Arguments for Phrasal Verbs in Croatian and Their Influence on Semantic Relations in Croatian WordNet

Daniela Katunar, Matea Srebačić, Ida Raffaelli, Krešimir Šojat

University of Zagreb, Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
E-mail: dkatunar@ffzg.hr, msrebaci@unizg.hr, ida.raffaelli@ffzg.hr, ksojat@ffzg.hr

Abstract

In this paper we introduce the category of phrasal verbs in Croatian lexicon and grammar description in order to show their influence on semantic relations, namely synonymy and polysemy in Croatian WordNet (henceforth CroWN). We discuss the practical and theoretical implications that arise from the introduction of the category of phrasal verbs in the description of the Croatian lexicon. We also address the interaction of synonymy and polysemy as manifested in the semantic relations of phrasal verbs to their monolexemic counterparts and facilitated by the structure of CroWN. The lemmatization of phrasal verbs in Croatian dictionaries and its modification for purposes of improving semantical relations in CroWN is also discussed. We also propose building of the Croatian phrasal verbs database, describe its structure and its further expanison which would facilitate extraction and incorporation of phrasal verbs into CroWN, and thus improve MT systems and information extraction via this computational lexical resource.

Keywords: phrasal verbs, semantic relations, synonymy, polysemy, Croatian WordNet

1. Introduction

Synonymy and polysemy are ubiquitous lexical semantic relations that continously structure the lexicon of a language. However, when it comes to their enumeration and notation within lexical resources, one is often faced with many caveats as to their valid representation. Particularly with regards to polysemy, the main problem seems to be a precise enumeration of various senses of a polysemous lexical unit, as well as their disambiguation from the various contexts they appear in (see Fellbaum, 2000, Fillmore & Atkins, 2000). On the other hand, though synonymy has been well described via thesauri as a very salient lexical relation, there is rarely an opportunity to study and represent the interaction of synonymy and polysemy within the format of tradicional dictionaries (see also Fellbaum, 1998). A fertile testing ground for such studies seems to be within the format of conceptual lexica such as WordNet. Since WordNet is conceived and built as complex network of lexical-semantic relations, it has a structure that necessitates the incorporation of various lexical-semantic relations, such as synonymy, antonymy, polysemy and hyperonymy/hyponymy in unison, i.e. it makes explicit their connections cross-cutting the structure of the lexicon of a language.

For instance, a polysemous unit in the Croatian WordNet (henceforth CroWN) *masa* 'mass' has seven distinct senses, three of which are *masa:1* 'a physical unit of weight', *masa:2, svjetina, puk, gomila* 'a crowd of people' and *vodena masa:3, vodena površina* 'lit. water mass, a body of water'. As the examples show, there is a three-way distiction between the senses in the way they interact with their surrounding lexical units. *Masa:1* 'a physical unit of weight' is a standalone lexical unit which denotes the source (or basic) meaning of 'mass' in general, that of weight. Conversely, *masa:2* is related to other lexical units in the same synset *svjetina, puk, gomila* 'a crowd', which clearly indicate the metaphorical shift in meaning that moved the particular sense of 'mass' into a different semantic domain. Furthermore, from the example we see how polysemy drives synonymy, i.e. by making semantic shifts lexical units are pushed into new synonymic relations with the lexical units profiling the same conceptual content in more-or-less the same way. The third sense of 'mass' ('a body of water') illustrates yet another principle by which polysemy structures the Croatian lexicon. Here not only has the semantic shift occured to indicate a specific homogenous and fairly large quantity of water (as in lakes and seas), but its specialization of meaning is further indicated by the collocation *vodena masa* 'lit. water mass, a mass of water'.

Although the example provided was from the category of nouns in CroWN, verbs behave in a similar manner, having even more polysemous senses entering into different synonymous relations and domains (Raffaelli&Katunar, 2010, in press). One notable property of verbs as opposed to nouns is their high degree of schematicity (Fellbaum, 1998), which accounts for a larger number of verb senses as well as smaller number of lexical units pertaining to the category of verbs. For this reason the makers of the original Princeton WordNet describe and categorize semantic verb relations in different terms from nouns, e.g. the relations of troponymy and entailment are considered as verbal counterparts of the noun relations hyperonymy/hyponymy and meronymy, respectively (Fellbaum, 1998). Polysemy of verbs is also described somewhat differently in WordNet. Peters et al. (1998), for instance, distinguish different criteria for sense disambiguation of verbs than that of nouns, such as transitivity/intransitivity, causativity/inchoativity etc. paired with the usage of different syntactic patterns that reflect the semantic shifts of verb lexemes. Miller (1999) and Fellbaum (1998, 2000) also point out repeatedly that polysemy operates under different principles when it comes to verbs as opposed to nouns. However, in the process of building CroWN (Raffaelli et al., 2008, Raffaelli&Katunar, 2010.) we have come face to face with certain regularities in both noun and verb relations that point to more general principles of polysemy working uniformly across categories, such as the ubiquitous mechanisms of metonymy and metaphor motivating the sematic shifts in both categories (see Lakoff, 1987, Langacker, 1987, Raffaelli&Katunar, 2010,

in press). Thus we believe that the aforementioned ways of interaction between synonymy and polysemy illustrated with the noun 'mass' are equally relevant for the same interaction for verbs, as we will show in the rest of the paper.

