Ljiljana Šarić
ILOS, University of Oslo, Norway
ljiljana.saric@ilos.uio.no

Svetlana Nedelcheva
Shumen University, Bulgaria
svetyo@yahoo.com

The verbal prefix o(b)– in Croatian and Bulgarian: The semantic network and challenges of a corpus–based study

This study compares the verbal prefix *o(b)*– in two South Slavic languages, Croatian and Bulgarian, from a cognitive linguistic perspective. We focus on the problems arising when constructing the semantic network of this polysemous prefix, particularly on 1) isolating the prefix's meaning from the meaning of the base verb and 2) identifying core/dominant sub–meanings for all verbs and giving them corresponding semantic labels. Our approach to morphology is based on extensive databases of verbs collected from dictionaries and a few corpora. However, our work with corpora led to a number of challenges. This study thus has two aims: a) presenting challenges encountered in working out semantic networks of prefixes, and b) presenting challenges related to obtaining reliable (quantitative) results from the corpora.

---

1. Introduction

1.1. Theoretical preliminaries and state of the art

Our approach to prefixes and prefixed verbs follows the basic theoretical assumptions of cognitive linguistics, which views spatial particles as networks of interrelated meanings (e.g., Janda 1986; Tabakowska 2003; Przybylska 2006; Belaj 2008; Šarić 2008, 2014; Janda et al. 2013). Cognitive linguistics assumes that category members (i.e., sub–meanings of a language unit) share different sets of attributes with one another and form a coherent meaning network. This approach allows fuzzy boundaries among concepts, and identifies more or

149

less prototypical meanings of language units to which other meanings directly or less directly relate (see, e.g., Langacker 1987; Tyler & Evans 2003).

Image schemas in cognitive linguistics depict two basic entities: a trajector (TR) and a landmark (LM), respectively defined by Langacker (1987) as the figure within a relational profile and another salient entity in a relational predication, prototypically providing a point of reference for locating the TR. An image schema is "a cognitive representation comprising a generalization over perceived similarities among instances of usage" (Barlow & Kemmer 2000: viii). Image schemas lack specificity and content, which makes them highly flexible preconceptual and primitive patterns used for reasoning (Johnson 1987: 30).

There are a number of twentieth–century studies dedicated to prefixes, prefixed verbs, and their aspectuality. Linguists are unanimous that prefixes are polysemous, but the question of their desemantization and grammaticalization is still open. Some authors claim very decisively that the only function of some prefixes is "to transform imperfective verbs into perfective" (e.g., Kostov 1939: 120). Andrejčin (1944: 198–199) expresses a similar view, stating that "in some cases the inherent meaning of the prefixes has bleached so much that the perfective verbs they form only differ from their imperfective counterparts in aspect." When discussing desemantized or empty verbal prefixes in Bulgarian, Ivanova (1966: 135) presents desemantization in a scale in which the opposite side is grammaticalization. In her analysis, the semantics of a grammaticalized prefix is never zero and it is the result of the interaction between the semantics of the prefix and the verbal base. Our study follows the trends in cognitive linguistics, and we suggest that "empty" affixes do indeed have semantic content; the illusion of semantic emptiness is due to a large amount of conceptual overlap (Langacker 1999) between the meaning of the stem and the affix (see also Janda et al. (2013) for further details on the Overlap Hypothesis). Furthermore, prefixes form radial categories (Lakoff 1987) organized around prototypes.

The Bulgarian prefix *o(b)*– has not received much attention in research to date.[1] (The same is true of the Croatian verbal prefix *o(b)*–. Belaj's (2008) cognitive linguistic study of Croatian prefixes does not include *o(b)*–. Grammars and word–formation manuals (e.g., Babić 1986: 483) and some dictionaries (such as the online dictionary *Hrvatski jezični portal*, hereinafter HJP[2]) provide basic information about the meaning of *o(b)*–. In traditional accounts, the meaning of prefixes in general is a challenge. It is unclear how the various meanings of prefixes are related. Equally unclear is the relation between their spatial and non–spatial meanings. Furthermore, there are no clear criteria for ordering the individual meanings. Interestingly, Croatian language manuals and research do not agree on whether *o*– and *ob*– are allomorphs or two prefixes.

