# Automatization of Valency Frames Prediction Based on Verb Prefixation with NooJ

**Bajac, Petra**

UNIVERSITY OF ZAGREB
FACULTY OF HUMANITIES AND SOCIAL SCIENCES
DEPARTMENT OF LINGUISTICS
ACADEMIC YEAR 2021/2022

Petra Bajac

# Automatization of Valency Frames Prediction Based on Verb Prefixation with NooJ

Master's thesis

Thesis advisors: dr. sc. Božo Bekavac, Asst. Prof.
dr. sc. Krešimir Šojat, Asst. Prof.

Zagreb, March 2022

# Abstract

By approaching the phenomenon of verbal prefixation in Croatian through the verb *dati* 'to give' and its 18 derivationally related forms, such as *prodati* 'to sell' and *dodati* 'to add', this thesis outlines the impact prefixation can have on the verb's valency frames. By utilizing the NooJ application to create a formal grammar which can detect and tag derivationally related verb forms, and with the help of a model <NP>/<PP> chunker for Croatian, our research aims to start the process of automatization of valency frames prediction and thus to speed up the analysis of prefixed verbs and their valency frames. Our analysis is focused specifically on those verbs which are not currently present in the Croatian valency lexicon CROVALLEX, and as a result, we provide the valency frame descriptions for 10 missing verbs from the *dati* derivational family. Furthermore, this thesis provides a template for the automatized creation of connections between derived and base verb forms.



Keywords: prefix, prefixation, derivation, valency frames, computational lexicons, NooJ



Pristupivši procesu glagolske prefiksacije u hrvatskom jeziku bazirajući se na primjeru glagola *dati* i njegovih 18 derivacijski povezanih oblika, poput *prodati* i *dodati*, ovaj rad prikazuje utjecaj prefiksacije na promjene u valencijskim okvirima derivacijski povezanih glagola. Cilj je ovog rada započeti proces automatizacije izrade valencijskih okvira pa smo tako izradili model formalne gramatike putem aplikacije NooJ koja može pronaći i označiti derivacijski povezane glagole. Uz pomoć modela razdjelnika za imenske fraze i prijedložno-padežne izraze, naš pristup ubrzava proces analize prefigiranih glagola i pripadajućih okvira. Naša se analiza primarno fokusira na one glagole koji trenutno nisu prisutni u CROVALLEX-u, valencijskom leksikonu glagola hrvatskoga jezika, a koji su dio derivacijske mreže glagola *dati*, te smo stoga opisali valencijske okvire 10 prefigiranih glagola koji nedostaju. Također, kao rezultat ovog rada nastao je i mogući predložak za automatsku izradu poveznica među deriviranim i osnovnim oblicima glagola.



Ključne riječi: prefiks, prefiksacija, derivacija, valencijski okviri, računalni leksikoni, NooJ

# Table of contents

**Abstract** ... **2**

**Table of contents** ... **3**

**List of tables** ... **4**

**List of schematics** ... **5**

**1 Introduction** ... **6**

**2 Resources and Tools** ... **8**

2.1 HJP ... 8

2.2 CroDeriV ... 8

2.3 CROVALLEX ... 9

2.4 NooJ ... 11

2.5 The Corpus: SketchEngine and HrWaC ... 11

**3 Theoretical framework** ... **12**

3.1 Meanings of Prefixes ... 12

3.2 The Role of Prefixes in Valency Frames ... 14

3.3 The Prefixation of the Verb *dati* ... 16

**4 Analysis** ... **28**

**5 Results and discussion** ... **35**

5.1 Frame Output for the Missing Verbs ... 35

5.2 Grammar evaluation ... 50

**6 Conclusion** ... **51**

6.1 Conclusions and final remarks ... 51

6.2 Future work ... 52

**7 References** ... **54**

**8 Resources** ... **56**

# List of tables

Table 1: Meanings of prefixes – adapted from Šojat et al. (2012) ..... 13
Table 2: *dati* ..... 17
Table 3: *dodati* ..... 19
Table 4: *predati* ..... 20
Table 5: *prodati* ..... 21
Table 6: *zadati* ..... 22
Table 7: *izdati* ..... 23
Table 8: *odati* ..... 25
Table 9: *nadodati* ..... 26
Table 10: *pridati* ..... 36
Table 11: *pridodati* ..... 36
Table 12: *udati se* ..... 37
Table 13: *udati* ..... 38
Table 14: *poudati se* ..... 39
Table 15: *poudati* ..... 40
Table 16: *preudati se* ..... 41
Table 17: *preudati* ..... 42
Table 18: *podati* ..... 42
Table 19: *podati se* ..... 43
Table 20: *razdati* ..... 44
Table 21: *razdati se* ..... 45
Table 22: *poprodati* ..... 45
Table 23: *preprodati* ..... 46
Table 24: *rasprodati* ..... 48
Table 25: Precision, recall, f-measure ..... 50

# List of schematics

Schematic 1: The main morphological grammar ..... 29

Schematic 2: Syntactic grammar for locating verbs within the derivational family of *dati* ..... 29

Schematic 3: Known dictionary entry tagged with DERIVFAMILY ..... 30

Schematic 4: Unknown dictionary entries tagged with DERIVFAMILY – one prefix ..... 31

Schematic 5: Unknown dictionary entries tagged with DERIVFAMILY – two prefixes ..... 31

Schematic 6: *Poprodati* tags ..... 31

Schematic 7: *Potražiti* tags ..... 32

Schematic 8: Concordance – syntactic grammar ..... 33

Schematic 9: XML-like representation of <NP> and <PP> chunks ..... 34

Schematic 10: *Dati* and its derivatives based on frequency and availability in CROVALLEX .. 51

Schematic 11: *Dati* and the connections between its derivatives ..... 52

# 1 Introduction

Today, it is important for a language to have extensive and high-quality computational resources such as corpora, dictionaries, and various types of computational lexicons. Building these resources manually can be very time-consuming, difficult, and, in turn, expensive. Therefore, developing tools which could aid in the making and/or improvement of such resources is an extremely important task for computational linguists, especially for languages such as Croatian, which have very underdeveloped language technologies (Tadić et al. 2012). Croatian is a morphologically rich language, not only in terms of inflection, but also in terms of derivation. For example, prefixation is the most productive derivational process in Croatian when it comes to verbs (Babić 2002). Furthermore, it seems that prefixation can not only change the meaning of some verbs in Croatian, it can also change some of their syntactic properties as well, such as their valency frames – which is the main focus of this paper. Naturally, we touch upon some lexical properties as well, such as the semantic selection which determines the semantic content of arguments of the verbs, since the two are closely connected. Unfortunately, not much research has been done to describe the effect of prefixation on verbs in Croatian, however, this thesis offers a short overview of the existing research in the field of Valency theory, namely, the work done by Mikelić Preradović, supported by some remarks from the field of derivational morphology and syntax, researched by Šojat and his colleagues. The results presented in this thesis could hopefully provide other researchers with a new and faster approach to study prefixation and its influence on valency frames.

The purpose of this paper is to start the development of an automatization tool for prediction of verb valency frames influenced by prefixes. Primarily, the aim is to start building a NooJ[1] grammar which will be able to automatically extract valency frames of prefixed verbs. We model the grammar based on the verb *dati* 'to give', covering its base form and its prefixed forms such as *rasprodati* 'to sell everything', *preprodati* 'to resell' etc. This approach is also beneficial for comparisons of changes in meanings and subcategorization frames between verbs derived from the same base form, as opposed to studying the effects of one particular prefix. To conclude, this grammar could potentially serve as a rule-based template appropriate for any verb in Croatian to extract derivational families, as a part of a larger statistical tool which would automatically

[1] NooJ – rule-based NLP tool - for more information – see 2.4

determine the valency frames by using large amounts of data, which, in turn, would need to be annotated in a way presented in this thesis. This would simplify the creation of lexicons such as CROVALLEX[2], and in turn, make future theoretical research into the intricacies of the process of prefixation easier as well, as all the needed data could then be gathered much faster.[3]

This thesis is divided into several sections, starting with *2 Resources and Tools*, where we enumerate and explain the roles of various lexicons and dictionaries as well as other tools used in this thesis. This section also outlines some theoretical aspects upon which those resources and/or tools are built, which is important in order to understand the theory presented in the following section *3 Theoretical Framework*. There we present the existing research on verb prefixation in general, as well as its influences on verb valency frames. Building on the theoretical outlines presented in these two sections, we give an overview of the prefixation of the verb *dati,* as well as its valency frames from CROVALLEX, which aims to explain our motivation and goals. After that, in section *4 Analysis*, we describe the functionalities and outputs of our NooJ grammars. Next comes *5 Results and discussion,* where we expand the analysis of the prefixation of the verb *dati* by providing new valency frames for derivatives which have not yet been covered in CROVALLEX. We also evaluate our grammar in terms of precision, recall and f-measure. Finally, in *6 Conclusion*, we give some final remarks and describe next possible steps in our automatization efforts.

[2] CROVALLEX - Croatian Valency Lexicon of Verbs - for more information – see 2.3

[3] Briscoe and Carroll's (1997) efforts in automatic extraction of subcategorization from corpora also demonstrated that such tools can improve other tools and resources by a significant amount.

# 2 Resources and Tools

As stated in *1 Introduction*, the goal of this thesis is to start building a grammar which could detect verb valency frames of prefixed verbs[4]. Since the ultimate goal is to create systematic connections between derived verbs and their base form following the available theoretical research presented in *3 Theoretical framework*, the starting point is a grammar that automatically creates verb families of prefixed verbs which share a base form. A complete overview of the grammar is presented in *4 Analysis*, and the output and final results of the analysis of valency frames are presented in *5 Results and discussion*. Before diving any deeper, it is prudent to first and foremost present an overview of all the tools and resources used to demonstrate the theory and develop our analysis in the following sections, and elaborate on their input and use in this thesis.

## 2.1 HJP

HJP, or *Hrvatski Jezični Portal* (*Croatian Language Portal*), is a joint project of the Croatian publisher *Znanje* and *SRCE*, University Computer Center, University of Zagreb. It is the first and currently the only dictionary database for Croatian available online. It is based on a dictionary database of 6 lexicographic resources for Croatian[5]. This electronic database provided the meanings and definitions which were relevant for the distinctions between the senses of most of the derived verbs, and the base form *dati* 'to give', as well. It also provided us with the etymology behind the verb *nadati se* 'to hope' which helped us determine that the connection between that verb and one of the derived forms of *dati – nadati se* 'to obtrude' is that of homonymy, and not polysemy, which in turn led us to exclude the verb from our analysis for practical reasons. Homonymous items create great difficulties in gathering a representative corpus since the search would provide us with numerous false positives, i.e. verbs which are not part of our analysis, and hence extensive manual revision would be needed, such as it would be the case if we wanted to include the verb *nadati se* (see 3.3).

## 2.2 CroDeriV

CroDeriV (Šojat, Srebačić & Štefanec 2013) is a Croatian morphological database describing 14 491 Croatian verbs. It segments the verbs into lexical and derivational morphemes

[4] Šojat et al. (2010a) describe valency frames of Croatian verbs of consumption in a similar manner.
[5] https://hjp.znanje.hr/index.php?show=baza, accessed on May 15th, 2020

and thus enables users to see derivational families either connected by the base verb, or any prefix/suffix. On top of the segmentation, the metadata also provides information about reflexivity and aspect[6]. The morphological segmentation and the information about aspect included in CroDeriV are used in the description of verbs covered in this thesis, as seen in Tables 2-19. The root of the verbs we are concerned with here is tagged as *da1* in CroDeriV, and the search yields numerous results, 18 of which were recognized as verbs derived by prefixation form the base form *dati 'to give'*. The database also contains verbs derived by suffixation, and the verbs derived by prefixation of those suffixed forms. As will be explained in more detail in *3 Theoretical framework*, those examples are not covered in this thesis.

## 2.3 CROVALLEX

CROVALLEX (Mikelić Preradović 2008) is a valency lexicon of Croatian verbs, developed as part of the PhD thesis by Nives Mikellić Preradović, titled *Approaches to the Development of the Machine Lexicon for Croatian Language*. She used the Functional Generative Description (FGD) to formalize the description of valency frames. The lexicon contains around 1740 verbs, and each verb entry contains one or more valency frame descriptions, each corresponding to a different sense. The lexicon served as a primary source for the retrieval of valency data for the verbs in this thesis, however, many derived forms we were looking for are missing from the lexicon. To be more precise, 10 derived forms do not have a formalized description of their valency frames in CROVALLEX.

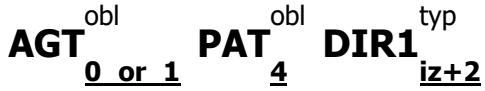
As mentioned previously, CROVALLEX presents multiple senses of verbs with their corresponding frames under one entry, or lemma. The valence frame is structured to include three attributes:

1) functor
2) list of morphemic realizations
3) type of complementation.