We will analyze polysemous lexical units in CroWN as defined in their senses by a) the surrouding lexical units of the same synset, b) by the semantic domain and hierarchy to which a particular sense belongs but also c) by specific constructional specifications in the lexical entries (one of these types being the example of 'mass' / 'water mass').

In this paper we deal especially with the last type of interaction mentioned, that of constructional specifications of lexical entries of verbs and we show how it serves to profile and specify the meaning of the category of the verb lexemes. While working on synsets in CroWN, it became apparent that some concepts are, along with one-word units, lexicalized as multi-word units. Though these are mostly mentioned as pertaining to idioms (also discussed in Fellbaum (1998) for English, e.g. 'kick' in *kick the bucket*), we will explore a more direct verbal construction, the V+Prep construction, in detail for the purposes of this paper (e.g. *poslati po* – to call, *zagrijati se za* – to be interested in). It is important to point out that the main verb has a completely different meaning without the preposition, and what is gained by adding a preposition to it is a holistic semantic unit[1] expressing a very different concept (e.g. *zagrijati se* 'warm up', zagrijati se za – to be interested in). We will also show that V+Prep. construction cannot be treated as an idiom, instead, this structure is consistent with what is called phrasal verbs in English (e.g. to make out, to run out). Introducing the concept of phrasal verbs is also very important, because it presents a novelty in the description of Croatian, as well as other Slavic languages. We believe that the incorporation the V+Prep constructions in the description of the Croatian lexicon is thus an important task that not only contributes to the fine-grained analysis of Croatian but also enriches the CroWN database and expands its applications in natural language processing tasks. Along with the incorporation of the V+Prep construction in CroWN, we set out to build a database of Croatian "phrasal verbs". We describe the methods used in building the database and demonstrate its applicability to the sense elaboration of verb synsets in CroWN as well as its benefits in the lemmatization of large corpora in the last section of the paper.

2. Phrasal Verbs

Phrasal verbs are a widely accepted phenomenon in languages such as English, and also in Dutch and German (Jackendoff, 2002), but as to our knowledge, there hasn't been a straightforward hypothesis about the existence of phrasal verbs in Slavic languages, including Croatian (cf. Sussex&Cubberley, 2006, Menac, 2007).

Descriptions of phrasal verbs vary from traditional approaches which interpret them as derivationally unpredictible, to cognitive approaches which point out the regularities of their meanings and formation through semantic shifts via metaphor and metonymy (see also Kovács, 2007). Taylor (2002) points out that the link between the verb and the preposition within the phrasal verb structure is notably different than a compositional V+ Prep. Thus in the example of 'look up' he shows that the interpretation can be twofold, depending on the compositionality or the bondedness[2] of 'look up':

1. look up the chimney – where 'look' can be replaced by 'peer' or 'gaze' up the chimney, or one can look down the chimney. In other words, the construction is compositional and its components can be replaced;

2. look up a word in the dictionary – where 'look' cannot be replaced by e.g. gaze (*gaze up a word) or any other lexical unit. In other words, „*look* and *up* coalesce to form a semantic unit in which the basic meaning of *up* has been coerced by a metaphorical meaning of *look* (Taylor, 2002: 330).

So, the criteria for identifying a phrasal verb are:
   a) the semantic unity of the V+Prep. construction;
   b) its distibutional properties which sanction the replacement of any of its parts by any other lexical unit.

Based on Taylor (2002) and other cognitive accounts (Lakoff, 1987, Langacker, 1987, Kovács, 2007 and others) we apply these criteria in the definiton and extraction of Croatian phrasal verbs. To our knowledge, nobody brought attention to the fact that phrasal verbs are not mentioned or described in Croatian. Furthermore, some authors even take the claim: "Phrasal verbs do not exist in Croatian language" (Geld, 2006) as some kind of a starting point in their papers. We find that the reasons for this ommision probably lie in (a) the contrastive analysis of Croatian and English, where prepositions are translationally equated with Croatian prefixes (eng. pull **out** – cro. **iz**vući; Arsenijević, 2004) (b) the fact that Croatian phrasal verbs form a smaller and more restrictive set than in English. However, as we will show in the following section, this set fits in the aforementioned criteria.

For the purposes of this paper two contemporary Croatian grammars[3] and two dictionaries[4] were consulted to see how they are dealing with verb constructions, namely V + Prep. constructions.

When it comes to Croatian grammars, phrasal verbs do not exist as a separate category, moreover, they're not even mentioned as a potential category in Croatian. Grammars that were taken into account mention verb government, but they do not give any detailed description, nor mention how different prepositions influence verb meaning. Government (rection) is simply presented as a verb[5] capacity to require a complement, namely object, in a predefined case. Such a classification is not cleary

---

[1] What we mean by the "holistic semantic unit" is a unit whose meaning is not simply a sum of its parts, i.e. compositional.