---

1   For an analysis of a few other Bulgarian prefixes in the cognitive linguistic framework, see Tchizmarova (2005, 2006), Šarić and Tchizmarova (2013), and Nedelcheva (2010, 2012).

2   Available at: http://hjp.srce.hr/index.php?show=search. HJP is based on six manuals, including Anić (see http://hjp.srce.hr/index.php?show=baza).

150

Babić (1986: 483) approaches *o*– and *ob*– as two prefixes, singling out two meanings of *o*–[3] and only one of *ob*–[4] (indicating that other meanings–e.g., perfectivity–are possible, although rare). HJP considers *o*– and *ob*– to be a single prefix. This dictionary lists four meanings of *o(b)*– in verbs (the second meaning has three sub–meanings),[5] the relation of which is unclear.

We discuss the issue of allomorphy in detail elsewhere (see Endresen et al. forthcoming; Šarić & Mikolić forthcoming). Here, we repeat the main arguments.

Regarding whether *o*– and *ob*– are semantically one unit (allomorphs) or two different prefixes, we consider their meaning similarities or differences to be a decisive factor, in addition to their common etymology, and tendencies in the historical and areal distribution of the two forms. For the hypothesis of *ob*– and *o*– as two units to be acceptable, their central meaning should not overlap. However, their central meaning does overlap in Croatian, as it does in Bulgarian, and other Slavic languages. That the central meaning of *o*– and *ob*– coincides is even confirmed by those analyses of Croatian that formally consider them to be two prefixes: for example, Babić (1986: 483) provides an almost identical definition of the meanings of *o*– and *ob*–: *o*–: "the action of the base verb revolves around something . . . encompasses all its sides"; *ob*–: "the action of the base verb revolves around something, from all its sides." The only apparent difference in Babić's account between the two forms is in *o*– having a resultative meaning. However, *ob*– also has a resultative meaning, and so this component is certainly not the factor that confirms the hypothesis of *o*– and *ob*– as two units.

Etymological information is also of outmost importance. Skok (1988: 533) describes the common Slavic prepositions *o* and *ob* as variants of the same unit. Trubačev (2001) provides comprehensive etymological information comparing many verbs with Proto Slavic **ob*– in various Slavic languages and includes information about various historical periods and dialects of Slavic: his extensive sources show that many otherwise identical verbs choose *ob*– in some Slavic languages, and *o*– in others, and that verbs prefixed with both *ob*– and *o*– exist in parallel in some languages. These verbs are either (near–) synonyms in the modern languages, or inter–dialectal synonyms, or one form is archaic and another one modern.[6]

---

3   (a) . . . the effect of the base verb goes around something . . . encompasses all sides; (b) lead to a result.

4   "The verbs with the prefix *ob*– mean that the action of the main verb happens around something, at all its sides in a concrete and abstract sense."

5   http://hjp.novi–liber.hr/index.php?show=search_by_id&id=eFxjXhE%3D&keyword=ob.

6   For instance, Proto Slavic **oblupiti* is realized with both *o*– and *ob*– in some Slavic languages, whereas some others only have a prefixed verb with *o*–. Verbs with both *ob*– and *o*– refer to removing a surface layer, most frequently to removing skin and bark (see Trubačev 2001: 30). The variation is not attested in verbs similar to **oblupiti* only (which relate to removing a surface layer), but also in verbs indicating physical motion (e.g., Proto Slavic **oblučiti* 'light up, illuminate (while carrying a torch around)'), and verbs indicating that an action affects all sides of an object with a substance (e.g., Proto Slavic **obmakati* 'soak, wet, moisten').

Our approach to *o*– and *ob*– as one unit is grounded in an account of their meaning network: they do not show any significant differences in their meaning. In this analysis, we follow Baydimirova's (2010) understanding of allomorphy, which takes into account a systematic relation of prefixal forms in all their meanings, spatial and non–spatial.[7] The central spatial meaning of *o*– and *ob*– overlaps, as well as their other extended meanings, and they follow the same paths of meaning extensions in the contemporary meaning network. Our approach is further grounded on unambiguous etymological information showing diachronic development, and it is confirmed by accounts of the equivalent prefix in other Slavic languages (e.g. Baydimirova 2010, Będkowska–Kopczyk & Lewandowski 2012).