The documentation available provides us with the following explanation[7]: functors are divided into two types, inner participants (AGT agent, PAT patient, REC recipient, RESL result, ORIG

[6] Šojat et al. (2016) deal with semi-automatic extraction and inclusion of verb valency frames in CroDeriV as well, however, their results have not been included in the current version of the database.
[7] http://theta.ffzg.hr/crovallex/doc/crovallex-doc.html, accessed on May 31st, 2020

origin) and numerous types of free modifications, such as LOC location, DIR3 direction-to, INST instrument etc. Morphemic realizations include explicit and implicit forms. For example, the pure prepositionless cases are numbered as 1 - nominative, 2 - genitive, 3 - dative, 4 - accusative, 5 - vocative, 6 - locative, and 7 - instrumental. Regarding the type of complementation, CROVALLEX differentiates between two main types: obligatory (obl), and typical, but optional (typ). These can be attributed to both inner participants and free modifications. The typical arguments can be both inner participants and free modifications. The non-obligatory free modifications are considered to be typical when they are related to some verbs, or whole classes of verbs, but not to others, and are hence deemed to be relevant enough to be included in the frame. This is a slight modification to the syntactic theory used to build the description, the FGD, which would only include free modifications in the frame description if they were obligatory. To demonstrate on a concrete example, let us examine the following annotation for the one of the senses of verb *izvaditi* 'to remove', as seen in CROVALLEX:




Example: *Vikinške brodove su izvadili iz mora.*
Translation: 'They removed the viking ships from the sea.'[8]

In this example, the CROVALLEX entry specifies that the sentence contains an AGENT (AGT), an obligatory element (obl), which can either be realized as a null-subject in a nominative case (0)[9], such as the case here, or as an explicit subject in a nominative case (1). Furthermore, we find a PATIENT (PAT), also an obligatory element (obl), realized as an accusative case (4), and a typical element (typ) denoting a DIRECTION-FROM (DIR1), realized as a combination of the preposition *iz* and the genitive case (2). To conclude, all valency frames examples presented in this thesis follow this same representation, both original, copied examples in 3.3 and novel ones in 5.1.

[8] CROVALLEX does not include English translations of Croatian examples in its entries, so for the purposes of this thesis, all sentences in Croatian are translated into English for better clarification.
[9] As English is not a null-subject language and does not permit omission of subjects, the translation includes the corresponding pronoun which was lacking in the original. Such instances can often be encountered in other examples and translations throughout this thesis.

## 2.4 NooJ

NooJ (Silberztein 2018) is a computational tool created by Max Silberztein, which is used to process text and corpora at any linguistic level needed. It has the power to combine morphological and syntactic grammars, which is essential for this kind of research where we look into valency frames (syntax) of derivationally related verbs (morphology). NooJ was the primary NLP tool used in the creation of the grammar for this thesis, and a more in-depth description of the features relevant for this research is presented in *4 Analysis*.

## 2.5 The Corpus: SketchEngine and HrWaC

The valency frames of our verbs were extracted from a small corpus collected from the Croatian Web Corpus HrWaC (Ljubešić & Klubička 2016). This corpus is a web corpus, collecting texts from the .hr domain and contains 1.9 billion tokens. The corpus was accessed through the SketchEngine tool[10], and a random sample of 100 sentences was taken for each of the verbs missing from CROVALLEX (*pridati* 'to assign [features]', *udati* '[women] to marry', *podati* 'to give in', *razdati* 'to dole out', *poprodati* 'to sell one by one', *poudati* '[women] to marry one by one', *preprodati* 'to resell', *preudati* '[women] to remarry', *pridodati* 'to add to another' and *rasprodati* 'to sell everything') However, some verbs have less than a 100 of total concordances in the entire HrWaC, with *poudati* '[women] to marry one by one' having 83 total concordances, and *poprodati* 'to sell one by one' having merely 13. This adds to a total of 896 sentences in our small corpus for these 10 verbs, amounting to a total of 38 265 tokens. As with any corpus, we expected certain problems caused by numerous homographs in the Croatian language, as the POS taggers can often yield faulty tags when dealing with ambiguous items.

[10] https://www.sketchengine.eu/, accessed on April 15th, 2020

# 3 Theoretical framework

As stated in *1 Introduction*, prefixation is the most productive derivational process when it comes to verbs, not only in Croatian, but other languages report the same as well, for example, Czech (Hrstková 2007). This section outlines some of the observed effects it can have on verbs in Croatian, namely their meanings, aspect, and valency frames. Meanings of individual, productive prefixes are given as a starting point to see how the meanings of prefixes interact with the meaning of the base form they attach to, which is illustrated through the verb *dati 'to give'* and its relationship with its observed derived forms.

## 3.1 Meanings of Prefixes

In Croatian, according to Šojat et al. (2012), one can find 19 productive prefixes[11]: *do-, iz-, na-, nad-, o-/ob-, obez-, od-, po-, pod-, pre-, pred-, pri-, pro-, raz-, s-, su-, u-, uz-,* and *za-* (118). The authors state that most of these prefixes retained the meaning of prepositions which they developed from. They go on to note that when combined with stems, some prefixes create entirely compositional combinations, where the two meanings are simply added together, creating a very transparent new meaning, while others create highly idiosyncratic combinations, where the new meaning cannot easily be deduced without resorting to analysis of metaphorical extensions or metonymical shifts.

Focusing on mostly compositional combinations, Šojat et al. (2012) formed four major groups of morphosemantic relations via prefixes between verbs, based on categories of prefixal meanings in Babić (2002) and Barić et al. (2003):

1) Location
2) Time
3) Quantity
4) Manner.

[11] The number varies from author to author when it comes to verbs – for example, Babić (2002) enumerates 28 different suffixes which are productive in the derivation of verbs, 5 out of which are vaguely productive (namely *de-, dis-, pred-, re-* and *su-*) and 4 which are unproductive (*mimo-, naj-, pa-, protu- and suprot-*), while the rest he considers to be highly productive. The difference in the number between the two authors stems from the fact that Šojat et al. (2012) consider only the listed 19 prefixes to be productive, and the rest are deemed unproductive, as well as not taking into consideration any prefixes of foreign origin.

They also establish a table of all 19 prefixes and their meanings within the four groups[12,13], as shown in Table 1.

Table 1: Meanings of prefixes – adapted from Šojat et al. (2012)[14]

| PREFIX | LOCATION | TIME | QUANTITY | MANNER |
|---|---|---|---|---|
| ***do-*** | to/toward | completion<br>finiteness | addition | |
| ***iz-*** | bottom-up<br>from | distributivity<br>completion | sufficiency<br>excessiveness | |
| *na-* | top-down<br>proximity<br>to/toward | inchoativity<br>distributivity | sufficiency<br>excessiveness<br>intensity<br>addition | |
| *nad-* | over | | exceeding | |
| ***o-/ob-*** | around | | | |
| *obez-* | | | deprivation | |
| *od-* | apart | completion | | |
| ***po-*** | top-down | inchoativity<br>distributivity | intensity | |
| *pod-* | under | | insufficiency | |
| ***pre-*** | over<br>re-location | completion | intensity<br>exceeding | change of property |
| *pred-* | | preceding | | |
| ***pri-*** | proximity<br>to/toward | | intensity<br>addition | connection |
| ***pro-*** | through<br>proximity | inchoativity<br>completion<br>preceding | intensity | |
| ***raz-*** | apart | | intensity | |
| *s-* | top-down | | | connection |

[12] The prefixes which can be added to the verb *dati* have been put into bold for easier and quicker cross-referencing in 3.3 and 5.1

[13] Even though the prefix *na-* can be added to *dati*, creating the verb *nadati se* 'to obtrude' it has been excluded from the analysis for reasons explained in 2.1 and 3.3

[14] Examples provided in the original table have been excluded in order to simplify the whole representation, and the column titles are added once more on the following page to improve readability.

| PREFIX | LOCATION | TIME | QUANTITY | MANNER |
|---|---|---|---|---|
| *su-* | proximity | | | connection<br>opposition |
| ***u-*** | into | finiteness | intensity | change of property |
| *uz-* | proximity | inchoativity | intensity | |
| ***za-*** | around<br>behind<br>to/toward<br>top-down | inchoativity | intensity | change of property |

What is clear from the table, since almost all prefixes denote at least one meaning in the *Location* domain, is that prefixes predominately denote spatial relations due to their prepositional origin (Šojat et al. 2012). Based on this, the majority of new senses added to base verbs are expected to correspond to some of the meanings in the *Location* domain[15].

## 3.2 The Role of Prefixes in Valency Frames

The impact prefixes have on verbs in contemporary Croatian varies. Šojat et al. (2012) enumerate three ways in which prefixes can alter the make-up of the Croatian verb:

1) Prefixes can cause a change in aspect
2) Prefixes can cause a change in aspect while adding a new semantic component to the verb
3) Prefixes can cause a change in meaning without a change in aspect.

They continue to note that prefixation of imperfective verbs can create both imperfective and perfective forms, but perfective verbs can only create perfective forms when added a prefix. This is because in Croatian, suffixation is used to derive imperfective forms from perfective verbs. The creation of aspectual pairs via prefixation and/or suffixation is not covered in this thesis[16], as our main example, the verb *dati* 'to give' is a perfective verb, and by adding prefixes, we only generate other perfective forms. Instead, the focus is on the lexical changes the prefixes impose on the verb *dati* and their consequence on its syntactic properties, i.e. valency frames. In this sense, prefixes have much more impact on meaning than suffixes, which are "used to form verbs denoting

[15] Metaphorical and metonymical shifts involved in processes such as these are presented in Lakoff (1987), Langacker (1987), Raffaelli (2007) and others. Šarić (2003, 2006a,b) and Belaj (2008a,b) focus specifically on Croatian prefixes and prepositions within the framework of cognitive linguistics.
[16] An extensive overview of verb valency changes can be found in Dixon & Aikhenvald (2000).

diminutive actions and pejorative attitudes" (Šojat et al. 2012, 117). The changes prefixes cause in Croatian verbs are "less predictable" than those caused by suffixes and "vary in terms of meaning from compositional to completely idiosyncratic" (Šojat et al. 2012, 117-118). For example, we can compare the influence prefixes *do-* and *u-* have on our verb *dati*. The former creates a transparent new meaning of "to add" by combining the original meaning of *dati*, "to give", and the meaning of addition in terms of quantity of *do-* (see Table 1). However, when we look at the latter, such a clear and straightforward analysis of meaning cannot be made when we look at *udati* "to marry".

Furthermore, Mikelić Preradović et al. (2013) suggest that different verbs with the same prefix often share the same valency frames, due to the similar added meaning, even though the valency frames of their respective base forms may be different. They state that "the effect of […] prefixes on the argument structure of verbs in Croatian exhibits regularity" and that "verbs sharing the same prefix (more precisely, the same meaning of the specific prefix) also share the same valency frames and the number of obligatory and typical optional complements"(37). They continue to conclude that there is a close connection between verb valency properties and their semantics. For example, they argue, that the typical complement DIR-1 (direction-from), syntactically realized as a combination of prepositions *iz*, *s*, and *od*, + the genitive case, can be expected to appear with verbs coming from the syntactic-semantic classes of *disappear*, *remove*, *push* etc. Based on this observation, they conclude that prefixes which share the same meaning as the aforementioned prepositions, *iz-*, *s-*, and *od-,* could also "[serve] as criteria for establishing a reasonably consistent syntactic-semantic class" (39).

While Mikelić Preradović et al. (2013) put the focus on the connection between multiple verbs through one specific prefix for the advancement of valency frames in CROVALLEX, Šojat et al. (2012) aim to enrich CroWN by researching the morphosemantic relations between derived forms of one single verb. This thesis takes a hybrid approach, focusing on a single verb and its derived forms, but trying to see if there are any changes in the valency frames of the derivatives, and if so, what are those changes, which added meanings contributed to the change, and if there are any predictable patterns at play. The verb *dati* is chosen as it allows numerous combinations and serves as an interesting case study for the connections between morphology, syntax and semantics, as explained in the following section(s).

## 3.3 The Prefixation of the Verb *dati*

According to Šojat et al. (2012), *dati* 'to give' is marked for 28 different senses in CroWN. This means that it is a highly polysemous verb and an exhaustive analysis of all its senses would be beyond the scope of this thesis and could be the topic of an entirely separate project, so in this section we give a short overview of only some of its senses, namely only those present in HJP, which should be sufficient to illustrate the motivation behind the creation of our grammar. The first entry in HJP was selected as the basic meaning of the verb, and each sense is presented alongside their valency frames extracted from CROVALLEX.