[2] Taylor (2002: 588) defines bondedness as a process „when units combine into a complex expression – especially when the composite form is entrenched and is characterized by coercion – it may be difficult to identify the expression's component units. The units become 'bonded' in a relatively unanalysible structure."

[3] Barić et al. (2003), Silić&Pranjković (2005).

[4] Anić (1991), Šonje (2000).

[5] As well as noun and adjective capacity (Silić&Pranjković, 2005: 263-264).

delimited as to the division between adverbials and object complements, and is sometimes confusing to discern to what it actually refers to. This problem arises from the fact that Croatian grammars do not delimit valency from government, instead they view them as synonymous (Silić&Pranjković, 2005:389) or do not mention valency at all (Barić et al., 2003).[6] As a consequence of this inadequate description of verb valencies Croatian dictionaries also don't include phrasal verbs, i.e. V+Prep. constructions with shift in meaning as separate lemmas. However, they do recognize a shift in meaning of verbs in different constructions, but list only the main verbs as lexical entries with different senses. Thus, the meaning 'to be interested in' is listed under the lemma *zagrijati se*, but the correlation of shift in meaning and preposition *za* isn't shown. In other words, the user of Croatian dictionaries cannot decode the fact that this particular shift of meaning occurs only in V+Prep. *za* construction.[7] In the only online dictionary of Croatian language[8] the situation is more or less the same, while it is based upon Anić (1991 and later) whose primary purpose was not conceived as a computational resource. It is therefore unhelpful, not only when it comes to individual users but also when it comes to disambiguating senses in machine translation (henceforth MT) systems or even in CroWN.

Thus, it needs to be shown how we can modify the current verb description and lemmatization in Croatian, in order to incorporate the set of phrasal verbs within its framework.

2.1.Semantics of Croatian Phrasal Verbs

In our analysis we were particularly interested in the change of the meaning of the main verb when followed by a particular preposition, in contrast to other prepositions which only function is to introduce several kinds of complements, namely objects or adverbials (see Taylor, 2002). For instance, the verb *zagrijati se* (to warm up) can be followed by different prepositions, among which are *pod* (under), *od* (from) and *za* (for):

1. (a) *zagrijati se pod pokrivačem* (to warm up under the blanket)
   (b) *zagrijati se od trčanja* (to warm up from running)
2. (a) *zagrijati se za lingvistiku* (to be interested in linguistics)
   (b) *zagrijati se za kuhanje* (to be interested in cooking)
   (c) *zagrijati se za Brada Pitta* (lit. to be interested in Brad Pitt; to have the hots for Brad Pitt)

---

[6] Conversely, we believe that the correct approach is to define government as referring solely to object complements, i.e. the verb governing the object case. On this account, valency is a broader term than government, and includes all sentence arguments, i.e. both subject, object and adverbial cases. For detailed description of valency in Croatian cf. Šojat (2009).

[7] Only in Šonje (2000) syntagmatic expressions are only vaguely noted in lexical entries as usage examples and not explained further.

[8] Hrvatski jezični portal (Croatian Language Portal), www.hjp.srce.hr. The fact is that HJP is slightly adapted Anić's dictionary.

3. *zagrijati se za utakmicu* (to warm up for the game)

It is obvious that in (1 a,b) the prepositions *pod* (under) and *od* (from) are part of the adverbials *pod pokrivačem* (under the blanket) and *od trčanja* (from running). They do not affect the verb's meaning, but only introduce a new circumstance of the action expressed by the main verb (in this particular case the location and the manner, respectively). On the contrary, the preposition *za* (for) in (2), apart from introducing a sentence object, completely changes the meaning of the main verb. *Zagrijati se* (to warm up) is metaphorically reinterpreted in accordance with what we may deem as the conceptual metaphor HAPPY IS WARM – SADNESS IS LACK OF HEAT (Kövesces, 2003), e.g. *ohladiti se od (koga)* (lit. to cool down, to loose interest (in somebody), *izgarati od (ljubavi, želje etc.)* (lit. burn with (love, desire)). Thus, the V+Prep. constuction in (2) expresses a very different concept than the V itself. Although it is clear that the metaphorical shift in meaning has happened and one can state that *to be interested in* is just one of the several meanings of the polysemous verb *zagrijati se*, what we claim is that the preposition is an explicit marker as well as an inherent part of that shift and thus should be a part of a lemma. As the examples in (2 a,b,c) also show, the meaning of the phrasal verb *zagrijati se za* is consistent regardless of the object complement following the preposition (it can be an abstract notion of science, e.g. linguistics or an activity, e.g. cooking or a person of romantic interest, e.g. Brad Pitt). Furthermore, one must be cautious to distinguish the compositional *zagrijati se za* (3) 'warm up' from the phrasal *zagrijati se za* (2 a,b,c) 'to be interested in'. Parallel to Taylor's (2002) description of 'look up' in English, these variants of *zagrijati se za* differ in their meaning in a way that (3) *za* is a part of the PP structure while in (2 a,b,c) is a part of the phrasal verb followed by an object. What follows from this distinction is the necessity to lemmatize *zagrijati se za* in (2a,b,c) 'to be interested in' separately from *zagrijati se* 'warm up'. Even though in (3) we see that *zagrijati se* 'warm up' can take *za* (for) as its complement it does not belong to its lemma because it is substitutable with any preposition and does not affect the verb's meaning. Such a semantic description argues for the separation of monolexemic and phrasal verbs in their lemmatization and notation in CroWN hierarchies.

