discuss only derivation which predominantly consists of affixation. Although there are some other processes like conversion or back formation, they are not as prominent as prefixation or suffixation (Šojat et al, 2013).

Croatian verbs can be thus divided into simple imperfectives (*pisati* "to write + imperfective") and prefixed perfectives (*na-pisati* "to write + perfective") denoting that the action is completed. Other prefixes used for the derivation of perfectives add different semantic features (*pisati* "to write + imperfective" – *pre-pisati* "to copy by writing + perfective" – *pot-pisati* "to sign + perfective") and enable further derivation, either through prefixation, suffixation or simultaneous prefixation and suffixation. These perfectives can be derived into secondary imperfectives denoting iterative actions through suffixation (*potpis-ivati* "to sign several / many times"). Other suffixes are used for the derivation of diminutive verbs (e.g. *pisati*– *pis-karati* "to scribble + imperfective") or verbs expressing punctual actions (*vikati* "to shout + imperfective" – *viknuti* "to shout once + perfective"). Further, some secondary imperfectives are derived via prefixation into perfectives denoting distributive actions (*is-potpisivati* – "to sign each one + perfective", e.g. each letter, every document etc.). In some cases aspectual distinctions are expressed by vowel variations or suppletive forms (e.g. *doći* "to come + perfective" – *dolaziti* "to come + imperfective"). Detailed account of morpho-semantic relations among Croatian verbal derivatives is found in Šojat et al. (2012).

In the following section we show how existing language resources can benefit from the information about verbal aspect in terms of their extension and enrichment. We demonstrate this on the inflectional dictionary for Croatian verbs in NooJ.

3 Verbs in NooJ Dictionary

The main language resources (LR) for Croatian, as prepared in Vučković (2009) and explained in Vučković et al. (2010), include the dictionary of Croatian verbs. Each verb was originally marked for the main category <V>, reflexive property <V+Prelaz=pov> and a paradigm rule **FLX** responsible for describing the rules used to build all the simple verb forms,¹ for example:

<i>pisati</i> , V+ FLX =PISATI	"to write"
<i>asistirati</i> , V+ FLX =SJATI	"to assist"
<i>sjati</i> , V+Prelaz=pov+ FLX =SJATI	"to shine".

In all cases, a name of a verb is used as a representative of a specific conjugation paradigm. For example, the verb *sjati* "to shine" uses the set of conjugation rules that we refer to as SJATI <FLX=SJATI>. All the other verbs that share the same set of conjugation rules will be associated with this name of the paradigm like the verb *asistirati* "to assist." In addition to these three markers, a small subset of verbs was also marked for valency (Vučković et al., 2010) to improve the performance of the Croatian chunker and some verbs were marked for aspect.

Although each Croatian verb has an aspect by its nature, either perfective, imperfective or bi-aspectual, this information was not originally encoded into the main dictionary entries. This is mainly due to

¹ Complex tenses in Croatian such as future I and II, perfect, pluperfect, or conditionals are described within syntax grammars and are beyond the scope of this paper.

the absence of that information in the resources built after the MULTEXT-East specifications prepared by Tadić, as explained in Erjavec (2001), from which NooJ resources have been adopted.

There are 4,168 main verb entries in the NooJ Croatian dictionary. At the beginning of our project, there were only 1,448 verbs that had been marked as either perfective <Aspect=perf>, imperfective <Aspect=imperf> or bi-aspectual <Aspect=bi> regarding the Category of aspect (Table 1). This means that little over 65% of verbs had no marker assigned for this category.

Total	+perf	+imperf	+bi	no marker
4,168	673	534	241	2,720

Table 1: Original distribution of Aspect markers

The importance of information on aspect encoded in the verb, lies, among others, in the list of possible verbal forms. For example, only perfective verbs can have aorist past tense and past adverbial participle forms, while only imperfective verbs can have the imperfect past tense and the present adverbial participle. Before we could add the tag for aspect automatically to the remaining verbs, it was important to check if the existing aspect markers were correctly assigned and what data could be used to correctly mark the remaining 65% of verbs. Since the list of possible tenses is embedded into the conjugation paradigms (FLX), we decided to start our investigation with them.