Based on the results from CroDeriV[17], there are 18 derived forms from the base verb *dati* 'to give', allowing for prefixation with 10 different prefixes: *do-, pre-, pri-, pro-, u-, za-, po-, raz-, iz, o-*. As seen in the results in CroDeriV, the verb also allows for combinations of up to two different prefixes: *na-+do-, po-+pro-, po-+u-, pre-+pro-, pre-+u-* and *pri-+do-*. Table 1 in 3.1 can be used to cross-reference the influence of the prefix on the meanings of the verbs by comparing the predicted added meaning of the prefix with the stem's original meaning without it, if, of course, the resulting meaning is (relatively) compositional.

The examples in Tables 2-9 below consist of morphological data and aspect information (*Derivational Analysi*s, *Aspect*) taken from CroDeriV[18], and syntactic data (*Valency Frames*) as well as example sentences in Croatian from CROVALLEX[19,20]. Furthermore, the base meaning of the verbs, and the meanings of their different senses have been translated based on entries in HJP[21], which were, on occasion, edited for clarification if the corresponding sense presented in CROVALLEX is worded differently. It is important to note that the tables below only show forms which have valency frame data available in CROVALLEX. The forms which are not present in the lexicon are covered in the later section 5.1, with the added valency frame data extracted as part of this research. These verbs are: *pridati* 'to assign [features]', *udati* '[women] to marry', *podati* 'to give in', *razdati* 'to dole out', *poprodati* 'to sell one by one', *poudati* '[women] to marry one by one', *preprodati* 'to resell', *preudati* '[women] to remarry', *pridodati* 'to add to another' and

[17] http://croderiv.ffzg.hr/Croderiv, accessed on May 17th, 2020
[18] Šojat et al. (2013, 2014) describe the morphological analysis presented in CroDeriV in more detail.
[19] http://theta.ffzg.hr/crovallex/data/html/generated/alphabet/index.html, accessed on May 21st, 2020
[20] The sentences in Croatian which serve as examples here have been translated into English for the purposes of this thesis and are not originally present in the lexicon.
[21] https://hjp.znanje.hr/index.php?show=baza, accessed on May 15th, 2020

*rasprodati* ‘to sell everything’. As shortly mentioned in 2.1, one particular derivative is purposefully not covered by this research either, the verb *nadati se* ‘to obtrude’. This is because it is a homonym to the verb *nadati se* ‘to hope’ and as such would create problems with the initial corpus search and research corpus creation, and subsequently the creation of the grammar, as well. Even though this could be dealt with, for the sake of simplicity and saving time, in this phase of the project, it was decided to simply exclude it.

As this research is primarily focused on prefixation of the base verb *dati* ‘to give’ and the aim of the grammar is not to search for aspectual pairs[22], the suffixed forms of *dati* and its derivates, such as *davati* ‘to give, iteratively’ or *prodavati* ‘to sell, iteratively’ are not included.

Since all Croatian verbs take on the infinitive suffix, either *-ti* or its allomorph *-ći*, and since this has no impact on our demonstration of connections between the verbs based on their prefixes, the derivational analysis does not delineate the infinitive suffix, which is the same for all the forms covered: *-ti*.[23] To demonstrate the connections between the verbs, let us consider Tables 2-9:

Table 2: *dati*

| **Form** | *dati* | **Meaning** | ‘to give’ |
|---|---|---|---|
| **Derivational analysis** | dati | **Aspect** | perfective |
| **Valency frames:** | | | |
| 1. to hand over | $\mathbf{AGT}^{\text{obl}}_{\mathbf{0\_or\_1}}$ $\mathbf{PAT}^{\text{obl}}_{\mathbf{4}}$ $\mathbf{REC}^{\text{typ}}_{\mathbf{3}}$ | | |
| EXAMPLE: | *Britanci su taj otok dali Amerikancima.* | | |
| TRANSLATION: | ‘The British handed the island over to the Americans.’ | | |
| 2. to do | $\mathbf{AGT}^{\text{obl}}_{\mathbf{0\_or\_1}}$ $\mathbf{PAT}^{\text{obl}}_{\mathbf{4}}$ $\mathbf{MANN}^{\text{typ}}_{\mathbf{kao+1}}$ | | |
| EXAMPLE: | *Ti si bila dala poljubac kao anđeo.* | | |
| TRANSLATION: | ‘You kissed like an angel.’[24] | | |

[22] In contrast, in their automatization efforts, Šojat et al. (2016) aim to include both perfective and imperfective forms of verbs involved in their research.
[23] This is a slight deviation from the morphological model which is used in CroDeriV.
[24] A more direct translation: ‘gave a kiss’

| | |
|---|---|
| 3. to produce, bring as a final result | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **TWHEN**$^{typ}_{u+6}$ |
| EXAMPLE: | *Obitelj je u XVII. st. dala misionare, pripadnike reda Družbe.* |
| TRANSLATION: | 'In the 17th century, the family produced missionaries, members of the order of the Society.' |
| 4. not to resist something | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{inf}$ |
| EXAMPLE: | *Dala se portretirati.* |
| TRANSLATION: | 'She let herself be portrayed.' |
| 5. to offer | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ |
| EXAMPLE: | *Iskaz su dala dva svjedoka.* |
| TRANSLATION: | 'Two witnesses offered a testimony.' |
| 6. to put on a show[25] | **AGT**$^{obl}_{0\_or\_1}$ **LOC**$^{typ}_{na+6}$ |
| EXAMPLE: | *Predstavu su dali na Gundulićevoj poljani.* |
| TRANSLATION: | 'They put on a show on Gundulić's field.' |
| 7. to entrust someone to do something[26] | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{inf-4}$ **LOC**$^{typ}_{na+6}$ |
| EXAMPLE: | *Na lijevom ramenu je prije godinu dana dala istetovirati malog dupina.* |
| TRANSLATION: | 'She had a small dolphin tattoed on her left shoulder a year ago.' |
| 8. to give into someone's influence[27] | **AGT**$^{obl}_{0\_or\_1}$ **REC**$^{obl}_{3}$ |
| EXAMPLE: | *Nismo se dali nikome.* |
| TRANSLATION: | 'We never gave in to anyone.' |

[25] For this valency frame and example, CROVALLEX doesn't provide a PAT argument, although it is explicitly stated in the example sentence that *a show* has been given. We think this to be a mistake in analysis.
[26] In this example, a WHEN modifier is not provided even though it is present in the example sentence. This may not necessarily be a mistake as it was probably deemed not specific or relevant enough for this sense.
[27] In this example, as in many others, CROVALLEX does not express the deep analysis behind reflexive verbs, where the reflexive pronoun *se* expresses the unity of AGENT and PATIENT arguments, but rather keeps to the surface level representation by providing only an AGENT within the valency frame.

| | |
|---|---|
| 9. to give money[28] | **AGT**$^{obl}_{0\_or\_1}$ **EXT**$^{obl}_{indeclinabilia\_2}$ **AIM**$^{typ}_{za+4}$ **TWHEN**$^{typ}_{adv}$ |
| EXAMPLE: | *Ove godine smo dali 12 milijuna kuna za uništavanje komaraca.* |
| TRANSLATION: | 'This year, we gave 12 million Croatian kunas for the eradication of mosquitoes.' |

The base verb *dati* has 8 different entries in HJP, some of which have one or two separate sub-entries, with a total of 12 meanings recorded in the dictionary. On top of that, it has a rich phraseology, with 22 phrases or sayings enumerated, such as *dati krv* 'to donate blood' and *dati krila* 'to encourage', literally 'to give wings'. *Dati* is also a very frequent verb, with more than a million occurrences in HrWaC[29]. In the table, we see extensive overlapping of obligatory elements of valency frames between most of the meanings.

Table 3: *dodati*

| **Form** | *dodati* | **Meaning** | 'to add' |
|---|---|---|---|
| **Derivational analysis** | do+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to add, to pass by throwing | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **REC**$^{typ}_{3}$ | | |
| EXAMPLE: | *Vranješ je dodao loptu Deranji.* | | |
| TRANSLATION: | 'Vranješ passed the ball to Deranja.' | | |
| 2. to add to a conversation, to say | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{sub\text{-}da}$ | | |
| EXAMPLE: | *Vilmos Sazbo dodao je da uspješni odnosi dviju stranaka mogu pridonjijeti i suradnji dviju država.* | | |

[28] It is important to note that it is very peculiar to have an EXT argument without a PAT or a REC, however, this sentence truly does reflect such a case. What is more peculiar is that CROVALLEX does not offer a frame for a common combination of AGT + REC + EXT, such as in a very simple and ordinary sentence *Dala je sestri 12 kuna* 'She gave her sister 12 kunas.'

[29] HrWaC 2.2 - Croatian Web Corpus - for more information, see 2.5

| | |
|---|---|
| TRANSLATION: | ‘Vilmos Sazbo added that successful relationships between two parties can contribute to the cooperation between two countries, as well.’ |

| | |
|---|---|
| 3. to enlarge by a certain amount, to combine | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **REC**$^{typ}_{3}$ **TWHEN**$^{typ}_{adv\text{-}ovih\_dana}$ |
| EXAMPLE: | *Pisanju drama ovih ste dana dodali i pisanje libreta.* |
| TRANSLATION: | ‘These days, you added writing librettos to writing Drama.’ |

*Dodati* is the most frequent derived form of *dati*, with 259 572 occurrences in HrWaC. It is not as highly polysemous, with only three different senses recorded in HJP – ‘to add [in terms of calculation]’, ‘to pass something over to someone’, and ‘to add something in a conversation’. We can see that one of the meanings of this verb is also the same as 1 – ‘to pass something over’ or ‘to hand something over’. Based on this example, it is logical to confirm the hypothesis that meanings and valency frames of verbs belonging to the same derivational family share a connection. Other valency frames present only a change in typical elements.

Table 4: *predati*

| **Form** | *predati* | **Meaning** | ‘to hand sth. over’ |
|---|---|---|---|
| **Derivational analysis** | pre+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to hand something over, to deliver | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **REC**$^{typ}_{3}$ | | |
| EXAMPLE: | *Ugledni veterinar je predao pismo japanskome caru.* | | |
| TRANSLATION: | ‘A respected vet handed a letter to the Japanese emperor.’ | | |
| 2. to accept defeat, to surrender[30] | **AGT**$^{obl}_{0\_or\_1}$ **REC**$^{typ}_{3}$ **TWHEN**$^{typ}_{adv\text{-}jučer}$ | | |

[30] As in previous examples, the PAT is not expressed for this reflexive form.

| | |
|---|---|
| EXAMPLE: | *Dvadesetogodišnji mladić D.G. predao se jučer zagrebačkoj policiji.* |
| TRANSLATION: | 'Yesterday, a twenty-year-old young man D.G. gave himself up to the Zagreb police.' |
| 3. to accept enemy's victory | **AGT**$^{obl}_{0\ or\ 1}$ **PAT**$^{obl}_{4}$ **REC**$^{typ}_{3}$ **REG**$^{typ}_{prema+6}$ |
| EXAMPLE: | *Borković je prema ratnim pravilima predao grad Šljivančaninu.* |
| TRANSLATION: | 'Per rules of war, Borković handed the city over to Šljivančanin.' |
| 4. to forfeit | **AGT**$^{obl}_{0\ or\ 1}$ **PAT**$^{obl}_{4}$ **CAUS**$^{typ}_{zbog+2}$ |
| EXAMPLE: | *Američki tenisač je predao meč zbog ozljede desnog zgloba.* |
| TRANSLATION: | 'The American tennis player forfeited the match due to an injury to his right wrist.' |
| 5. to give into something[31] | **AGT**$^{obl}_{0\ or\ 1}$ **REC**$^{obl}_{3}$ **ACMP**$^{typ}_{sa+7}$ |
| EXAMPLE: | *Djevojka se predala užicima sa svojim izabranikom.* |
| TRANSLATION: | 'The girl indulged in pleasures with her chosen one.' |
| 6. to extradite, hand over prisoners | **AGT**$^{obl}_{0\ or\ 1}$ **PAT**$^{obl}_{4}$ **REC**$^{typ}_{3}$ **LOC**$^{typ}_{u+6}$ |
| EXAMPLE: | *Young je zarobljenike predao kontinentalnim agentima u Lewesu.* |
| TRANSLATION: | 'Young extradited the prisoners to the continental agents in Lewes.' |

Just as the previous verb, *predati* also shares its basic meaning and valency frame with 1. However, the senses vary in terms of added typical complements. Interestingly, the web of closely related meanings of 'surrender' have different valency frames even between each other and share only certain obligatory complements with their base form.