On the other hand, we need to distinguish such phrasal verb constructions from idioms, i.e. other multi-word units (henceforth MWU). Idioms vary in their components and complexity, whereas phrasal verbs have only the V+Prep structure. Moreover, phrasal verbs illustrate the continuum of linguistic constructions (Fillmore, 1987), falling between the monolexemic verbs and full-fledged idiomatic constructions. Also, phrasal verb meaning is still, as we will demonstrate later, closely related and motivated by the schema of the polysemous structure of the verb itself.

2.2. Croatian Phrasal Verbs Database

Since, as we pointed out, phrasal verbs do not exist as lemmas in Croatian dictionaries, we weren't sure how to include them as literals in CroWN, but keeping them out of CroWN would significantly impoverish our resource.

So the first step we made was to write them down and create a small database of so called Croatian phrasal verbs.

Main verb	Prep.	Case	Synonyms
ciljati	na	ACC. a./i.	misliti
dovesti	do	GEN. i.	uzrokovati
držati	do	GEN. a./i.	cijeniti
ići	na	ACC. i.	poduzeti, namjeravati
ići	za	INST. i.	nastojati, težiti
patiti	od	GEN. i.	bolovati
plivati	u	LOC. i.	snalaziti se
poslati	po	ACC. a.	pozvati
privoljeti	na	ACC. i.	pristati
skinuti se	s	GEN. i.	odviknuti se
tući	po	LOC. a./ i.	pucati
ubiti se	od	GEN. i.	izmoriti se
zagrijati se	za	ACC. a./i.	zainteresirati se
zakačiti se	s	INST. a.	posvađati se
zapaliti se	za	ACC. a./i.	zainteresirati se

Main verb	Prep.	Case	Synonyms
to aim (at)	at	ACC. a./i.	to think
to bring (to)	to	GEN. i.	to cause
to hold	to	GEN. a./i.	to value
to go	on	ACC. i.	to opt for
to go	for	INST. i.	to aim at
to suffer (from)	from	GEN i.	to be ill, to suffer from
to swim (into)	into	LOC. i.	to get along
to send (for)	over	ACC. a.	to call
to persuade (to)	on	ACC. i.	to accept
to take (off)	with	GEN. i.	to quit
to beat	over	LOC. a./i.	to shoot
to kill (oneself)	from	GEN. i.	to exhaust oneself
to warm (up)	for	ACC. a./i.	to be interested in
to attach (to)	with	INST. a.	to fall out with
to burn (up)	for	ACC. a./i.	to be interested in

Figure 1 Sample of the Croatian phrasal verb database followed by an English translation

Our database includes the following data:
1. main verb – lemma in current dictionaries of Croatian language;
2. preposition – only the particular preposition which changes the meaning of the main verb in a specific way is listed;
3. case of the complement following the preposition along with its animacity (a.)/inanimacity(i.);
4. synonym(s). (for the sample see. Figure 1 below)
For example:
zagrijati se za A (a.)/(i.) zainteresirati se, zanijeti se
    V   Prep. case   synonyms
'to be interested in'.
Since there is no such thing as a lexicon or dictionary of Croatian verbs including prepositions following them, we weren't able to automatically extract all V+Prep. constructions, in order to find possible candidates for Croatian phrasal verbs database. Thus the manual making of the database is also a prerequisite for automatic

extraction of phrasal verbs from corpora. Since our primary goal is to enrich CroWN with phrasal verbs we started out by manually examining the list of about 2 300 verb synsets currently present in CroWN and extracting possible candidates for phrasal verbs. Those were primarily verbs with several senses whose synonyms in the same synset were indicative of a semantic shift occuring in the phrasal verb candidate. For instance, *ciljati* 'to aim at' appears in two synsets, one being defined as 'the act of aiming a weapon at somebody/something' and its synonym being the verb *nišaniti* 'to aim a weapon at'; the other synset contains the units *ciljati (na)* but also *misliti* 'to think', clearly indicating that *ciljati (na):2* has undergone a semantic shift into the domain of cognition and is also followed by the particular preposition, in this case *na* 'on'. So the second sense of *ciljati na* was treated as a phrasal verb candidate. The candidates extracted from the list of verbs in CroWN were then cross-referenced with their occurences in the CNC[9] in order to establish their syntactic patterns and distribution, i.e. to check whether they satisfy the two criteria for defining phrasal verbs (as listed above), the semantic unity of the MWU and its distribution. Its distributional pattern, i.e. the case occuring with a particular phrasal verb was also added to the database.[10] Furthermore, we started to develop a lexicon of Croatian verbs containing their derivational and inflexional forms, as well as their valency frames. This will facilitate detection of an even greater number of phrasal verb candidates in two ways:
1. when construing verb valency frames[11], we could recognize V+Prep. constructions which form holistic semantic units and include them in our database;
2. after construing verb valency frames, we could more easily extract all V+Prep. constructions in order to detect phrasal verbs among them.
This will be an important step towards expanding the database, since we have managed to manually extract 76 candidates so far, which may seem as a small sample, but it still comprises 3,2% of the current CroWN verb synsets and is highly indicative of a more widespread phenomenon in the Croatian lexicon.
The database will then be used to incorporate all detected phrasal verbs into CroWN, more precisely into synsets