The situation in Croatian is somewhat additionally complicated by the fact that the ablative prefix *od*– is reduced to *o*– in some environments, and the fact that some semantic subgroups of verbs prefixed with *o(b)*– are similar in meaning to some verbs prefixed with *od*–. We reflect on this later in this paper when we discuss these subgroups of *o(b)*– verbs.

In Bulgarian, *o*– and *ot*– are not discussed as related.[8] Most Bulgarian grammars treat *ob*– as an allomorph of *o*– (Pashov 1966; Gerov 1977; Tilkov et al. 1993; etc.).[9] According to Gerov (1977: 288) *o* is used as a preposition and is written separately, or together with other words as a prefix, whereas *ob*– is only used as a prefix. The Bulgarian etymological dictionary claims that *ob* exists as a preposition (in some dialects) and as a prefix, and is a variant of *o*– before consonants (*Bălgarski etimologičen rečnik* 1995: 735–735). *Rečnik na bălgarskija ezik* (Bulgarian Dictionary) compiled by the Bulgarian Academy of Sciences (2002: 7–9) lists *o*– and *ob*– separately, each of them with a number of meanings, part of which overlap.

Our analysis pays particular attention to how seemingly unrelated meanings of *o(b)*– relate to each other, and how they all relate to the central spatial meaning. Such an approach to spatial particles has clear advantages in L2 contexts.

7　Interestingly, Baydimirova (2010) considers *o*– and *ob(o)*– in Russian allomorphs of the same prefix, although they add rather different meanings to some verbs and yield minimal pairs. She calls this phenomenon "irregular allomorphy" (Baydimirova 2010: 10). Her argumentation is that O, OB, and OBO should be treated as "one morpheme with a non–complementary but at the same time statistically significant distribution of allomorphs" (Byadimirova 2010: 110). Both spatial and non–spatial meanings of these prefixes are systematically related and can be discussed within the same semantic model. This, according to her, implies that the traditional understanding of allomorphy is "too narrow and should take into account the gradient and complex nature of this linguistic phenomenon" (ibid.).

8　Moreover, when studying Bulgarian prepositions, which are seen as cognate with prefixes, Konstantinova (1982: 100) points out that *o* is sometimes wrongly substituted by *ot*: for example, *Podprja se ot zida* (He leaned *ot* 'from' the wall). The change of the preposition changes the meaning of the sentence and makes the spatial configuration unclear. The person is not leaning against the wall but moves in a direction opposite the wall, as *ot* suggests, which is impossible in this context.

9　There are a number of verb stems that can be prefixed by both *o*– and *ob*–. They either bear the same meaning (e.g., *okiča* 'adorn, decorate' – *obkiča* 'adorn, decorate') or are slightly differ in meaning (e.g., *ogradja* 'enclose, fence off' – *obgradja* 'surround, enclose'). The discussion of those pairs of verbs will be part of a separate future study.

152

A few cognitive linguistic studies analyze prefixes equivalent to the Bulgarian and Croatian *o(b)*– in Polish, Russian, and Slovenian. Twardzisz (1994) identifies six configurations of the Polish verbal prefix *ob(e)*–, some of which have metaphorically extended variants: the result is a total of fourteen configurations. The most prominent and central configuration is the notion of circularity or roundness, in which a TR prototypically moves around a LM. Another central configuration in the meaning network is the one in which a TR performs an action around a LM so that the LM is either "wrapped" or "framed" (Twardzisz 1994: 221–225). In their corpus–based analysis of the Polish and Slovenian verbal prefix *ob*–, Będkowska–Kopczyk and Lewandowski (2012) provide statistical evidence for the meanings of Polish and Slovenian prefixed verbs. The authors identify eleven meanings for Polish verbs, and ten meanings for Slovenian verbs. The main difference between Slovenian and Polish seems to relate to the meaning "metaphorical remaining in proximity," which is identified in 162 Slovenian verbs, but not attested at all in Polish. In her corpus– and experiment–based study, Baydimirova (2010) analyses the Russian prefixes *o*–, *ob*–, and *obo*–, which can be synonyms but also carry strikingly different meanings that even yield minimal pairs (e.g., *o–sudit'* 'condemn' vs. *ob–sudit'* 'discuss'; see Baydimirova 2010: iv). She works out a common semantic network with fifteen interrelated meanings of *o*–, *ob*–, and *obo*–, in which 'move around an object' is the central one (see Baydimirova 2010: 42). The author shows that the meanings of the three forms that seem unrelated are actually parts of a single semantic network and that all sub–meanings can be expressed by each of the three forms.