Table 5: *prodati*

| **Form** | *prodati* | **Meaning** | 'to sell' |
|---|---|---|---|
| **Derivational analysis** | pro+dati | **Aspect** | perfective |
| **Valency frames** | | | |

[31] In this example we also have a missing representation of the PAT with the reflexive form.

| | |
|---|---|
| 1. to sell property in exchange for money | **AGT**$^{\text{obl}}_{\text{0 or 1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **EXT**$^{\text{typ}}_{\text{za+indeclinabilia 2}}$ |
| EXAMPLE: | *Ljahnicky je prodao stan u Crnčićevoj ulici za 310.000 DEM.* |
| TRANSLATION: | 'Ljahnicky sold the apartment on Crnčićeva street for 310 000 DEM.' |

| | |
|---|---|
| 2. to switch sides for money, to betray your own[32] | **AGT**$^{\text{obl}}_{\text{0 or 1}}$ **REC**$^{\text{typ}}_{\text{3}}$ |
| EXAMPLE: | *Nisam se prodao Austrijancima.* |
| TRANSLATION: | 'I didn't sell myself to the Austrians.' |
| 3. to exchange goods with someone | **AGT**$^{\text{obl}}_{\text{0 or 1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **REC**$^{\text{typ}}_{\text{3}}$ |
| EXAMPLE: | *Seljaci su prodali svoju pšenicu Čakovečkim mlinovima.* |
| TRANSLATION: | 'The peasants sold their wheat to the Čakovec Mills.' |

When we look at the derivative *prodati*, we see that the shift in meaning is a bit greater than with the previous two verbs. However, upon comparing the valency frame to the ones of *dati*, we can see the conceptual connection between meaning 9 and therefore a shared valency frame is not surprising. Moreover, the meaning of 'betray' shares the same valency frame as the corresponding meaning of *izdati* in Table 7 also suggesting a close connection between semantics and syntax.

Table 6: *zadati*

| **Form** | *zadati* | **Meaning** | 'to assign, give a task' |
|---|---|---|---|
| **Derivational analysis** | za+dati | **Aspect** | perfective |

[32] Another case of a missing PAT in reflexive verbs.

| **Valency frames** | |
|---|---|
| 1. to order, to assign, to give a task | AGT$^{obl}_{0\ or\ 1}$ PAT$^{obl}_{4}$ REC$^{typ}_{3}$ TWHEN$^{typ}_{prilikom+2}$ |
| EXAMPLE: | *Gradonačelnik Milan Bandić je prilikom obilaska gradilišta izvođačima radova zadao novi rok za završetak radova na gradnji podvožnjaka na križanju Škorpikove ulice i Ljubljanske avenije* |
| TRANSLATION: | 'During the tour of the construction site, Mayor Milan Bandić gave the contractors a new deadline for the completion of works on the construction of the underpass at the intersection of Škorpikova Street and Ljubljanska Avenue.' |
| 2. to do harm | AGT$^{obl}_{0\ or\ 1}$ PAT$^{obl}_{4}$ REC$^{typ}_{3}$ |
| EXAMPLE: | *Galbraith je zadao udarac haaškom tužiteljstvu.* |
| TRANSLATION: | 'Galbraith dealt a blow to the Hague prosecution. [metaphorically]' |

*Zadati* seems to have an added typical complement denoting time, but it shares the basic valency frame with meaning 1 of *dati*. It is not difficult to see the connection here, as this verb is specific for its connection with words such as task, or problem, so we can see the partial synonymy between *dati zadatak* 'to give a task' and *zadati zadatak* 'to assign a task', confirmed by 537 and 333 concordances in HrWaC for the phrases, respectively. Interestingly, the other meaning, of 'to do harm', in a sense to *zadati udarac* 'to strike someone, literally: to deliver a blow' also shares a conceptual connection to the meanings of the base form, and therefore, also the valency frame.

Table 7: *izdati*

| **Form** | *izdati* | **Meaning** | 'give away (a secret), to betray ' |
|---|---|---|---|
| **Derivational analysis** | iz+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to reveal (a secret) | AGT$^{obl}_{0\ or\ 1}$ PAT$^{obl}_{4}$ | | |

| | |
|---|---|
| EXAMPLE: | *Izdala je tajnu svoje gospodarice.* |
| TRANSLATION: | 'She revealed her mistress' secret.' |
| 2. to publish | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ |
| EXAMPLE: | *John Greenberg iz Berkeleya izdao je 21 Salingerovu priču.* |
| TRANSLATION: | 'John Greenberg from Berkeley published 21 Salinger's stories.' |
| 3. to rent something to someone | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **REC**$^{typ}_{3}$ |
| EXAMPLE: | *Ona izdaje stan onom šuftu i niškoristi Paku.* |
| TRANSLATION: | 'She is renting to that loser, good-for-nothing Pak.' |
| 4. to abandon someone and switch sides | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ |
| EXAMPLE: | *On je izdao Hrvatsku.* |
| TRANSLATION: | 'He betrayed Croatia.' |
| 5. to betray someone for something | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **AIM**$^{typ}_{za+4}$ |
| EXAMPLE: | *Za sultanov dar izdao je svoju braću i prijatelje.* |
| TRANSLATION: | 'He betrayed his brothers and friends for a sultan's gift.' |
| 6. to decide on a legal matter | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ |
| EXAMPLE: | *Sud je izdao nalog za premetačinu stanova.* |
| TRANSLATION: | 'The court issued a search warrant.' |
| 7. to reveal yourself or your secret by doing something[33] | **AGT**$^{obl}_{0\_or\_1}$ **MANN**$^{typ}_{7}$ |
| EXAMPLE: | *Izdao se svakom kretnjom svojom, cijelim bićem svojim.* |
| TRANSLATION: | 'He revealed himself with every move he made, with his every being.' |

[33] Here we also have a missing PAT, as the AGENT indeed revealed someone, in this case, himself.

| | |
|---|---|
| 8. to issue something in accordance with the law | **AGT** (obl) **0 or 1** **PAT** (obl) **4** |
| EXAMPLE: | *MUP je izdao potvrdu o preuzimanju tijela.* |
| TRANSLATION: | 'The Ministry of Internal Affairs issued a certificate of transfer of body remains.' |

*Izdati* has 8 separate entries in CROVALLEX and 3 separate entries in HJP. However, 5 of those 8 entries have the same valency frame, AGENT + PATIENT, and the rest present different typical complements, such as RECIPIENT, AIM or MANNER. However, it must be pointed out that in some of the examples of valency frames with only an AGENT and a PATIENT, the same sentence can often appear with a RECIPIENT, too, which may not be obvious from the examples CROVALLEX offers. For example, when something is issued, it can almost always also be issued *to someone*. Based on all this, we can conclude that even though the metaphorical shift behind the meanings is not always transparent, there are no significant changes to the core of the valency frame, merely an additional typical complement may be present.

Table 8: *odati*

| **Form** | *odati* | **Meaning** | 'to pay (respects), to disclose' |
|---|---|---|---|
| **Derivational analysis** | o+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to pay respects | **AGT** (obl) **0 or 1** **PAT** (obl) **4** **BEN** (typ) **3** **INST** (typ) **7** | | |
| EXAMPLE: | *Đapić je u Osijeku odao počast palima za NDH minutom šutnje.* | | |
| TRANSLATION: | 'In Osijek, Đapić paid tribute to the fallen for the NDH with a minute of silence.' | | |
| 2. to indulge oneself[34] | **AGT** (obl) **0 or 1** **PAT** (obl) **3** | | |
| EXAMPLE: | *Posljednjih mjeseci odao se alkoholu.* | | |
| TRANSLATION: | 'In recent months, he has indulged himself with alcohol.' | | |

[34] Another case of a reflexive verb not being analyzed as an AGT + PAT combination.

| | |
|---|---|
| 3. to disclose something confidential | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **BEN**$^{typ}_{3}$ |
| EXAMPLE: | *Anti Markoviću i Slobodanu Miloševiću odao je tajnu o hrvatskim financijskim mogućnostima kupnje oružja.* |
| TRANSLATION: | 'He disclosed a secret about Croatia's financial capabilities to buy weapons to Ante Marković and Slobodan Milošević.' |
| 4. to clumsily reveal your own secret[35] | **AGT**$^{obl}_{0\_or\_1}$ |
| EXAMPLE: | *Djevojčica se nije odala.* |
| TRANSLATION: | 'The girl didn't reveal herself.' |

*Odati* seems to be closely connected to some meanings with *izdati* with the meaning 'to betray' and to *predati* with the meaning 'to give into something'. Let us compare, for example, the meanings of *odati* 'to clumsily reveal a secret' and *izdati* 'to reveal yourself or your secret by doing something'. The only difference in meaning, and therefore in the valency frame, is that *izdati* may contain an expressed manner in which a person reveals themselves. The frames of 'give into something' in verbs *odati* and *predati* are interestingly completely different, despite the similarity in meaning.

Table 9: *nadodati*

| | | | |
|---|---|---|---|
| **Form** | *nadodati* | **Meaning** | 'to add extra' |
| **Derivational analysis** | na+do+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to add, to enlarge | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ | | |
| EXAMPLE: | *Urednik je nadodao dva posljednja retka.* | | |
| TRANSLATION: | 'The editor added the last two rows.' | | |
| 2. to add something extra while saying something | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{sub\text{-}da}$ | | |

[35] Same as the previous.

| | |
|---|---|
| EXAMPLE: | *Gucićeva braniteljica nadodala je da nema ni saznanja da su otac i sin Gucić izručeni.* |
| TRANSLATION: | 'Gucić's defense attorney added that she did not even know that father and son Gucić had been extradited.' |

The verb *nadodati* has two prefixes, and has been created from another derived form, which is why it is fitting to only compare it to that form – *dodati* – rather than the base form directly. The added meaning of *extra*, regarding a conversation, as opposed to just *add*, does not result in differences in the type of arguments between the meaning 'to add to a conversation' in *dodati*, nor their realization. However, the other valency frame is not present in *dodati*, and as both elements are obligatory, we can conclude that this change in the frame is relatively significant.

To conclude, some connections between meanings, and therefore valency frames, are relatively transparent, expected, and straightforward. Others show a more nuanced differentiation. This is not surprising, though, as it is expected that every prefix will have something to add to the meaning of base form, and if the end result is different enough, the valency frame reflects it as well. It is quite interesting that most changes can be seen in the typical elements, suggesting that the power prefixes have over valency frames is relatively limited.

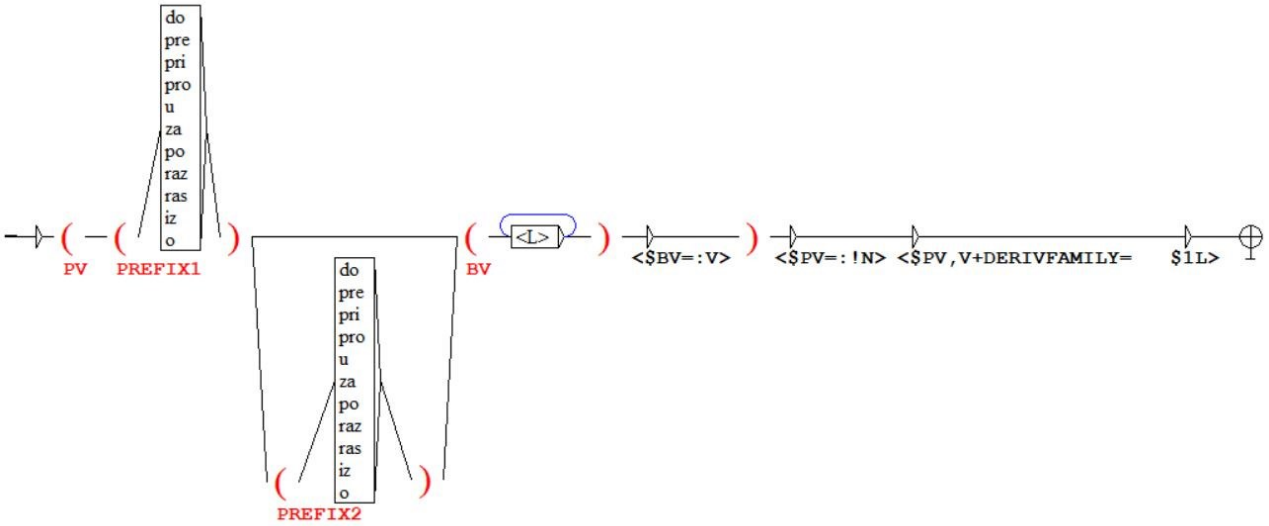
# 4 Analysis

This section provides an overview of our NooJ grammars, its functionalities, and its output.

The first and the main component of our research is a morphological grammar whose purpose is to detect and tag derivationally related verbs. Again, we note that the relationships are based only on prefixation, and verbs related to each other via suffixation are not meant to be tagged with this grammar.

As we based our preliminary research on the verb *dati*, whose derivatives only take up to two prefixes, the grammar includes batches of prefixes within variables PREFIX1 and PREFIX2. The number of these batches can easily be increased, and the list of prefixes within it can also be extended. The list currently includes only prefixes which have been found on 10 derivatives which are not described in CROVALLEX. Note how there are 11 prefixes on the list, not 10. This is because the prefix *raz-* can have two allomorphs: *raz-* and *ras-*. This is due to the effect of assimilation between voiced and voiceless consonants, which presents itself in the Croatian orthography[36].