---

[9] Croatian National Corpora, www.hnk.ffzg.hr.
[10] Another important aim is to get a general list of prepositions that can stand as a prepositional part of phrasal verbs in Croatian language. So far eight prepositions are extracted in our database, among them *za* (for) and *na* (on) being most frequent. The current list of prepositions could help us to extract more phrasal verbs from CNC by listing V+Prep constructions in more narrow way - we don't have to include all prepositions in Croatian, but only those that appear in the existing database.
[11] Construing verb valency frames in Croatian is almost completely manual work. There is only one printed Croatian valency dictionary *Rječnik valentnosti hrvatskih glagola* (Croatian Valency Dictionary), which is restricted to a very small set of verbs and does not give a complete description of valency frames, especially when it comes to the prepositions required by the verb. Much larger in size and quantity is *Crovallex* (Mikelić Preradović et al., 2009), an electronic lexicon of Croatian verbs which resembles in its structure to Czech lexicon *Vallex* (Žabokrtský, Lopatková, 2007). See Šojat (2009).

which contain synonyms listed next to the them in the database. This implies that we would treat phrasal verbs as a separate lemmas in CroWN which would also have different synonyms, hyperonyms etc. than the main verb of the phrasal construction. It also means that once the list of phrasal verbs is complete and added to CroWN we could simply add the list to the list of lemmas in CNC and thus lemmatize the entire corpus. In the next chapter we illustrate the interaction of polysemy and synonymy as reflected in the CroWN structure pertaining to phrasal verbs and their semantic relations.

2.3. Semantic Relations of Phrasal Verbs in CroWN

There are two important aspects of the interaction of synonymy and polysemy with regards to phrasal verbs. First, phrasal verbs are specifications of the more schematic meaning denoted by the main verb via prepositions. Secondly, since polysemy drives synonymy, these verbs are also placed in different synsets as well as different lexical hierarchies, which implies a whole new set of semantic relations gained by the semantic shift in specialization. On the other hand, their relation to the other senses of the main verb in CroWN is preserved through the inclusion of the sense of a phrasal verb as one of the senses of its main verb. To illustrate this point, we will describe and show the semantic relations of the verb *držati* 'to hold'. Since this is a highly polysemous verb in Croatian, it has a plentitude of senses registered in CroWN, one of them being specified by a phrasal verb *držati do* 'to value'. All together, the verb *držati* 'to hold' has 13 senses, the thirteenth being the sense *držati do* 'to value'. Below in Figure 2 is the entire synset to which it belongs, along with its synonym pairs, definition and usage examples (followed by its PWN counterpart).

<SYNSET>
<ID>ENG20-00670967-v</ID>
<POS>v</POS>
<SYNONYM>
<LITERAL>**cijeniti**<SENSE>2</SENSE></LITERAL>
<LITERAL>**štovati – poštovati - poštivati**<SENSE>2</SENSE></LITERAL>
<LITERAL>**držati do**<SENSE>13</SENSE></LITERAL>
<LITERAL>**respektirati**<SENSE>9</SENSE></LITERAL>
</SYNONYM>
<DEF>imati visoko mišljenje o komu ili čemu; uvažavati čije mišljenje</DEF>
<USAGE>Visoko cijenim njezino sposobnosti.</USAGE>
<USAGE>Poštujem tvoju slobodu govora.</USAGE>
<USAGE>Držim do tvojeg mišljenja.</USAGE>
<BCS>2</BCS>
<DOMAIN>factotum</DOMAIN>
<SUMO>IntentionalPsychologicalProcess<TYPE>+</TYPE></SUMO>
<CROWN>1</CROWN>
</SYNSET>