3.1 Data Analysis of the original dictionary

At the very beginning of our analysis, we found two paradigms that were using both Aorist and Imperfect endings [FLX=POKLONITI and FLX=BIRATIDODATI]. Two does not sound like a large number, but if we take into account that the POKLONITI paradigm is responsible for the conjugation of 106 verbs and the BIRATIDODATI paradigm for 204, we are not talking small numbers any more. We proceeded with the analysis by double-checking each of the 310 verbs. We learned that all the verbs using the POKLONITI paradigm are actually perfective in aspect (as is the paradigm itself), while the other verbs are all dual which makes the presence of aorist and imperfect endings appropriate. Thus, we were able to annotate these verbs with appropriate markers <Aspect=perf> and <Aspect=bi>, respectively. In addition, we revisited the POKLONITI paradigm and removed the endings for Imperfect. At this point, we also decided to recheck the aspect category of all the verbs.

To avoid individual checking each verb, we cross referenced all of the assigned aspects with the aspect of a verb used to mark the conjugation paradigm. From that list (Table 2), after removing all the verbs whose aspect matched the aspect of a paradigm, and all the verbs that had no aspect defined, we were left with only 62 verbs that had been marked as mismatched and that needed to be checked manually.

Since the reason for the mismatched aspect could be due to either an incorrectly marked aspect of a verb or to an incorrectly described paradigm, we checked both, starting with the verbs. Paradigm analysis revealed some missing aorist and/or past participle descriptions in the category of perfective verbs, and missing imperfect and/or present participle descriptions of imperfective verbs. Also, some descriptions of bi-aspectuals were missing, either aorist and past participle descriptions, or imperfect and present participle descriptions. All of these cases were corrected to mirror the related aspect, either by correcting the value of an Aspect attribute or changing the paradigm name.

For the final analysis we wanted to make sure that there are no duplicates among our paradigms, i.e. that we do not have different paradigm names describing the same conjugation occurrences. We detected 16 such occurrences that we have replaced choosing the one paradigm that was used more often in the dictionary.

3.2 The new verb dictionary and paradigms

There are 209 paradigms that describe the conjugation rules for Croatian verbs. Some rules describe only one verb in a dictionary (there are 67 such rules), while others describe more (Figure 1). The largest number of verbs (538) described by the same paradigm BIRATI make up 13% of all the verbs in the dictionary, while the second runner up (SJATI) describes 7% of verbs.

FLX Aspect	FLX	ASPECT			
		perf	bi	imperf	Null =
bi	ČUTI	6	1		12
	VIDJETI	2	1		4
perf	DODATI	48		5	164
	BLJESNUTI	17		1	72
	ZASJATI	15	2	1	69
	KAZATI	8	1		23
	DAROVATI	3	3	2	13
	UGASITI	5		1	13
	ZAPAZITI	5		1	11
	UGOSTITI	5		1	10
	PRILEĆI	3		1	5
	OTPUTOVATI	3	1	1	1
imperf	BIRATI	4	5	140	402
	SJATI	1	3	71	236
	POSJEDOVATI		2	13	38
	GUBITI	2		8	19
	SMIJATI	1		10	18
	BRANITI	1		11	17
	CVILITI		1	2	16
	RAZLIKOVATI		6	5	11
	OKRETATI	1		3	9
	KLICATI	1		3	7
	UMJETI	3		2	4
	IZLAGATI	1		7	3
	ODMICATI	1		1	