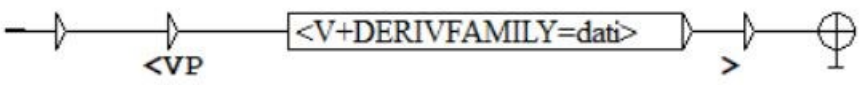
[36] In this case, the principle states that the consonant /z/ becomes /s/ before another voiceless consonant – such as in the verb *rasprodati*, where /z/ comes in contact with /p/ and assimilates into /s/.




Schematic 1: The main morphological grammar

The grammar in *Schematic 1* utilizes a few constraints in order to get the most accurate results as possible. Firstly, it makes sure that the BV (base verb) variable is a verb, and the <L> (lemma) node makes sure the grammar works universally, not only for one specific verb which could have been added manually. Secondly, it demands that the whole PV (prefixed verb) variable is not a noun to carry out the necessary disambiguation. Finally, the grammar assigns a special tag to any verb that fits the grammar's criteria: V+DERIVFAMILY. It also includes information about which base they are related to. In this case, the tag for our verbs is V+DERIVFAMILY=dati. This makes it possible to automatically create derivational families around a common base verb, which is precisely our goal.

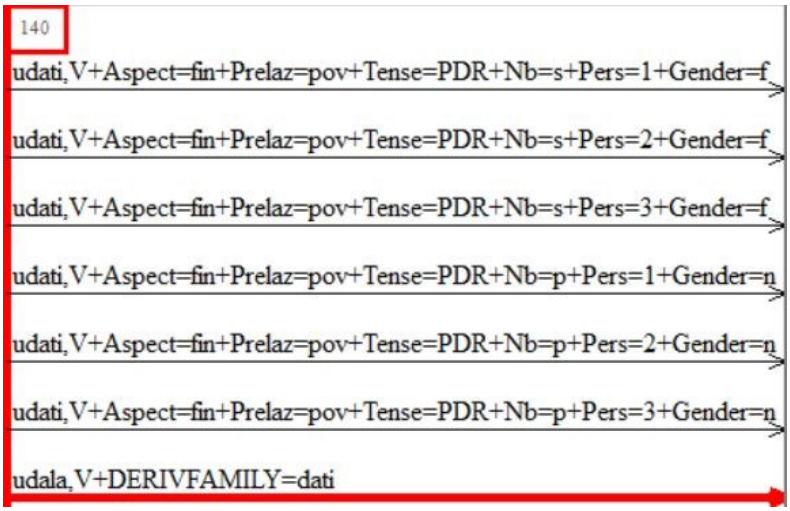
As the corpus, naturally, contains many verbs derived not only from the verb *dati*, but others as well, we needed a way to clean up the concordances to include only our target verbs, those related to the verb *dati*. We utilize the syntactic grammar shown in Schematic 2 to do that.




Schematic 2: Syntactic grammar for locating verbs within the derivational family of *dati*

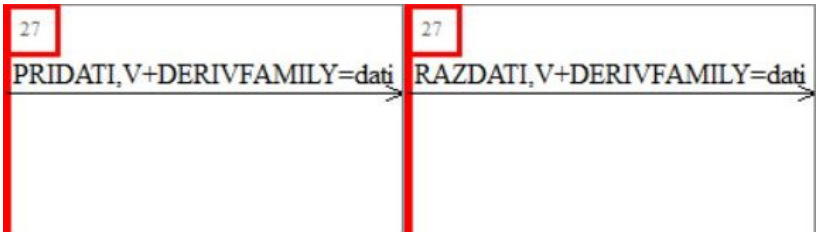
After the corpus is processed with our main morphological grammar, the verbs are tagged accordingly. The syntactic grammar shown in Schematic 2 utilizes those tags to select only the verbs tagged with V+DERIVFAMILY=dati. The grammar can be modified to look for any other verb instead of *dati*, making it quite versatile. Moreover, the grammar automatically tags the verbs as <VP> which would be important later when the <NP>/<PP> chunker is introduced to aid in valency frames recognition.

Let us now consider the efficiency of the two grammars in Schematics 1 and 2 and the results they yield. When we look at Schematic 3, we are presented with an example of a verb which had been present in the NooJ dictionary before. So, the verb contains two sets of tags, the original tag, and our added tag seen in the last row.



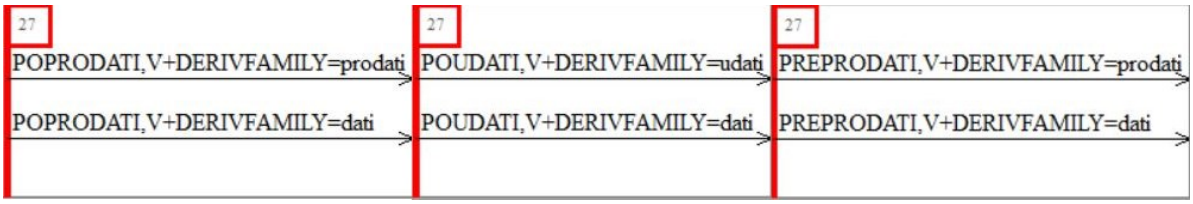

Schematic 3: Known dictionary entry tagged with DERIVFAMILY

When dealing with unknown verbs, we only get our added tag as shown in Schematic 4. After our main morphological grammar analyzes a corpus, it is of no significance whether or not the verb was previously known to NooJ, it can nevertheless be found and tagged. In Schematic 4, we see what the tag looks like on unknown verbs with one prefix.



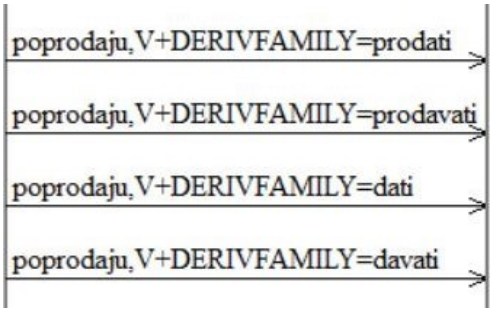
Schematic 4: Unknown dictionary entries tagged with DERIVFAMILY – one prefix

A very interesting thing happens when we examine the tags of verbs with two prefixes. Those verbs contain not only the tag of the initial base verb *dati*, but also the verb which we can classify as its derivational predecessor. This is extremely useful because verbs with two or more prefixes are more closely related to their predecessors than the base verb, as the second prefix modifies the meaning of an already prefixed verb, and not the base directly. Therefore, these tags are an excellent way to create a sort of a derivational network which would clarify the path the verb took during the prefixation process from the base form to the final form. We see this in Schematic 5 below, where the verbs *poprodati*, *poudati* and *preprodati* have annotated connections to both the base and their predecessors, *prodati*, *udati* and *prodati*, respectively.




Schematic 5: Unknown dictionary entries tagged with DERIVFAMILY – two prefixes

As an unintentional coincidence, our grammar seems to even create connections to suffixed verbs within the family. The verbs are indeed connected, however, as noted, this was purely unintentional as the focus was never on the suffixed forms of the base verb.

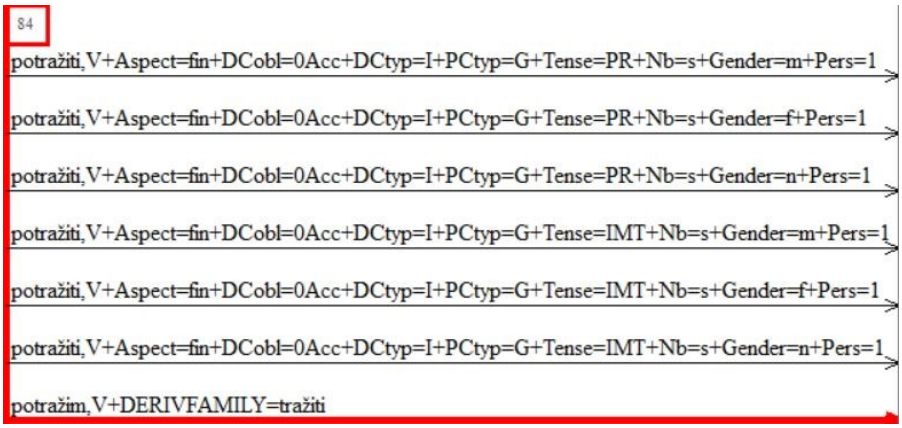



Schematic 6: *Poprodati* tags

This is, however, probably the result of overlaps within the paradigms of *poprodati* (perfective) and *poprodavati* (imperfective). When we check both paradigms in the Croatian Morphological

Lexicon[37], we see that the form *poprodaju* is present as 3rd person plural under the lemma *poprodavati,* as well as *poprodati*. We can conclude that the same overlap is present within NooJ, too.

The grammar was created with a certain dose of universality in mind. Instead of focusing on merely the lemma *dati*, we instructed the grammar to look for any lemma in the dictionary, which is also a verb. Therefore, the grammar also correctly tagged verbs of other origin, such as the one shown in Schematic 7.




Schematic 7: *Potražiti* tags

*Potražiti* 'to look up' is also a prefixed verb, its base being the verb *tražiti* 'to look for, to seek'. The tag shown in Schematic 7 illustrates the ability of the grammar to tag any verb whose base form is present in the NooJ dictionary.

Finally, let us look at the concordance lines retrieved by the syntax grammar. In Schematic 8, we see the generated <VP> tags:

[37] Hrvatski morfološki leksikon - http://hml.ffzg.hr/hml/ - a lexical database which generates word-forms from its repository of lemmas, accessed on May 31st, 2020

| | | |
|---|---|---|
| nastao s izgubljenim d od | poda/<VP> | i nosa Nos do poda |
| poda i nosa Nos do | poda/<VP> | sigurno nema veze s ponosom |
| nastao s izgubljenim d od | poda/<VP> | i nosa Nos do poda |
| poda i nosa Nos do | poda/<VP> | sigurno nema veze s ponosom |
| slike presvučena od stropa do | poda/<VP> | plemenitim drvom, razne radne i |
| se proširio zgradom nakon prekopavanja | poda/<VP> | kupaonice sve govori, a to |
| si i spustio auto do | poda/<VP> | da se ništa ne njiše |
| kad stisnes gas pedalu do | poda/<VP> | , izraz cujes cesto kad se |
| size: 13 query: Query:[lempos='(?i)( | POPRODATI/<VP> | )-v'] unizg.hr\| vinograde još |
| setim na ' gospoštinu '. Inače ispiju, | poprodaju/<VP> | i naliju vode. - Ali čemu |
| izbavio. Čekaj, dok to siromaštvo | poprodam/<VP> | , povratit ću ti pošteno novce |
| slavoljubpenkala.hr\| sve svoje obligacije | prodao/<VP> | bio, plijenio nemilo do zadnje |
| ljudi krvavo znojili, koješta založili, | poprodali/<VP> | , samo da se riješe pijavice |
| evo, npr. recimo da Linić | poproda/<VP> | sve živo i odjebe te |
| je jako važno da se | poprodaju/<VP> | oni silni prazni objekti na |
| Grad na škole poslao i | poprodao/<VP> | svu zemlju da bi ga |
| dobio prvenstvo 2003. i nakon toga | poprodao/<VP> | sve igrace. Covjeka koji, kao |
| samo radi relativno brze zarade | poproda/<VP> | da će opet krenut prema |
| Šta je vrijedilo - već je | prodano/<VP> | . Da poprodamo otoke i obalu |
| vrijedilo - već je prodano. Da | poprodamo/<VP> | otoke i obalu, dobili bi |
| u sustini dionice drzavnih poduzeca | poprodali/<VP> | strancima, a tvrke ostale kakve |
| vidjeli rast cijena sve su | poprodali/<VP> | , a oni koji nisu stigli |
| ce se cijene ' vratiti ', hahahaha | Pogledajmo/<VP> | vecernji.hr\| a u životu |
| a na kraju, kad sve | poprodaju/<VP> | , možda i iz 1. HKL. To |
| koliko je on popljackanih stvari | poprodao/<VP> | - on i NJEGOVA familija VSG |
| size: 83 query: Query:[lempos='(?i)( | POUDATI/<VP> | )-v'] tportal.hr\| nego skroz |
| nam se frendica i frendova | poudalo/<VP> | i poženilo, tko se čak |
| a mnogi su se već | poudali/<VP> | i poženili. S nekima se |
| (iako su mu djeca odrasla, | poudala/<VP> | se i poženila...). Á izio |
| sam ti se već odavno | poudala/<VP> | , evo četvrto na putu, a |
| vam se djeca požene i | poudaju/<VP> | .To su ta famozna obiteljska |
| se ' mnogi sretno poženili i | poudali/<VP> | '... Proslava 25 godina splitske diskoteke Master |
| tradiciji, ' mnogi sretno poženili i | poudali/<VP> | '. Glazbeni dio, pak, otvorio je |

Schematic 8: Concordance – syntactic grammar

However, we must emphasize the number of concordance lines yielded by the search: 1008. Our corpus was designed to have 896 instances of our target verbs, based on 896 retrieved sentences/concordance lines from HrWaC. This leads us to conclude that there may be some false positives present here and/or verbs which are also part of the *dati* derivational family, for example, the verb *prodati* found itself as part of a sentence within the concordance of *podati*. We see this clearly in Schematic 8, where we find words such as *poda,* the genitive singular form of the noun *pod* 'the floor', which is a homograph to the 3rd person singular form of the verb *podati*, hence the false positives. This was an expected problem even in the beginning stages while compiling the corpus itself, as the POS taggers used both by HrWaC and NooJ are not fool-proof and disambiguation is needed. The calculation of the number of true/false positives - the recall, precision and f-measure are provided in 5.2.