The findings by Twardzisz (1994), Baydimirova (2010), and Będkowska–Kopczyk and Lewandowski (2012) are relevant for our analysis. When discussing the semantic network of the prefix *o(b)*– in Bulgarian and Croatian, we indicate differences we noticed in relation to other Slavic languages. Although there are unquestionable similarities in the semantic networks of the Slavic *o(b)*–, there are also some differences worth examining.

1.2. Methodology and databases

Cognitive linguistics recently experienced an "empirical turn"; that is, a trend towards the use of empirical methods, (see, e.g., the introduction to Glynn & Fischer 2010) along with a "quantitative turn" (see Janda 2013). The use of quantitative methods in studying cognitive semantics has become increasingly popular. Quantitative methods, according to their proponents, contribute to turning cognitive semantics into a truly usage–based model, connecting it to corpus linguistics. The "quantitative turn" can help introduce more methodological rigor into the discipline, and many studies seek to demonstrate the advantages of quantitative analyses. However, using corpora and carrying out quantitative analyses poses challenges that are perhaps more relates to some (less commonly studied and taught) languages than others. We believe that it is important to address these challenges, and we do so in the second part of this analysis (Section 4).

153

Our analysis relies on an extensive inventory of verbs collected in our two databases: a database of Croatian verbs compiled on the basis of two dictionaries and two corpora, and a database of Bulgarian verbs compiled on the basis of the Bulgarian National Corpus and two online dictionaries.

We compiled the initial database of Croatian *o(b)*– verbs using two comprehensive dictionaries that reflect contemporary usage: 1) Željko Bujas: *Veliki hrvatsko–engleski rječnik*, 2001 (bilingual) and 2) Vladimir Anić: *Rječnik hrvatskoga jezika*, 2004 (monolingual). Bujas lists more colloquial verbs than dictionaries usually do. We collected around five hundred *o(b)*– verbs in these dictionaries. Additional verbs were found in the two publicly available corpora (both around one hundred million tokens) which we used in the second step: 1) Corpus 1 (CNC): Croatian National Corpus (version 2.5),[10] and 2) Corpus 2 (CLR): Croatian Language Repository.[11]

The main source of the Bulgarian corpus used for this study is *The Bulgarian Verb* (Pashov 1966), which presents an exhaustive list of prefixed verbs in Bulgarian. His data included 299 verbs was compared with the Bulgarian National Corpus[12] and two dictionaries: *Eurodict*[13] (an online multilingual dictionary containing 60,000 Bulgarian words) and *Rečnik na bălgarskija ezik* (Bulgarian Dictionary)[14] (the details are discussed in Section 4).

We chose to compare two South Slavic languages, Bulgarian and Croatian,[15] because such a comparison contributes to the knowledge about each of them and may yield interesting, unexpected results that can be applied in other comparative studies.

The analysis is structured as follows: Section 2 presents a semantic network of the verbal prefix *o(b)*– in Croatian and Bulgarian, the central meaning and its extensions, and challenges related to defining sub–meanings of *o(b)*–. Section 3 addresses challenges in identifying additional meanings. Section 4 discusses sources of our databases (dictionaries and corpora), specifically questioning frequencies and addressing other challenges and the advantages related to using corpora. Section 5 provides some conclusions.

---

10   Available at: http://filip.ffzg.hr/bonito2/run.cgi/first_form.

11   Available at: http://riznica.ihjj.hr/.

12   Available at: http://search.dcl.bas.bg/. Currently, the corpus core consists of approximately 1.2 billion words and more than 240,000 texts.

13   Available at: http://www.eurodict.com/.

14   Available at: http://ibl.bas.bg/rbe/. This is a multivolume dictionary of Bulgarian and it is the largest and most representative thesaurus of Bulgarian. Fourteen volumes have been published so far that include a total of 112,686 lemmas. The dictionary exhaustively presents the richness of Bulgarian lexis in the last 150 years in its stylistic and functional diversity (literary vocabulary, colloquial vocabulary, dialects, terminology, old words, etc.). Croatian does not have such a dictionary source that would truly illustrate its stylistic and functional diversity.