<SYNSET>
<ID>ENG20-00670967-v</ID>
<POS>v</POS>
<SYNONYM>
<LITERAL>**respect**<SENSE>1</SENSE></LITERAL>
<LITERAL>**esteem**<SENSE>1</SENSE></LITERAL>
<LITERAL>**value**<SENSE>3</SENSE></LITERAL>
<LITERAL>**prize**<SENSE>3</SENSE></LITERAL>
<LITERAL>**prise**<SENSE>3</SENSE></LITERAL>
</SYNONYM>
<DEF>regard highly; think much of </DEF>
<USAGE>I respect his judgement.</USAGE>
<USAGE>We prize his creativity.</USAGE>
<BCS>2</BCS>
<DOMAIN>factotum</DOMAIN>
<SUMO>IntentionalPsychologicalProcess<TYPE>+</TYPE></SUMO>
</SYNSET>
Figure 2 Phrasal verbs in CroWN synsets

In the example it is clear that *držati do* 'to value' enters into rather different synonymic relations than for instance *držati* 'to hold (in ones hand)'. As the synonyms surrounding *držati do* 'to value' indicate, the meaning of the phrasal verb *držati + do* 'lit. hold to, to value' is far removed from the domain of physically grasping on object (as in 'to hold in ones hand') and pertains to the domain of psychological processes, namely those including respect and judgement. The semantic shift here is cleary metaphorical, as it includes a movement from a concrete domain (physical object interaction of 'holding') to the abstract domain of judgement. The connotations added to the abstract notion of 'holding to or valuing' are further motivated by the domain of judgement. Thus we see that the polysemous shift motivated the verb to specialize in meaning and enter synonymic relations with 'respect' and 'value', which otherwise would not be possible. Futhermore, it is important to stress that the monolexemic verb *držati* 'to hold' would not be able to enter these relations because it would not have been specified enough as to its meaning, i.e. the only possibilty is to have a phrasal verb as lemma in CroWN since the option of entering only the main verb would leave the relations in this particular synset understated and vague.

To further stress the importance of proper specification of lemmas and their polysemous relations in CroWN, we will present the entire polysemous structure of the polysemous verb *držati* 'to hold', taken and modified from Raffaelli&Katunar (2010, in press). Raffaelli&Katunar (2010, in press) do not include in their analysis phrasal verbs and do not treat them as separate lemmas in CroWN, although they discuss in detail the ways of presenting polysemous verbs as radial structures. Thus we modify the existing graph (see Figure 3. below) of *držati* 'to hold' in order to show how the inclusion of phrasal verbs adds relevant information about parts of the radial structure containing phrasal verbs as well as the structure of the Croatian lexicon.
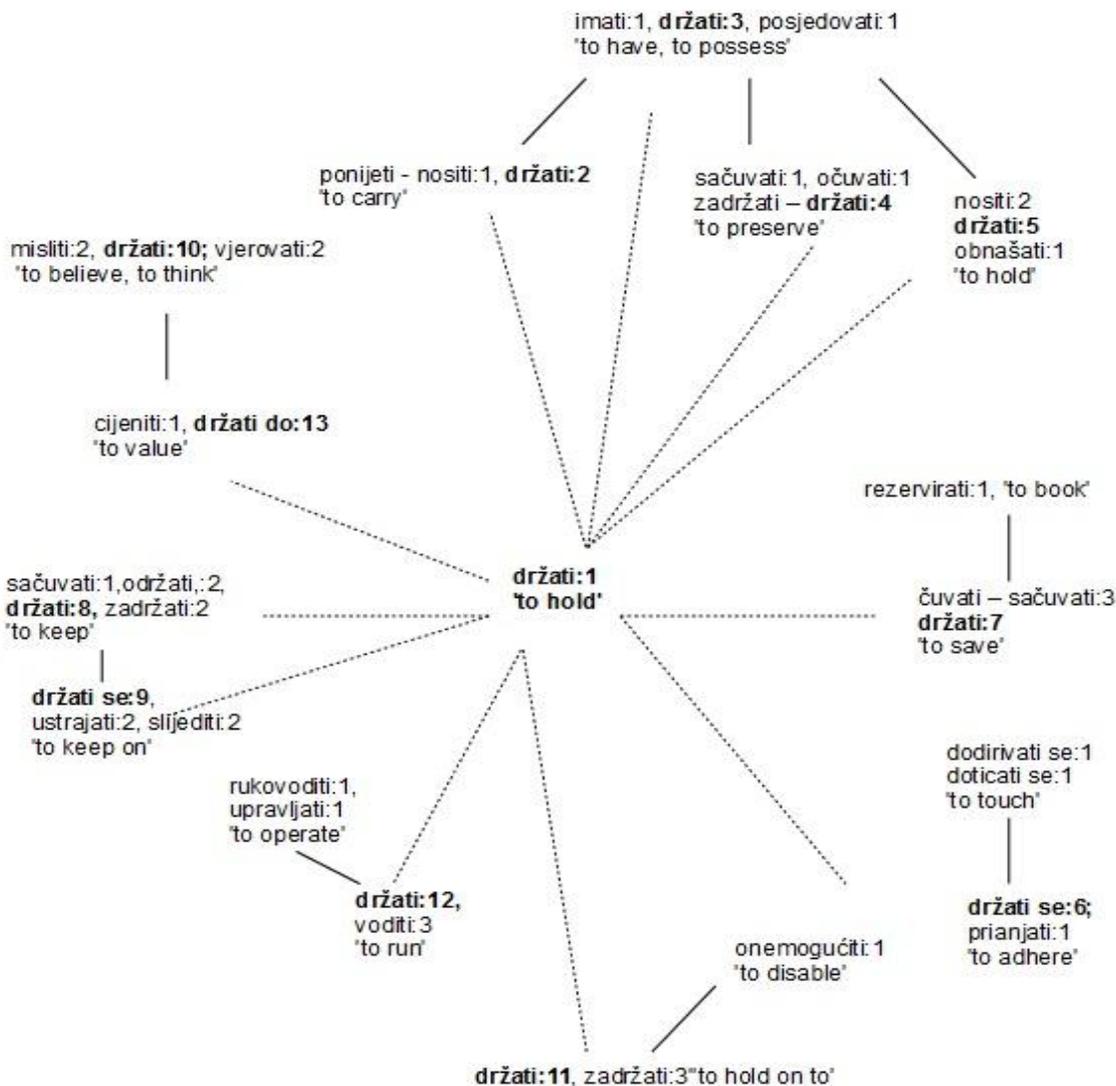
imati:1, **držati:3**, posjedovati:1
'to have, to possess'