Table 2. Distribution of different aspects assigned to paradigms

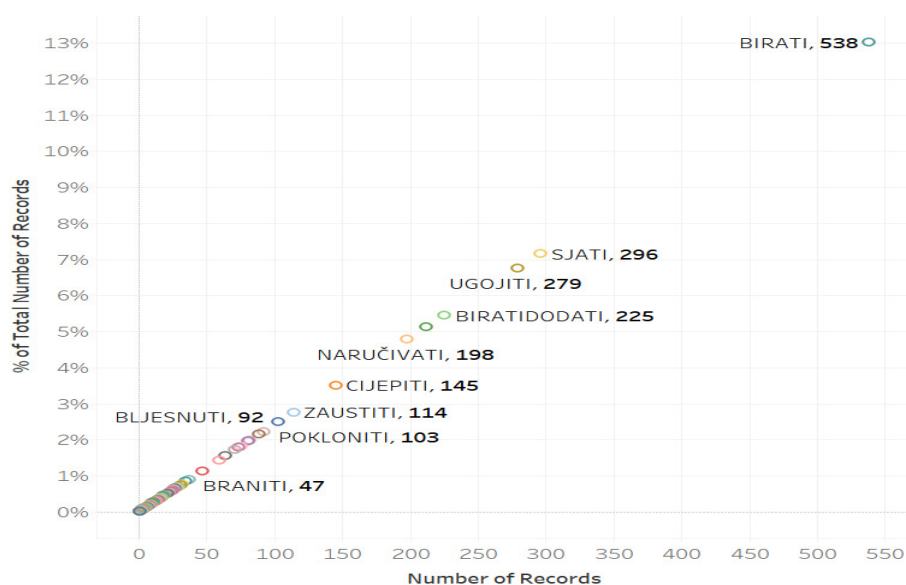


Figure 1. Distribution of paradigms among Croatian verbs

However, a closer look shows us that these descriptions are not as different as they might first appear. We will show this through a detailed analysis of simple verb forms by tense category, starting with present tense.

What makes these paradigms different is the list of tenses they describe, but also how each tense is described. These two categories (list of tenses, form of tenses) are linked by an OR operator, meaning that the difference in either one or both from the existing list of paradigms, results in a new paradigm.

To demonstrate, let us compare the UGOJITI and POKLONITI paradigms. They both have the same list of tenses they can form (present, imperative, PDR - verbal adjective active, PDT - verbal adjective passive, aorist and GPP - past adverbial participle), but the way they make PDT differs [$\langle B3 \rangle_{en}(\text{PDT})$ and $\langle B3 \rangle_{jen}(\text{PDT})$ respectfully] and thus they form different paradigm rules.

However, since NooJ allows multiple usage of its grammars (Silberstein, 2016), each tense rule is described as a separate sub-grammar and then called from a paradigm, where needed. In order to describe derivations of Croatian simple verb forms that are in the NooJ dictionary, we built 280 such sub-grammars whose different combinations build 209 paradigms that describe 4,168 verbs and recognize 377,603 forms, taking into account simple verb tenses (long/short infinitive, present, imperfect, aorist, passive/active verbal adjectives, imperative, present/past adverbial participles), gender (masculine, feminine, neutral), number (singular, plural) and person (1st, 2nd, 3rd).

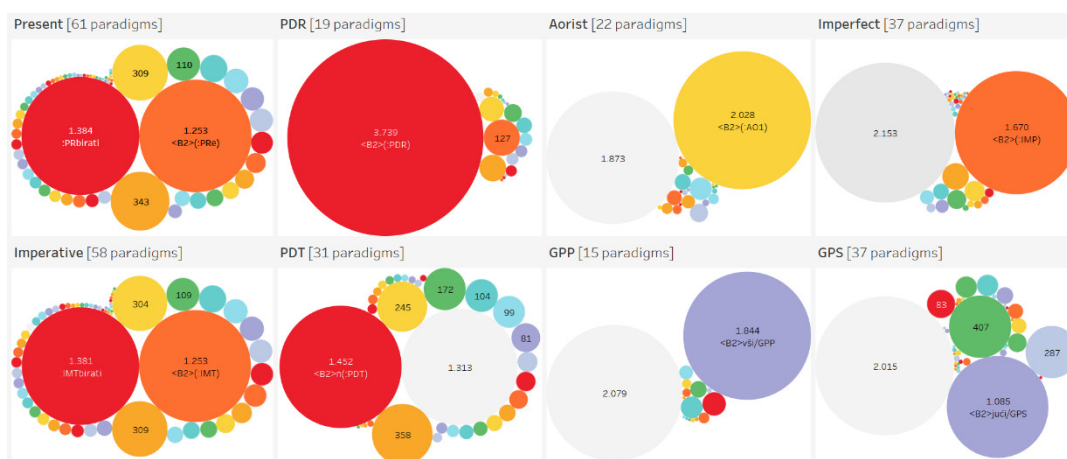


Figure 2. Distribution of sub-grammars per tense

As expected, inside each tense category, some descriptions are more common than others. This distribution (Figure 2) is different for each of the tenses. The same is true for the number of rules which ranges from 15 (*GPP*) to 61 (*Present*). The number of paradigms that do not have rules for a tense are marked in gray (for example: 2,153 paradigms do not have Imperfect, 2,079 do not have GPP, 2,015 do not have GPS etc.). *Present* is the only tense that is found in all paradigms.