The final stage of this part of the research involved implementing a <NP>/<PP> chunker (Vučković et al. 2008b) which would speed up the process of determining the best candidates for constituents around the main verb and therefore aiding the process of describing the valency frames

of the missing verbs. The chunker was created using the NooJ tool and its primary purpose is to separate sentences into NP and PP chunks. The <VP> annotation, however, is provided by our syntactic grammar (Schematic 2). Schematic 9 provides an example of the adapted annotation of the sentence *Sumnjiči ga se da je na račun duga svoje tvrtke jednoj zagrebačkoj tvrtki preprodao dva vozila* 'He is suspected to have resold two vehicles to one company in Zagreb at the expense of his company's debt.', extracted from our corpus.

```
<SENTENCE>
        <VP> Sumnjiči <VP>
        <NP> ga </NP>
        <PRO> se </PRO>
        <C> da </C>
        <VP>je </VP>
        <PP> na
                <NP> račun duga svoje tvrtke </NP>
        </PP>
        <NP> jednoj zagrebačkoj tvrtki </NP>
        <VP>preprodao </VP>
        <NP> dva vozila
                <PP> u
                        <NP>vlasništvu Hypo leasinga</NP>
                </PP>
        </NP>.
</SENTENCE>
```

Schematic 9: XML-like representation of <NP> and <PP> chunks

The chunker provided us with the nearest structures to the VP[38], which are presumed to be the best candidates for its arguments. However, at this point, the final syntactic analysis was done manually as this version of the chunker is not capable of differentiating between obligatory and optional and/or typical arguments, and if those arguments took the form of a clause, as is often the case, problems arose. Also, the final stages of valency frames extractions based on CROVALLEX's model involve a semantic analysis, determining the semantic roles for each argument[39]. This part of the job was also done manually.

To conclude, this section illustrated the automatic process of locating and tagging the desired verbs, making the morphological level completely automated. The analysis on the syntactic level is partially aided by automation, while the semantic level remains completely manual. The

[38] The structures in question have been bolded in the XML-like representation for clarity.
[39] These two issues were, in fact, dealt with by Vučković et al. (2008a), however, at this stage in our project, these solutions were not yet utilized.

results and a more detailed account of syntactic and semantic analyses are presented in the next section.

# 5 Results and discussion

With the help of the chunker and our grammar, the analysis of the valency frames was significantly sped up. The observed valency frames are presented in tables 10-19, exactly like those in 3.3, but with added examples from HrWaC and their translations. The decision to include the examples here, and not in 3.3 lays in the simple fact that the examples of the verbs in 3.3 can be easily found in CROVALLEX, and since the verbs below are not available there, the inclusion of the examples was, naturally, a necessary part of valency frame analysis.

## 5.1 Frame Output for the Missing Verbs

The frame output of our verbs was intended to be as consistent as possible with the methodology used in CROVALLEX, so that they can be directly transferred into its database, with perhaps only slight revision of the formatting. Meanings were translated from entries in HJP, when possible. It is important to note, even though it may go without saying, that the translations of

examples to English often lose the similarity to the original valency frame/tense/word order etc. and the translations only serve the purpose of clarifying the meanings of the sentences. The tables include the structured information about the valency frames, and they are also accompanied by short comments regarding any additional insights and/or problems with the analysis.

Table 10: *pridati*

| **Form** | *pridati* | **Meaning** | 'to assign [features]' |
|---|---|---|---|
| **Derivational analysis** | pri+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to assign [features] | obl obl obl<br>**AGT PAT REC**<br>**0 or 1 4 3** | | |
| EXAMPLE: | *Sve više kupaca namještaja pridaju vrijednost tome da namještaj, ne samo dobro izgleda, nego da je i praktičan.* | | |
| TRANSLATION: | 'More and more furniture buyers assign value to the furniture not only for looking nice, but also for being practical.' | | |

This verb, alongside some others described in this section, does not have its own entry in HJP - we are directed to the entry of its imperfective counterpart to see the meaning. The difference between the two really is just in their perfectiveness, so we saw no problem in including the meaning provided by HJP. Interestingly, there were some rogue verbs present within the concordance, and we noticed that they are dialectalisms of the verb *predati se* 'to surrender' in the form of *pridati se*. These were, naturally, disregarded. The difference between reflexive and non-reflexive forms of this verb does not play any role in the semantic or syntactic properties[40]. Furthermore, when it comes to the influence of the prefix on the meaning of the verb, the metaphorical shift and the connection with the base verb is visible, but the meaning is not entirely compositional.

Table 11: *pridodati*

[40] In most cases, the reflexive properties of a verb in Croatian influence the overtness of arguments such as PATIENT or AGENT. However, there are some verbs which have different meanings and different valency frames between their reflexive and non-reflexive forms, as seen with the verb *podati*/*podati se* in Tables 15 and 16.

| **Form** | *pridodati* | **Meaning** | 'to add to another' |
|---|---|---|---|
| **Derivational analysis** | pri+do+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to add to another | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **REC**$^{typ}_{3}$ | | |
| EXAMPLE: | *Fizičkoj dimenziji mi smo pridodali duhovnu.* | | |
| TRANSLATION: | 'We added a spiritual dimension to the physical one.' | | |
| 2. to add to a conversation, to say | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{sub\text{-}kako}$ | | |
| EXAMPLE: | *Valja pridodati kako čvorno pitanje programske sinergije bitno ovisi o izboru tipa društva.* | | |
| TRANSLATION: | 'It should be added that the main question of program synergy highly depends on the choice of company.' | | |

*Pridodati* shares the same valency frame with *pridati* in its first meaning. The second meaning share approximately the same meaning and the same valency frame with *nadodati* and *dodati*.[41] HJP, however, only lists the first meaning in its entry, but we strongly believe that the second one should be included as well, as the new sense is clearly visible and presents itself with a different valency frame as well.

Table 12: *udati se*

| **Form** | *udati se* | **Meaning** | '[women] to marry' |
|---|---|---|---|
| **Derivational analysis** | u+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. [women] to marry | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ | | |
| EXAMPLE: | *E sad ti je vrijeme da se udaš.* | | |
| TRANSLATION: | 'Now is the time for you to get married.' | | |

[41] The difference is in the realization of the subordinate clause, starting with either *kako* or *da*, which are in this case interchangeable (Silić & Pranjković 2007, 333)

| | |
|---|---|
| | **AGT**$^{\text{obl}}_{\text{0 or 1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **BEN**$^{\text{typ}}_{\text{za+4}}$ |
| EXAMPLE: | *[…] Keto koja se udala za vlastitog brata.* |
| TRANSLATION: | '[…] Keto, who married her own brother.' |
| | **AGT**$^{\text{obl}}_{\text{0 or 1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **DIR3**$^{\text{typ}}_{\text{u+4}}$ |
| EXAMPLE: | *Druga se udala u Italiju.* |
| TRANSLATION: | 'The other one married (and moved) (in)to Italy.' |
| | **AGT**$^{\text{obl}}_{\text{0 or 1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **TWHEN**$^{\text{typ}}_{\text{u+3}}$ |
| EXAMPLE: | *U međuvremenu sam se udala.* |
| TRANSLATION: | 'In the meantime, I got married.' |
| | **AGT**$^{\text{obl}}_{\text{0 or 1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **CAUS**$^{\text{typ}}_{\text{zbog+2}}$ |
| EXAMPLE: | *Svi pričaju da sam se udala zbog para.* |
| TRANSLATION: | 'Everyone is saying that I got married because of money. ' |

| | |
|---|---|
| | **AGT**$^{\text{obl}}_{\text{0 or 1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **BEN**$^{\text{obl}}_{\text{za+4}}$ **TWHEN**$^{\text{typ}}_{\text{u+3}}$ |
| EXAMPLE: | *[…] za kojeg se udala u siječnju prošle godine.* |
| TRANSLATION: | '[…] whom she married in January last year.' |

Table 13: *udati*

| **Form** | *udati* | **Meaning** | 'to marry [a woman]' |
|---|---|---|---|
| **Derivational analysis** | u+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to marry a woman away | **AGT**$^{\text{obl}}_{\text{0 or 1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **BEN**$^{\text{obl}}_{\text{za+4}}$ | | |

| | |
|---|---|
| EXAMPLE: | *A zle sestre budu kažnjene tako što ih udaju za kuhara ili stolara.* |
| TRANSLATION: | 'And evil sisters are punished by being married away to a chef or a carpenter.' |

We mentioned earlier that CROVALLEX does not differentiate between reflexive and non-reflexive verb forms and does not assign a PAT argument to those reflexive verbs. We decided to take a different approach. The verb *udati* carries two different meanings, both explicit in HJP. The first meaning is closely tied to the reflexive properties of *udati*, as a woman is presumably giving *herself* away, and the second meaning is not reflexive, requiring another person to give the woman away. For this reason, even when the PAT is only indicated by the reflexive pronoun *se*, we decided to include it in the valency frame, as well as to separate the corresponding analyses into two lemmas. Since CROVALLEX does not specify any special type of annotation for reflexive pronouns, we kept the annotation for a prepositionless accusative case - 4. Regarding the concordances of this verb, there was observed homography between the 3rd person plural form *udaju* and the accusative singular form of the noun *udaja* 'marriage'[42]. It would be interesting to go deeper into the etymology and the choice behind the verb *dati* as a base verb for this meaning, which would reveal certain cultural conventions of the Croatian people. On a short note, the conceptual metaphor behind *udati* could go along the lines of WOMEN ARE PROPERTY[43]. However, this is, of course, a topic for separate, more thorough research.

Table 14: *poudati se*

| **Form** | *poudati se* | **Meaning** | '[women] to marry one by one' |
|---|---|---|---|
| **Derivational analysis** | po+u+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. [women] to marry one after another | **AGT**$^{\text{obl}}_{\text{0 or 1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ | | |
| EXAMPLE: | *[…] koliko nam se frendica i frendova poudalo i poženilo.* | | |
| TRANSLATION: | '[…] how many of our friends married.' | | |
| | obl obl typ | | |

[42] Croatian differentiates between the female and male versions of words connected to getting married, having two different forms for verbs, nouns, and adjectives. The male forms are *oženiti* (v), *ženidba* (n) and *oženjen* (adj).
[43] For more about these types of metaphors see Lakoff & Johnson (1981).

| | |
|---|---|
| | **AGT**$_{\text{0 or 1}}$ **PAT**$_{\text{4}}$ **BEN**$_{\text{za+4}}$ |
| EXAMPLE: | *[…] se sve zgodne cure poudaju za pomorce.* |
| TRANSLATION: | '[…] all the pretty girls marry sailors.' |
| | **AGT**$^{\text{obl}}_{\text{0_or_1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **MANN**$^{\text{typ}}_{\text{adv-sretno}}$ |
| EXAMPLE: | *Svi su se ranije i sretnije poudali/poženili od mene.* |
| TRANSLATION: | 'Everyone married sooner and more happily than me.' |
| | **AGT**$^{\text{obl}}_{\text{0_or_1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **TWHEN**$^{\text{typ}}_{\text{adv-odavno}}$ |
| EXAMPLE: | *Ja sam ti se već odavno poudala.* |
| TRANSLATION: | 'I married a long time ago.' |
| | **AGT**$^{\text{obl}}_{\text{0 or 1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **DIR3**$^{\text{typ}}_{\text{u+4}}$ |
| EXAMPLE: | *[…] al se zato sve Dalmatinke poudaju u Zagreb.* |
| TRANSLATION: | '[…] but that's why all the Dalmatian girls marry (and move) (in)to Zagreb.' |

Table 15: *poudati*

| **Form** | *poudati* | **Meaning** | 'to marry [women] one by one' |
|---|---|---|---|
| **Derivational analysis** | po+u+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 2. To marry women one after another | **AGT**$^{\text{obl}}_{\text{0 or 1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **BEN**$^{\text{obl}}_{\text{za+4}}$ | | |
| EXAMPLE: | *Sve svoje prijateljice sam poudao za svoje muške prijatelje.* | | |
| TRANSLATION: | 'I married all of my (female) friends to my male friends.' | | |

The derivative *poudati* follows the same pattern as *udati*. Its (non)reflexiveness has the same effects on the valency as with *udati* '[women] to marry', so it too was separated into two

different lemmas. It does not have any HJP entries whatsoever, but the added meaning is transparent – to marry one by one/one after another. Also, the sentences in which this verb occurs were riddled with emphatic dative constructions, and as these serve very little syntactic purpose but rather have only pragmatic implications, the dative was not included in the valency frames.