ponijeti – nositi:1, **držati:2**
'to carry'

sačuvati:1, očuvati:1
zadržati – **držati:4**
'to preserve'

nositi:2
držati:5
obnašati:1
'to hold'

misliti:2, **držati:10**; vjerovati:2
'to believe, to think'

cijeniti:1, **držati do:13**
'to value'

rezervirati:1, 'to book'

držati:1
'to hold'

čuvati – sačuvati:3
držati:7
'to save'

sačuvati:1, održati,:2
držati:8, zadržati:2
'to keep'

držati se:9,
ustrajati:2, slijediti:2
'to keep on'

dodirivati se:1
doticati se:1
'to touch'

rukovoditi:1,
upravljati:1
'to operate'

držati:12,
voditi:3
'to run'

držati se:6;
prianjati:1
'to adhere'

onemogućiti:1
'to disable'

držati:11, zadržati:3"to hold on to'

Figure 3 Semantic relations of the verb *držati* 'to hold' and its senses in CroWN. Above each sense are the hyperonymic synsets noted by the continous lines. The dotted lines represent sense extensions from the source meaning *držati:1* 'to hold physically in one's hands'.

The polysemy of *držati* 'to hold' is very clearly shown in Figure 3., where the verb has 13 senses that vary from the concrete sense of 'holding in ones hands' to the senses of 'keeping', 'thinking', 'possesion', 'adhering' etc. What Figure 3. also indicates is the path of the semantic shift from the source meaning *držati* 'to hold' to *držati do* 'to value'. The shift is not a direct one, but it includes a) the metaphorical shift from *držati:1* 'to hold' to *držati:10* 'to think, to believe' motivated by the fact that one can 'hold an opinion or belief' in the abstract sense, and b) the specialization of *držati:10* 'to think, to believe' by the features of judgement and esteem added by the preposition *do* 'to' in *držati do:13* 'to value'. In other words, the link between *držati:10* 'to think' and the more specific *držati do* 'to value' is best described in the way that *držati do* 'to value' specifies a particular manner of thinking, that of 'holding on to' a person, opinion etc., which implies the relevancy of the entity one is 'holding

on to' or 'thinking of', allowing it to have a value component of its meaning.

It is clear from the example in Figure 3. that by adding the V+Prep construction we describe the properties of the entire radial structure in more detail, and represent the semantic shifts, especially specification in this case, as processes transparently noted in the lemmas themselves, i.e. in the preposition added to the main verb. What this allows is an expansion of synonymic and polysemous relations in CroWN, as well as (in some cases) the inclusion of phrasal verb into new hierarchical relations with which they otherwise had no relation at all as monolexemic units (see example above *zagrijati se za* 'to be interested in').

3. Conclusion and Future Work

In this paper we presented the description of phrasal verbs in Croatian, which to our knowledge are ommitted from

any current and past descriptions of the Croatian lexicon and grammar. We emphasized the importance of this description from the viewpoint of a) a fine-grained analysis of semantic relations in CroWN, and b) the interaction of synonymy and polysemy as manifested in the semantic relations of phrasal verbs to their monolexemic counterparts and facilitated by the structure of CroWN, and c) current lemmatization of phrasal verbs in Croatian dictionaries and its modification for the necessities of CroWN. For these purposes we proposed building a database of Croatian phrasal verbs, described its structure and the methods of its further expansion. Future work includes building valency frames which would enable this expansion, but also the extraction of V+Prep constructions in large corpora and incorporation of the extracted phrasal verbs into CroWN verb hierarchies. We believe that this work will contribute to (a) the theoretical aspects of the interaction between polysemy and synonymy; (b) description of the Croatian verb system; (c) the enrichment of semantic relations in CroWN; (d) lemmatization of verbs in CroWN and other resources such as CNC; (e) facilitating MT applications and information extraction via CroWN.