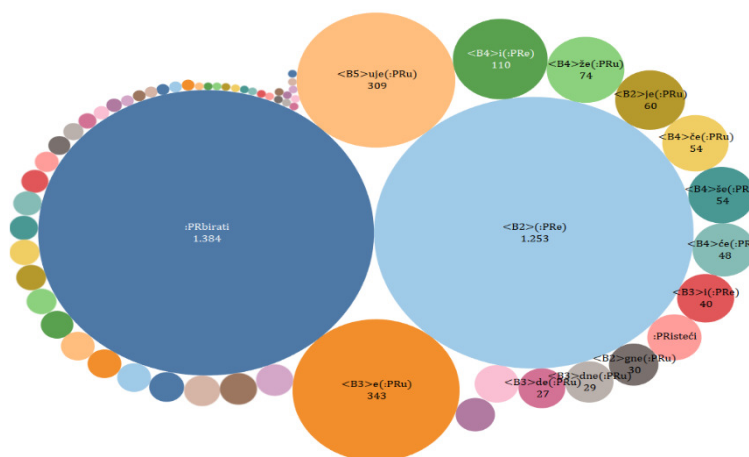


Figure 3. Distribution of Present rules used among existing verbs

Figure 3 shows the distribution of rules² used to build the present tense found in the existing 209 paradigms. It may look as if there are 61 different suffixes for the Croatian present, but this is not the case. Throughout the paradigms, the same set of suffixes is used for all three genders (male, female, neutral), for all three persons singular [-*m*, -*š*, -], and for the first and second person plural [-*mo*, -*te*]. The third person plural may have the ending [-*e*] or [-*u*]. So, why are there 61 different present tense descriptions in this figure?

Although the set of suffixes is the same (with two possible alterations for 3rd person plural), changes that occur before the suffix differ. In some cases it is enough to only remove (<B2>) the infinitive ending

² Each circle represents one rule; the bigger the circle, the more verbs the rule describes.

(-*ti* or -*ći*) and add the suffix, but in some cases more letters need to be removed (<B3>..<B5>) and more letters prior to the suffix need to be added, like in the following examples:

```

isteći -> <B2>će(:PRsingular) => istečem [removes last 2 characters, inserts 'će' before adding suffix for Present]
izleći -> <B2>že(:PRsingular) => izležem [removes last 2 characters, inserts 'že' before adding suffix for Present]
otprijeti -> <B5>e(:PRsingular) => otprem [removes last 5 characters, inserts 'e' before adding suffix for Present].

```

After removing duplicate verbs from the dictionary and sorting out the paradigm sets, we were able to automatically add the missing aspect information for all unmarked verbs. Their total now amounts to 4,134. The largest aspect category are perfective verbs, followed by imperfective verbs and bi-aspectual verbs. Their distribution is visualized in Figure 4.

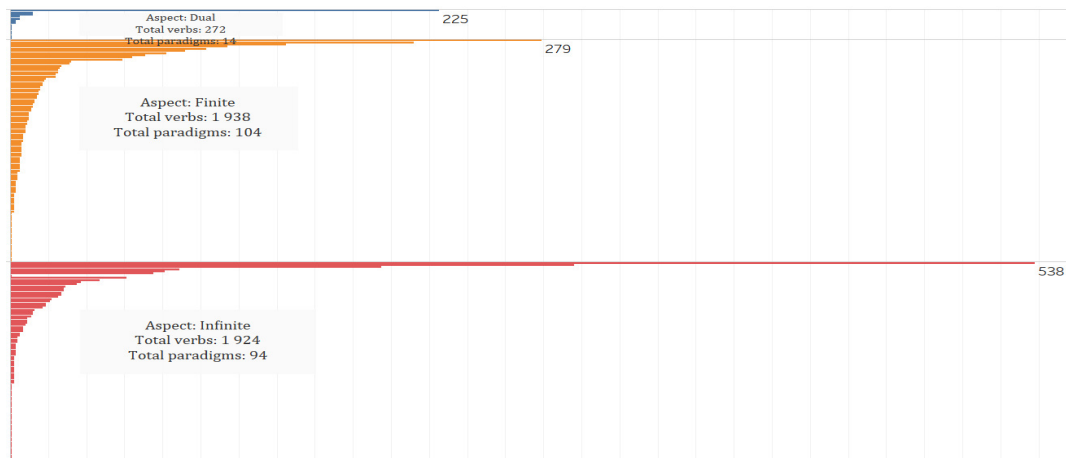


Figure 4. Distribution of verbs per aspect category in the dictionary and number of paradigms used to describe each category