Table 16: *preudati se*

| **Form** | *preudati se* | **Meaning** | '[women] to remarry' |
|---|---|---|---|
| **Derivational analysis** | pre+u+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to remarry | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **BEN**$^{obl}_{za+4}$ **TWHEN**$^{typ}_{adv}$ | | |
| EXAMPLE: | *Majka se ubrzo preudala za glumca.* | | |
| TRANSLATION: | 'Mother soon remarried an actor.' | | |
| | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **BEN**$^{obl}_{za+4}$ | | |
| EXAMPLE: | *Svježa informacija o tome kako sam se preudala za mlađeg Srbina sigurno ne bi pomogla.* | | |
| TRANSLATION: | 'Fresh information on how I remarried a younger Serb would definitely not help.' | | |

| | |
|---|---|
| | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ |
| EXAMPLE: | *Basaut kaže svima da želi da se Gehna preuda.* |
| TRANSLATION: | 'Basaut is telling everyone that she wants Gehna to remarry.' |
| | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **DIR3**$^{typ}_{u+4}$ |
| EXAMPLE: | *[…] a kad joj se majka preudala u Čulete.* |
| TRANSLATION: | '[…] and when her mother remarried (and moved) (in)to Čulete.' |
| | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **TWHEN**$^{typ}_{adv\text{-}kasnije}$ |
| EXAMPLE: | *Žena se kasnije preudala i ima još jednu kćer.* |

| | |
|---|---|
| TRANSLATION: | 'The woman later remarried and has another daughter.' |
| | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **BEN**$^{obl}_{za+4}$ **MANN**$^{typ}_{adv\text{-}sretno}$ |
| EXAMPLE: | *Sen se sretno preudala za novoimenovanog gospodara Himeji-jo.* |
| TRANSLATION: | 'Sen happily remarried a newly appointed lord Himeji-jo.' |

Table 17: *preudati*

| **Form** | *preudati* | **Meaning** | 'to remarry [a woman]' |
|---|---|---|---|
| **Derivational analysis** | pre+u+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 2. to remarry [a woman] | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **BEN**$^{obl}_{za+4}$ | | |
| EXAMPLE: | *Njezin otac ju je preudao za bogatog trgovca svile.* | | |
| TRANSLATION: | 'Her father remarried her to a rich silk trader.' | | |

*Preudati* has the same reflexive/non-reflexive differentiation as *udati* and *poudati* and shares most of its valency frames with them, so the same remarks apply to it, too. This group of verbs seems to be very closely related both in terms of syntax and semantics. The concordance of this verb was littered with spelling mistakes of the noun *presuda* 'verdict' in the form of *preuda*, spelled the same as 3rd person singular of *preudati*. This clearly illustrates some of the problems with using a web corpus, as opposed to a corpus of the standard where there would be fewer mistakes such as these, if any.

Table 18: *podati*

| **Form** | *podati* | **Meaning** | 'to give, distribute' |
|---|---|---|---|
| **Derivational analysis** | po+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to give | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **AIM**$^{typ}_{sub\text{-}kako}$ | | |

| | |
|---|---|
| EXAMPLE: | *Sva gradska krv podala bi svoje granule kako bi srce bolesnika ponovno sretno zakucalo.* |
| TRANSLATION: | 'All the urban blood would give their granules so that the hearts of the sick would beat happily again.' |
| | **AGT**$^{obl}_{0\ or\ 1}$ **PAT**$^{obl}_{4}$ **REC**$^{typ}_{3}$ |
| EXAMPLE: | *Usliši milostivo molitve naroda svoga i podaj svoj mir vremenima našim.* |
| TRANSLATION: | 'Hear mercifully the prayers of your people and give your peace to our times.' |

Since HJP provides us with two separate lemmas for the non-reflexive and reflexive forms of the verb, we separated the two analyses which show the motivation behind this separation:

Table 19: *podati se*

| **Form** | *podati se* | **Meaning** | 'to give in' |
|---|---|---|---|
| **Derivational analysis** | po+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to give into something | **AGT**$^{obl}_{0\ or\ 1}$ **PAT**$^{obl}_{4}$ **REC**$^{obl}_{3}$ | | |
| EXAMPLE: | *Podao se stvarima, kakove jesu da jesu.* | | |
| TRANSLATION: | 'He gave in to things, whatever they may be.' | | |

| | |
|---|---|
| 2. to have sexual intercourse | **AGT**$^{obl}_{0\ or\ 1}$ **PAT**$^{obl}_{4}$ **REC**$^{obl}_{3}$ **AIM**$^{typ}_{sub\text{-}da}$ |
| EXAMPLE: | *[…] a ti se podala mladim atenskim ratnicima da bi ostala na stazi slobode.* |
| TRANSLATION: | 'And you gave yourself to the young Athenian warriors to stay on the path of freedom.' |
| | **AGT**$^{obl}_{0\ or\ 1}$ **PAT**$^{obl}_{4}$ **REC**$^{obl}_{3}$ |
| EXAMPLE: | *Ne troši vrijeme na one bez love Zakon velikih brojeva, podaj se njima 100 i jedan će upaliti.* |
| TRANSLATION: | 'Do not waste time on those without money, the Law of Large Numbers, give yourself to a hundred of them and one will pay off.' |

*Podati* has considerable differences in meaning between its non-reflexive form and reflexive form *podati se*. For this reason, we believe that CROVALLEX, and other resources in general, should explicitly note reflexivity as a feature in cases when it can have a great impact not only on the syntactic properties of a verb, but also semantic and pragmatic features as well, or even separate the words into two different lemmas, as we have done in the previous examples. Moreover, there were many observed homographs with this verb, both with the genitive form of the noun *pod* 'the floor' and the preposition *poda* 'underneath'. Furthermore, the combination of the prefix and the non-reflexive base verb in meaning 1 is very compositional and the added meaning is minimal, but with the reflexive form, the meaning is significantly changed, and this is evident in the different valency frames as well.

Table 20: *razdati*

| **Form** | *razdati* | **Meaning** | 'to dole out' |
|---|---|---|---|
| **Derivational analysis** | raz+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to dole out | **AGT**$^{\text{obl}}_{\text{0\_or\_1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ | | |
| EXAMPLE: | *Ništa sam i kad bi razdao sav svoj imetak […]* | | |
| TRANSLATION: | 'I am nothing and if I doled out all my possession […]' | | |
| | | | |
| | **AGT**$^{\text{obl}}_{\text{0\_or\_1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **REC**$^{\text{typ}}_{\text{3}}$ | | |
| EXAMPLE: | *Kada bi netko razdao sva dobra siromašnima […]* | | |
| TRANSLATION: | 'If someone would dole out all the goods to the poor […]' | | |
| | **AGT**$^{\text{obl}}_{\text{0\_or\_1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **AIM**$^{\text{typ}}_{\text{sub-kako}}$ | | |
| EXAMPLE: | *Razdaj sve kako bi ti se ruke ispunile.* | | |
| TRANSLATION: | 'Dole out everything so that your hands would be fulfilled.' | | |

| | |
|---|---|
| | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **REC**$^{typ}_{3}$ **MANN**$^{typ}_{adv\text{-}nesebično}$ |
| EXAMPLE: | *Sve što su bili i imali nesebično su drugima razdali.* |
| TRANSLATION: | 'Everything they were and everything they had they selflessly doled out to others.' |

Table 21: *razdati se*

| **Form** | *razdati se* | **Meaning** | 'to dole out' |
|---|---|---|---|
| **Derivational analysis** | raz+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to dole out | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **BEN**$^{typ}_{za+4}$ | | |
| EXAMPLE: | *[…] Božjom ljubavlju koja se razdala za druge.* | | |
| TRANSLATION: | '[…] with divine love which doled itself out for others.' | | |

Regarding *razdati*, we would like to note that this verb exhibited some prominent pragmatic features. It seems that the religious context is prevalent within the concordances. That, coupled with the fact that there are merely 112 instances of it in HrWaC, leads us to believe it is primarily used in that context – giving out something, generally to the poor. *Razdati* also comes in reflexive and non-reflexive forms, however, no significant differences in meaning resulting from that have been found. Nevertheless, for consistency's sake, it too has been divided into two lemmas, following the example of previous such verbs.

Table 22: *poprodati*

| **Form** | *poprodati* | **Meaning** | 'to sell one by one' |
|---|---|---|---|
| **Derivational analysis** | po+pro+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to sell one after another | **AGT**$^{obl}_{0\_or\_1}$ **PAT**$^{obl}_{4}$ **AIM**$^{typ}_{sub\text{-}da}$ | | |

| | |
|---|---|
| EXAMPLE: | […] poprodao svu zemlju da bi ga školovao. |
| TRANSLATION: | '[…] he sold all his land to afford an education for him.' |
| | **AGT**$^{obl}_{0\ or\ 1}$ **PAT**$^{obl}_{4}$ **REC**$^{typ}_{3}$ |
| EXAMPLE: | […] i u suštini dionice državnih poduzeća poprodali strancima. |
| TRANSLATION: | '[…] and basically, sold shares of state-owned enterprises to foreigners.' |
| | **AGT**$^{obl}_{0\ or\ 1}$ **PAT**$^{obl}_{4}$ |
| EXAMPLE: | […] čovjeka koji je na ruletu dobio prvenstvo 2003. i nakon toga poprodao sve igrače. |
| TRANSLATION: | '[…] the man who won the 2003 roulette championship and then sold all the players.' |

Same as *pridati*, this verb also did not have a separate HJP entry, so the meaning was taken from the entry of its imperfective counterpart. The meaning of 'one after another' added by the prefix is evident in the meaning of the verb and the combination is highly compositional. Also, the valency frames remain largely unchanged by adding the prefix to *prodati*.

Table 23: *preprodati*

| **Form** | *preprodati* | **Meaning** | 'to resell' |
|---|---|---|---|
| **Derivational analysis** | pre+pro+dati | **Aspect** | perfective |

| **Valency frames** | |
|---|---|
| 1. to resell, second hand | AGT$^{obl}_{0_or_1}$ PAT$^{obl}_{4}$ |
| EXAMPLE: | *[...] mogu preprodati svoje pravo na kupnju od pravog vlasnika.* |
| TRANSLATION: | '[...] I can resell my right to purhase from the real owner.' |
| | AGT$^{obl}_{0_or_1}$ PAT$^{obl}_{4}$ THL$^{typ}_{adv\text{-}dok}$ |
| EXAMPLE: | *[...] dok traje taj kontrakt ja mogu preprodati tu istu nekretninu.* |
| TRANSLATION: | '[...] while the contract is valid, I can resell that same real estate.' |
| | AGT$^{obl}_{0_or_1}$ PAT$^{obl}_{4}$ REC$^{typ}_{3}$ INST$^{typ}_{putem+3}$ |
| EXAMPLE: | *[...] imaju priliku preprodati je drugom fanu putem Viagogo internet platforme.* |
| TRANSLATION: | '[...] they have the opportunity to resell it to another fan via Viagogo internet platform.' |
| | AGT$^{obl}_{0_or_1}$ PAT$^{obl}_{4}$ REC$^{typ}_{3}$ |
| EXAMPLE: | *[...] da smo oružje legalno preprodali nekoj od zemalja u regiji* |
| TRANSLATION: | '[...] that we resold the weapons legally to one of the countries within the region.' |
| | AGT$^{obl}_{0_or_1}$ PAT$^{obl}_{4}$ LOC$^{typ}_{na+6}$ |
| EXAMPLE: | *Informacije sa hakiranih profila se preprodaju na forumima.* |
| TRANSLATION: | 'The information from the hacked profiles is resold on forums.' |
| | AGT$^{obl}_{0_or_1}$ PAT$^{obl}_{4}$ EXT$^{typ}_{po+3}$ |
| EXAMPLE: | *[...] pa ga nedavno preprodao po tržišnoj cijeni.* |
| TRANSLATION: | '[…] so recently resold it at market price.' |

| | obl obl typ typ |
|---|---|

| | |
|---|---|
| | **AGT$_{0_or_1}$ PAT$_{4}$ MANN$_{preko+2}$ EXT$_{po+3}$** |
| EXAMPLE: | *[…] da ga vlasnici sada preprodaju preko oglasa po znatno skupljim cijenama.* |
| TRANSLATION: | '[…] so that the owners are now reselling it via ads at significantly higher prices.' |

Even though HJP offers only one meaning, *preprodati* is a verb which allows for many different typical arguments, and, as it seems, only two obligatory ones – AGENT and PATIENT. It is interesting to note that CROVALLEX only offers 3 different valency frames for 8 specified meanings of its derivational parent – *prodati*, which leads us to believe that either CROVALLEX deemed some typical arguments irrelevant, the entry is incomplete, or the additional prefix opens up the possibility of occurrence of these new frames. For this verb, reflexiveness does not play any relevant roles either.