4. References

Anić, V. (1996). *Rječnik hrvatskoga jezika* [The Dictionary of Croatian Language]. Zagreb: Novi Liber.

Arsenijević, B. (2004). *Non-predicative Particles as Adjuncts to Abstract Arguments*. An abstract from the CHRONOS 6 Conference in Geneva, 2004. http://www.unige.ch/lettres/latl/chronos/Arsenievic.pdf

Barić, E. et al. (2003). *Hrvatska gramatika* [Croatian Grammar]. Zagreb: Školska knjiga.

Dehé, N., Jackendoff, R. et al. (eds.) (2002). *Verb-Particle Explorations (= Interface Explorations 1)*. Berlin / New York: Mouton de Gruyter.

Geld, R. (2006). Strateško konstruiranje značenja engleskih fraznih glagola [Strategic Construal: English Particle Verbs]. *Jezikoslovlje*, 7(1-2), pp. 67 – 111.

Fellbaum, Ch. (ed.) (1998). *WordNet. An Electronic Lexical Database. With a preface by George Miller.* Cambridge, MA: MIT Press.

Fellbaum, Ch. (1998). Towards a Representation of Idioms in WordNet. In *Proceedings of the COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems*. Montreal: COLING/ACL, pp. 52 – 57.

Fellbaum, Ch. (2000). Autotroponymy. In Ravin, Y. and Leacock, C. (Eds.), *Polysemy*, pp. 52 – 67. Cambridge: Cambridge University Press.

Fillmore, Ch. and Atkins, B.T. (2000). Describing Polysemy: the Case of Crawl. In Ravin, Y. and Leacock, C. (Eds.), *Polysemy*, pp. 91 – 110. Cambridge: Cambridge University Press.

Fillmore, Ch., Kay, P., O'Connor, M. K. (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of *let alone*. *Language* 64, pp. 501 – 538.

Kovács, É. (2007). The Traditional vs. Cognitive Approach to English Phrasal Verbs. *English.*: http://epa.oszk.hu/02100/02137/00022/pdf/EPA02137_I SSN_1219-543X_tomus_16_fas_1_2011_141-160.pdf

Kövecses, Z. (2000). *Metaphor and Emotion: Language, Culture, and Body in Human Feeling.* Cambridge: Cambridge University Press.

Lakoff, G. (1987). *Women, fire and dangerous things: What Categories Reveal about the Mind.* Chicago/London: The University of Chicago Press.

Langacker, R.W. (1987). *Foundations of Cognitive Grammar. Volume 1. Theoretical Prerequisites.* Stanford: Stanford University Press.

Menac, A. (2007). *Hrvatska frazeologija* [Croatian Phraseology]. Zagreb: Knjigra.

Mikelić Preradović, N., Boras, D., Kišiček, S. (2009). CROVALLEX: Croatian Verb Valence Lexicon. In *Proceedings of the ITI 2009 31st International Conference of Information Technology Interfaces.* Zagreb: SRCE, pp. 533 – 538.

Miller, G.A. (1999). Nouns in WordNet. Fellbaum, C. (Ed.), *WordNet. An Electronic Lexical Database.* Cambridge/Massachussetts/London : MIT Press, pp. 23 – 47.

Peters, W. et al. (1998). Cross-linguistic Alignment of Wordnets with an Inter-Lingual_Index. Vossen, P. (Ed.), *EuroWordNet: A multilingual database with lexical semantic networks.* Dordrecht/Boston/London: Kluwer Academic Publisher, pp. 221 – 225.

Raffaelli, I. et al. (2008). Building Croatian WordNet. In Tanács, A. et al. (Ed.) *Proceedings of the 4th Global WordNet Conference.* Szeged: Global WordNet Association.

Raffaelli, I., Katunar, D. (2010). Leksičko-semantičke strukture u Hrvatskom WordNetu [Lexical-semantic structures in the Croatian WordNet]. *Filologija* (in press).

Silić, J., Pranjković, I. (2005). *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta* [Croatian Grammar for High Schools and Universities]. Zagreb: Školska knjiga.

Sussex, R., Cubberley, P. (2006). *The Slavic languages. (Cambridge Language Surveys).* Cambridge: Cambridge University Press.

Šojat, K. (2009). Morfosintaktički razredi dopuna u Hrvatskom WordNetu [Morphosyntactic Annotation in the Croatian WordNet]. *Suvremena lingvistika* 68, pp. 305 – 339.

Šonje, Jure (Ed.) (2000). *Rječnik hrvatskoga jezika* [The Dictionary of Croatian Language]. Zagreb: Leksikografski zavod Miroslav Krleža: Školska knjiga.

Taylor, J. R. (2002). *Cognitive Grammar.* New York : OUP.

Žabokrtský, Z., Lopatková, M. (2007). Valency Information in VALLEX 2.0.: Logical Structure of the Lexicon. *The Prague Bulletin in Mathematical Linguistics*, 87, pp. 41 – 60.