4 Grammar modeling for aspectual derivatives

Computational derivation is a well-known process when new words need to be created in order to economically enlarge the dictionary (Trost, 2003). NooJ provides two routes to describe derivations.

The first one uses a derivational module that allows a direct link between the dictionary with the list of words that can be derived and a grammar that provides rules for their derivation either graphically in a form of an enhanced recursive transition network (ERTN) or via formal grammar rules as context-free grammar (CFG). This link is defined via an attribute DRV that holds the name of the paradigm responsible for the allowed derivation(s).

The second one uses a morphological grammar module that may simulate the dictionary entries via ERTN. It can recognize a defined set of letters and tag them in the same manner that we would manually do in the dictionary. The difference is that in the grammar can have a few graphs describing multiple dictionary entries (e.g. if we wanted to recognize and tag roman numerals, we can do it by a minimum of five graphs or by 3,999 dictionary entries).

Since our main objective is to produce derivational paths in a format that we can use to populate the web-based database, we have opted for the second approach that left us more room to accommodate the output to the database design (see Figure 5). To avoid recognizing words that start with the same set of letters as prefixes used in derivations, we introduced the constraint that the dictionary check and validate if the primary verb first exists. So, for example, *sufinancirati* "co-finance" will be recognized, since there is a verb *financirati, V* in the dictionary. On the other hand, the word *suncobran* "parasol" will not be recognized, since there is no dictionary entry marked as *ncobran, V* (nor any such word in Croatian).

For the preliminary grammar model, we used all the derivations for the verb *pisati* "to write" (Table 3a) and the verbs *bacati / baciti* "to throw away" (Table 3b). All the pairs have both perfective and

imperfective forms. This is not true in only two cases: for the derived form *napisati* that has no aspectual pair, and the aspectual pair *ispotpisati* - *ispotpisivati* that share the same aspect (perfective).³ If we put aside these two exceptions, from the remaining pairs we can conclude that if there is a verb in the dictionary to which a prefix is added, then the newly derived verb will be perfective in aspect. If a verb derived in such a manner adds the suffix2 (SUF 2), then the new verb will be imperfective in aspect if the length of suffix2 is 3 or 4 and perfective if its length is 0, 1 or 2.

a) PISATI				b) BACATI - BACITI				
PREF2	PREF1	a	iva	PREF2	PREF1	a	i	iva
0	0	PISATI		0	0	BACATI	BACITI	
	do	dopisati	dopisivati		do		dobaciti	dobacivati
	is	ispisati	ispisivati		iz		izbaciti	izbacivati
	na	napisati			na	nabacati	nabaciti	nabacivati
	o	opisati	opisivati		od		odbaciti	odbacivati
	ot	otpisati	otpisivati		po	pobacati	pobaciti	pobacivati
	po	popisati	popisivati		pod		podbaciti	podbacivati
	pot	potpisati	potpisivati		pre		prebaciti	prebacivati
	pre	prepisati	prepisivati		pred		predbaciti	predbacivati
	pri	prispisati	prispisivati		raz	razbacati		razbacivati
	pro	propisati	propisivati		u		ubaciti	ubacivati
	ras	raspisati	raspisivati		uz		uzbaciti	uzbacivati
	u	upisati	upisivati		z		zbaciti	zbacivati
	za	zapisati	zapisivati		za		zabaciti	zabacivati
is	pot	ispotpisati	ispotpisivati	po	raz	porazbacati		porazbacivati
na	do	nadopisati	nadopisivati					
su	pot	supotpisati	supotpisivati					

Table 3. Aspectual derivatives of the root verbs
a) *PISATI* "to write" and b) *BACATI* / *BACITI* "to throw away"

We have applied these rules to the morphological grammar built in NooJ. Figure 5 illustrates how the grammar works, using the verb *ispisivati* "to write out."