Table 24: *rasprodati*

| **Form** | *rasprodati* | **Meaning** | 'to sell everything' |
|---|---|---|---|
| **Derivational analysis** | raz+pro+dati | **Aspect** | perfective |
| **Valency frames** | | | |
| 1. to sell everything | **AGT$^{obl}_{0_or_1}$ PAT$^{obl}_{4}$** | | |
| EXAMPLE: | *A mi? Rasprodali sve i sad idemo […]* | | |
| TRANSLATION: | 'And us? Sold everything and now we're going […]' | | |
| | **AGT$^{obl}_{0_or_1}$ PAT$^{obl}_{4}$ REC$^{typ}_{3}$** | | |
| EXAMPLE: | *Pobacat sve krpice van i rasprodat prijateljima.* | | |
| TRANSLATION: | 'To throw all clothes away and sell to friends.' | | |
| | **AGT$^{obl}_{0_or_1}$ PAT$^{obl}_{4}$ MANN$^{typ}_{u+4}$** | | |
| EXAMPLE: | *Poduzetnici su rasprodali zvučnike u tren.* | | |
| TRANSLATION: | 'Entrepreneurs sold out the speakers in an instant.' | | |

| | |
|---|---|
| | **AGT**$^{\text{obl}}_{\text{0\_or\_1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **AIM**$^{\text{typ}}_{\text{sub-kako}}$ |
| EXAMPLE: | *On je rasprodao sve kako bi bolesne mogao liječit i hraniti.* |
| TRANSLATION: | 'He sold everything to be able to heal and feed the sick.' |
| | **AGT**$^{\text{obl}}_{\text{0\_or\_1}}$ **PAT**$^{\text{obl}}_{\text{4}}$ **TWHEN**$^{\text{typ}}_{\text{adv-nedavno}}$ |
| EXAMPLE: | *Nedavno je rasprodao koncert u Kinu Šiška.* |
| TRANSLATION: | 'He recently sold out a concert in Kino Šiška.' |

For *rasprodati*, HJP specifies two meanings, both being 'to sell out', but one specifying that the things are being sold for lower price. Working with our corpus, the cotext was not enough to differentiate between the two meanings, as the price was not specified in any of the concordances. This is not to say that the differentiation is irrelevant or nonexistent, but that the random sample perhaps failed to account for this nuance, which would be expected to have an EXTENT argument, as in *prodati* and *poprodati*.

To conclude, this section provided the valency frames of the missing prefixed verbs, with additional remarks regarding the problems with the concordance, some interesting syntactic, semantic and/or pragmatic features, and when prominent, the specific influence the prefix has on the meaning and/or the valency frames of the prefixed verb.

## 5.2 Grammar evaluation

Sections *4 Analysis* and *5.1 Frame Output for the Missing Verbs* outlined some of the expected (and unexpected) problems with the corpus and the output of the grammar, and this final section provides the concrete numbers about the statistics of the grammar's efficiency, and in turn, the success of our automatization efforts, presented in Table 21:

Table 25: Precision, recall, f-measure

| | |
|---|---|
| **True positives** | 870 |
| **False positives** | 138 |
| **False negatives** | 0 |
| **Total in text** | 870 |
| **Precision** | 0.863 |
| **Recall** | 1 |
| **f-measure** | 0.926 |

The recall for the morphological grammar is 100% meaning that no prefixed forms slipped our detection, however, as expected, the precision is a bit lower given the amount of homography found within our corpus. This calls for disambiguation methods to differentiate nouns, prepositions, and other parts of speech from verbs to improve our precision.

# 6 Conclusion

## 6.1 Conclusions and final remarks

The theoretical overview in this research presented some key observations about prefixation of Croatian verbs and by focusing on the verb *dati* 'to give' and its numerous derivatives, we have illustrated the connections between prefixes and valency frames of derivationally related verbs. We have also created two NooJ grammars whose purpose is to detect and automatically connect derivationally related verbs to their base forms. With the aid of a model chunker for Croatian and our grammar, we have sped up the valency frame extraction process, and provided the missing valency formalizations for CROVALLEX[44].

Also, it is very interesting to notice that many of the verbs not present in CROVALLEX are those which have two prefixes, which points to the low prevalence of such combinations relative to those with just one prefix in Croatian, given that CROVALLEX is a frequency-based lexicon. Schematic 10 offers a visual representation of the connections between the frequency of the verbs in HrWaC and their inclusion in CROVALLEX[45,46]:


Schematic 10: *Dati* and its derivatives based on frequency and availability in CROVALLEX

[44] In a different direction, Vučković et al. (2008a) also utilize CROVALLEX and its valency data to enhance recognition of VP, NP and PP chunks in Croatian sentences with NooJ.

[45] The size of the font in the schematic is relative to the number of concordances found for the verbs in HrWaC. The black-colored verbs are present in CROVALLEX, while the grey ones are not.

[46] The infographic was created with the infogram app (https://infogram.com/), which unfortunately is not able to illustrate the differences in frequencies below a certain treshold, for example, between *poprodati* with only 13 hits and *preprodati* which has 1640. However, the size differences between the included and not included verbs are enough to get the point across, as the only included verb with two prefixes is *nadodati*, whose font size indicates that it is among the verbs with the lowest frequency.

Furthermore, the grammars and the results of this research can be used to improve not only CROVALLEX by adding the missing verb entries described in 5.1 but also aid the theoretical research into prefixation by exploiting the automatized connections generated by our grammars to create derivational families, much like the representation in the graph below:


Schematic 11: *Dati* and the connections between its derivatives

## 6.2 Future work

The path this research could take in the future can be divided into two directions, theoretical and practical.

Firstly, when it comes to theoretical implications of this thesis, it would be interesting to expand the research onto examples of prefixation of imperfective verbs to see the possible impacts of prefixes. Then aspect would also be a feature subject to change and not only meaning as is the case with perfective verbs, possibly creating more prominent discrepancies between the valency frames. In addition, the research into the prefixation of perfective verbs and its impact on syntactic properties could be expanded by focusing on other verbs, and not only *dati*, as well. Moreover, certain insights into meanings of some verbs that arose from our analysis, mainly the metaphorical extensions and metonymical shifts, could be researched further within the context of Cognitive Linguistics.

Secondly, on a more practical note, the number of batches within our morphological grammar should be increased and the list of prefixes should be expanded to include all the productive prefixes found in Croatian, keeping in mind their different allomorphs. Also, to continue the path of automatization, a more in-depth version of the syntactic parser for Croatian should be utilized within this context to make the valency frames extraction faster and more precise. Furthermore, the final step in the automatization process would be to use a semantic parser which would be able to tag the semantic roles of the syntactic arguments of the verbs, utilizing the existing resources for Croatian, such as HOBS (Tadić, 2007)[47].

Finally, to combine the theoretical and practical efforts on this topic, a slightly different approach could be taken to describe and formalize the similarities and differences in valency frames between verbs sharing the same prefix but differing in regard to the base verb.

[47] HOBS – Hrvatska ovisnosna banka stabala – Croatian Dependency Treebank, a corpus containing sentences tagged on morphosyntactic, dependency and semantic role levels. It has previously been utilized by Šojat et al. (2010b) in their efforts to extract verb valency frames.

# 7 References


Babić, S. (2002). *Tvorba riječi u hrvatskome književnom jeziku*. HAZU: Nakladni zavod Globus. Zagreb.

Barić, E., M. Lončarić, D. Malić, S. Pavešić, M. Peti, V. Zečević, M. Znika (2003). *Hrvatska gramatika*. Školska knjiga. Zagreb.

Belaj, B. (2008a). *Jezik, prostor i konceptualizacija. Shematična značenja hrvatskih glagolskih prefiksa*. Faculty of Humanities and Social Sciences, University of Osijek. Osijek.

Belaj, B. (2008b). "Pre-locativity as the schematic meaning of the Croatian verbal prefix pred-." *Jezikoslovlje* 9(1-2), 123–140.

Briscoe, T. & J. Carroll (1997). "Automatic extraction of subcategorization from corpora." *Proceedings of the 5th ACL Conference on Applied Natural Language Processing*, 356-363.

Dixon, R. M. W. & A. Aikhenvald (2000). *Changing Valency: Case Studies in Transitivity.* Cambridge University Press.

Hrstková K. (2007). "Czech Prefixed Verbs in a Valency Lexicon. Preliminary Study." *WDS'07 Proceedings of Contributed Papers* 1, 131-136.

Lakoff, G. & M. Johnson (1981). *Metaphors We Live By.* The University of Chicago Press. Chicago.

Lakoff, G. (1987). *Women, Fire and Dangerous Things. What Categories Reveal about the Mind.* The University of Chicago Press. Chicago.

Langacker, R. (1987). *Foundations of Cognitive Grammar*. Stanford University Press. Stanford.

Mikelić Preradović, N., D. Boras & T. Lauc (2013). "Verbal Prefixes and Syntactic- Semantic Classification of Verbs in CROVALLEX." *Recent Advances in Computer Science* 14, 35-40.

Raffaeli, I. (2007). "Neka načela ustroja polisemnih leksema." *Filologija* 48, 135-172.

Silić, J. & I. Pranjković (2007). *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta*. Školska knjiga. Zagreb.

Šarić, Lj. (2003). “Prepositional categories and prototypes: Contrasting some Russian, Slovenian, Croatian and Polish examples.” *Jezikoslovlje* 4(2), 187–204.

Šarić, Lj. (2006a). “A preliminary semantic analysis of the Croatian preposition u and its Slavic equivalents.” *Jezikoslovlje* 7(1-2), 1–43.

Šarić, Lj. (2006b). “On the meaning and prototype of the preposition pri and the locative case: A comparative study of Slavic usage with emphasis on Croatian.” *Rasprave instituta za hrvatski jezik i jezikoslovlje* 32, 225–248.

Šojat, K., B. Bekavac & K. Kocijan (2016). “Detection of Verb Frames with NooJ.” *Automatic Processing of Natural-Language Electronic Texts with NooJ*, *CCIS* 667, 157-168.

Šojat, K., K. Vučkovič & M. Tadić (2010a). “Extracting Verb Valency Frames with NooJ.” *NooJ 2009 International Conference and Workshop*.

Šojat, K., M. Srebačić & M. Tadić. (2012). “Derivational and Semantic Relations of Croatian Verbs.” *Journal of Language Modelling* 0, 1, 111-142.

Šojat, K., M. Srebačić & V. Štefanec (2013). “CroDeriV i morfološka raščlamba hrvatskoga glagola.” Suvremena lingvistika 39, 75-96.

Šojat, K., M. Srebačić, T. Pavelić & M. Tadić (2014). “CroDeriV: a New Resource for Processing Croatian Morphology.” *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14),* 3366-3370.

Šojat, K., Ž. Agić & M. Tadić (2010b). “Verb Valency Frame Extraction Using Morphological and Syntactic Features of Croatian.” *Proceedings of the 7th International Conference on Formal Approaches to South Slavic and Balkan Languages,* 119-126.

Tadić M. (2007). “Building the Croatian Dependency Treebank: the initial stages.” *Suvremena lingvistika* 63, 85-92.

Tadić, M., D. Brozović-Rončević & A. Kapetanović (2012). *The Croatian Language in the Digital Age.* Springer.

Vučković, K., Z. Dovedan & N. Mikelić Preradović (2008a). "Verb Valency Enhanced Croatian Lexicon." *Applications of Finite-State Language Processing: Selected Papers from the 2008 International NooJ Conference,* 52-60.

# 8 Resources

Hrvatski jezični portal (2006). http://hjp.znanje.hr/

Hrvatski morfološki leksikon (2005). http://hml.ffzg.hr/hml/

infogram (2012). https://infogram.com/

Ljubešić, N. & F. Klubička (2016). "Croatian web corpus hrWaC 2.1." *Slovenian language resource repository CLARIN.SI*. http://nlp.ffzg.hr/resources/corpora/hrwac/

Mikelić Preradović, N. (2008). *Pristupi izradi strojnog tezaurusa za hrvatski jezik.* Ph.D. dissertation, University of Zagreb. http://theta.ffzg.hr/crovallex/

Silberztein M. (2018). *NooJ Manual*. http://nooj-association.org

SketchEngine (2004). https://www.sketchengine.eu/

Šojat, K., M. Srebačić & V. Štefanec (2013). "CroDeriV i morfološka raščlamba hrvatskoga glagola." *Suvremena lingvistika* 39, 75, 75-96. http://croderiv.ffzg.hr/

Vučković, K., M. Tadić & Z. Dovedan (2008b). "Rule Based Chunker for Croatian." *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, 2544-2549.