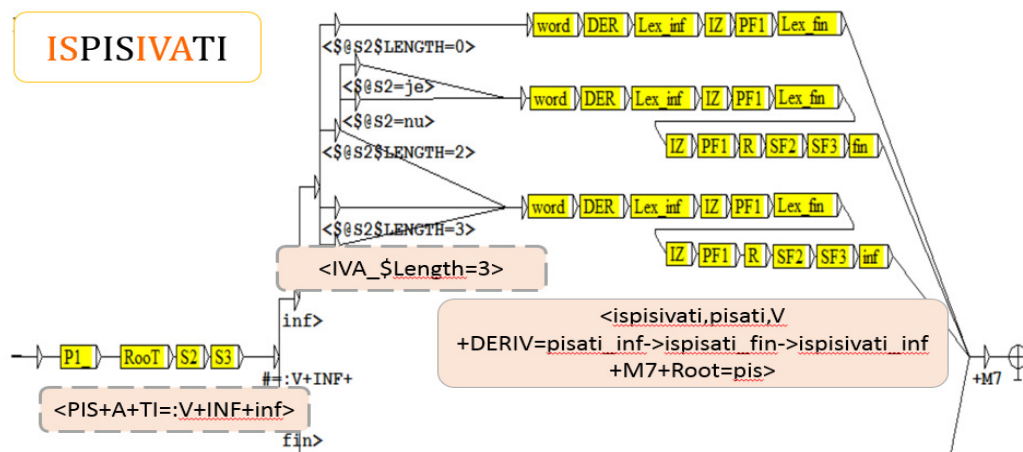


Figure 5. Morphological grammar that recognizes and annotates a verb derived by a single prefixation and suffixation (example of the verb *ispisivati*)

Possible prefixes in the first position (i.e. the position closest to the root), such as the prefix *is*, are listed in the P1_ node. The following node holds any set of letters which are recognized as the root of the verb used to build the constraints that check if such a root concatenated with *a + ti* exists in the dictionary as a verb in infinitive form whose Aspect is defined as INF <pis+a+ti=:V+INF+inf>. If this constraint is validated, we check against the length of suffix2. Since the length in our example is 3, we proceed with the path where <S2\$LENGTH=3>. It then leads us to the annotation section that ads

³ This may be due to the fact that the verb *ispotpisivati* is actually derived from *potpisivati*, while the verb *ispotpisati* is redundant in semantical meaning of its prefixes i.e. the prefix *is* does not bring anything new to the meaning of *pot* in this context. In the hrWaC 2.2 web corpus (Ljubešić & Klubička, 2014) it shows up only 7 times, mostly in an informal web setting.

the recognized lemma as the superlemma of the derived verb, and marks the POS, Root and Derivational chain with the aspect marker for each derivation [+DERIV=pisati_inf->ispisati_fin->ispisivati_inf]. This information is then exported from NooJ and added to the *Specifics* table of our web database as discussed in Section 5.

In Table 3b, there is an imperfective verb BACATI "to throw away" from which the perfective verb *nabacati* "to throw onto" is derived by prefixation. Its imperfective pair is the verb *nabacivati* "to throw onto" derived via suffixation. The perfective verb BACITI uses prefixation to produce a perfective form *nabaciti* whose imperfective pair is also *nabacivati*. This means that the imperfective form *nabacivati* should be found in both aspectual derivational chains. But, ambiguity is not a stranger to language.

5 Database and interface design

After extracting data into usable chunks, we wanted to present it in a way usable to others. To reach the widest possible audience with our tool, we focused on bringing it to the web. In that way it will be available to everyone with basic Internet access and can be dynamically updated as new chains are prepared within NooJ. However, to accomplish that, we needed to create a searchable interface backbone in a well-structured database, whose main function is to support our information system as defined in Gunjal (2003).

Due to the nature of our data, we decided to split it into three separate semantic data-groups (Figure 6). The *Main data* set stores all the data at the morphological level that will be searchable through the online tool, with various levels of granularity. The *Specifics* provides derivational data focusing on one model used for the derivation. The *Examples* set unifies all the semantics and should provide a better understanding of the verb's usage.

Main data										Specifics			
Form	Root 1	Suffix 1	Suffix 2	Suffix 3	Prefix 1	Prefix 2	Prefix 3	Prefix 4	Infinitive	Form	Root	Model	Derivation
ispisivati	pis	NULL	iva	ti	is	NULL	NULL	NULL	True	ispisivati	pis	M7	pisati_inf->ispisati_fin->ispisivati_inf

Examples						
Form	Sense 1	Sense 2	Example Croatian 1	Example Croatian 2	Example English 1	Example English 2
ispisivati	write out	print	Kako polaznici budu odgovarali, njihove odgovore možete ispisivati na ploču.	Ako je sadržaj cijelog reda nepoznat ili ga ne treba ispisivati, tada se izostavlja cijeli red.	As students respond, you may want to write their answers on the board.	If the content of a complete line is unknown, or need not to be printed, then the complete line is omitted.

Figure 6. Data structure of three data groups

Since we are using csv to represent our data, it is already in a rather structured state. Thus, our data can be described as a structure in the first normal form. This means that the data itself can fit into tabular format and that it always contains only one value for each cell. The first normal form also assumes the usage of primary keys for the unique representation of every row of data. This can be performed automatically while importing it into the database as described in Gilfillan (2015).

As can be seen, there is still some redundant data (Figure 6) since the *Main Data* file contains both Form and Root components, present in other files as well. The idea behind the interface is to allow users to search either by *Form* or *Root* fields with additional (optional) *Suffix* and *Prefix* information, and then conditionally showing *examples* and *specifics* depending on which *Form* is selected. Thus we can reduce the clutter in the other two files by imposing a foreign key constraint after importing it into our database and removing *Form* and *Root* information from the other two tables (since that data will already be in the search results from querying the *Main Data* file).

We used MySQL (Coulter, 2017) to store the data and created primary keys, as well as foreign key constraints. After importing all the data and imposing foreign key constraints, we are left with 3 tables (Figure 7) with almost identical structure as we had at the beginning (Figure 6). Now the *Form* and *Root* fields are present exclusively in the *Main Data* table. In order to access it from other tables, we use the foreign key named *mainID* constrained to the *ID* field. It also acts as the primary key of our *Main Data* table. The foreign key constraint is set up in a way that easily enables us to update or delete multiple records at once without ever leaving the *Main Data* table.

Main data											Specifics			
ID	Form	Root 1	Suffix 1	Suffix 2	Suffix 3	Prefix 1	Prefix 2	Prefix 3	Prefix 4	Infinitive	ID	mainID	Model	Derivation
79	ispisivati	pis	NULL	iva	ti	is	NULL	NULL	NULL	True	2	79	M7	pisati_inf->ispisati_fin->ispisivati_inf

Examples							
ID	mainID	Sense 1	Sense 2	Example Croatian 1	Example Croatian 2	Example English 1	Example English 2
3	79	write out	print	Kako polaznici budu odgovarali, njihove odgovore možete ispisivati na ploču.	Ako je sadržaj cijelog reda nepoznat ili ga ne treba ispisivati, tada se izostavlja cijeli red.	As students respond, you may want to write their answers on the board.	If the content of a complete line is unknown, or need not to be printed, then the complete line is omitted.

Figure 7. Representation of tables and their data in the database

Furthermore, a search is performed on only one (*Main Data*) table with all the additional information retrieved only upon the user's request. Figure 8 depicts one such instance when the verb form *ispisivati* is selected for search.⁴ The derivational chain [Derivacija: *pisati* -> *ispisati* -> *ispisivati*] holds additional information on aspect that is color-coded (perfective verbs are marked in red, imperfectives in blue) and available via a hover feature.

Search by typing in form, or root with other attributes.

ISPISIVATI

PRETRAGA

PRETRAGA

Prefix 1

Prefix 2

Prefix 3

Prefix 4

Suffix 1

Suffix 2

Suffix 3

Oboje
 Svršeni
 Nesvršeni

Detalji za glagol ispisivati

ispisivati

Derivacija: pisati -> ispisati -> ispisivati

Prefix 1: is **Korijen:** pis **Sufiks 2:** iva **Sufiks 3:** ti

Značenje 1: write out
Značenje 2: print

Primjer 1 hrvatski: Kako polaznici budu odgovarali, njihove odgovore možete ispisivati na ploču.

Primjer 2 hrvatski: Ako je sadržaj cijelog reda nepoznat ili ga ne treba ispisivati, tada se izostavlja cijeli red.

Primjer 3 hrvatski: Ispisati, kopirati ili skenirati jednu sliku.

Primjer 1 engleski: As students respond, you may want to write their answers on the board.

Primjer 2 engleski: If the content of a complete line is unknown, or need not to be printed, then the complete line is omitted.

Primjer 3 engleski: Print, copy, or scan a single image.

Figure 8. Web interface showing additional information on derivation and examples for the verbs *ispisivati*

6 Conclusion

In this paper we showed some preliminary steps taken in the processing of Croatian aspectual pairs. This phase of the project consists of the extension of the existing verb dictionary and its enrichment with the information on verbal aspect. This preliminary step resulted in the significant expansion and improvement of its coverage. The second step that was taken in the preliminary stage was to design and to build the web-based database of aspectual derivatives. The database will enable various types of queries and provide information about affixes used in a specific derivational process, full derivation chains of verbs, basic meaning definitions, contextual examples, etc. Our next goal is to populate the database with aspectual derivatives from other derivational families. The database will remain free for on-line search.

⁴ The web interface is deployed on a Heroku server at <https://vidski-parnjaci.herokuapp.com>.

References

- Tom Coulter. 14.12.2017. *Why MySQL is still King*, Retrieved 22.06.2018. from <https://www.free-lancer.com/community/articles/why-mysql-is-still-king>
- Tomaž Erjavec (ed.). 2001. Specifications and Notation for MULTEXT-East Lexicon Encoding. *MULTEXT-East Report*, Concede Edition D1.1F/Concede, Jožef Stefan Institute, Ljubljana. <http://nl.ijs.si/ME/Vault/V2/msd/msd.pdf>
- Ian Gilfillan 13.06.2015. Database Normalization Overview, *MariaDB Docs*, Retrieved 22.06.2018. from <https://mariadb.com/kb/en/library/database-normalization-overview/>
- Bhojaraju Gunjal. 2003. Database Management: Concepts and Design, In *Proceedings of 24th IASLIC-SIG-2003*. Dehradun: Survey of India. Retrieved 20.06.2018. from https://www.researchgate.net/publication/257298522_Database_Management_Concepts_and_Design
- Nikola Ljubešić and Filip Klubička. 2014. {bs,hr,sr}WaC – Web corpora of Bosnian, Croatian and Serbian, In *Proceedings of the 9th Web as Corpus Workshop (WaC-9) @ EACL 2014*, (eds.) F. Bildhauer & R. Schäfer, Association for Computational Linguistics, pages 29–35, Gothenburg, Sweden.
- Max Silberztein. 2016. *Formalizing Natural Languages: The NooJ Approach*, Cognitive science series, Wiley-ISTE, London, UK.
- Krešimir Šojat, Matea Srebačić, and Marko Tadić. 2012. Derivational and Semantic Relations of Croatian Verbs. *Journal of Language Modelling*, 00 (2012), 1: 111-142.
- Krešimir Šojat, Matea Srebačić, and Vanja Štefanec. 2013. CroDeriV i morfološka raščlamba hrvatskoga glagola, *Suvremena lingvistika*. 39 (2013), 75: 75-96.
- Harald Trost. 2003. Morphology. *The Oxford Handbook of Computational Linguistics* (ed. R. Mitkov), Oxford University Press: 25-47.
- Kristina Vučković. 2009. *Model parsera za hrvatski jezik*. PhD dissertation, Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb.
- Kristina Vučković, Marko Tadić, and Božo Bekavac. 2010. Croatian Language Resources for NooJ. *CIT: Journal of computing and information technology*, 18(2010):295-301.
- Kristina Vučković, Nives Mikelić-Preradović, and Zdravko Dovedan. 2010. Verb Valency Enhanced Croatian Lexicon. In T. Varadi, J. Kuti, M. Silberztein (eds.) *Applications of Finite-State Language Processing*, pages 52-60, Cambridge Scholars Publishing, Newcastle upon Tyne.