15   We used Croatian corpora and dictionaries in this analysis, hence the term *Croatian*. Regarding prefixed verbs in general and o(b)– verbs in particular, the language varieties used in Croatia, Bosnia and Herzegovina, Montenegro, and Serbia show no significant differences. Therefore, our general claims about the semantic profile of *o(b)*– verbs in Croatian apply to all of the standard languages based on neo–štokavian.

154

2. The semantic network of the verbal prefix o(b)–: the central meaning and its extensions

We worked on the semantic network of the prefix *o*(*b*)– using a database of 564[16] Croatian verbs and 396 Bulgarian verbs[17] collected from dictionaries and corpora (see Section 4).[18] Our first aim was to analyze the prefix *o*(*b*)– and its semantic contribution to the verbs it combines with, which means semantically "isolating" the meaning of the prefix from the meanings of the verbs. However, in many situations the meaning of *o*(*b*)– is hardly separable from the meanings of the base verb; that is, their meanings overlap (see the discussion of the Overlap Hypothesis in Janda et al. 2013); for example, BG[19] *peka* 'bake, roast' – *opeka* 'bake, roast'; *tărkolja* 'roll, tumble' – *otărkolja* 'roll, tumble', Cro. *kružiti* 'encircle' – *okružiti* 'encircle'.

Second, after examining all the verbs, their dictionary descriptions, and corpus samples, we identified several frequently occurring sub–meanings. We aimed to identify a core or dominant sub–meaning for each individual verb, and to provide all the verbs in our databases with corresponding semantic labels. However, it turned out that a few sub–meanings–"move around", "surround", and "envelop"–blend in many verbs. Assigning a single semantic label is tricky because the label covers only part of a verb's complex meaning. This challenge relates to the general issue of establishing discrete lexical senses: recall Glynn's (2014: 141) recommendation of resisting the temptation to assume that discrete reified lexical senses exist.

The main point with all classifications is separating entities into groups. Ideally, one entity belongs to only one group. Cognitive linguists also try to carry out this type of categorization, although at the same time they emphasize the fuzzy borders between categories. Clear overviews indicate convincing analyses, but such overviews contrast with the cognitive organization of categories. In idealized (and artificial) classifications of meaning of a lexical unit, we reduce polysemy to a "dominant" meaning, and that reduction seems problematic to us.

Various researchers (e.g., Geeraerts 1997; Tyler & Evans 2003; Janda 2007) have posited a variety of criteria for establishing the prototypicality of a word's sense; the criterion we followed in this analysis was identifying a sense from

---

16 The Croatian database by and large contains perfective verbs (e.g., *obliti* 'pour all over'). Secondary imperfectives are not included (e.g., *oblijevati* is not included) unless they show some difference in meaning in relation to the prefixed perfective (which is the case with, e.g., *obigrati* 'dance/go around' and *obigravati* 'solicit, annoy, ply'), in which case both are included. However, similar cases are rare.

17 The Bulgarian database does not include secondary imperfective verbs. We only included prefixed perfective verbs.

18 This section provides a condensed overview. The semantic networks and the relation of the individual sub–meanings are extensively discussed in two submitted publications (Šarić & Mikolić (forthcoming); Endresen et al. (forthcoming)).

19 Abbreviations used in this study: BG = Bulgarian, BulNC = Bulgarian National Corpus, Cro. = Croatian, CNC = Croatian National Corpus, CLR = Croatian Language Repository, HJP = Hrvatski jezični portal.

155

which most others can be derived. According to that criterion, a spatial image schema "around"–the meaning of circular movement–is the prototype in the semantic network of *o*(*b*)– in Croatian and Bulgarian (as expected, this coincides with findings for other Slavic languages; see Twardzisz 1994; Dobrušina & Paillard 2001; Baydimirova 2010; Będkowska–Kopczyk & Lewandowski 2012; Janda et al. 2013; Endresen 2013; Endresen et al. (forthcoming); for Croatian, see Šarić & Mikolić (forthcoming)). In this schema, a TR prototypically moves (all) around a LM and encircles the LM. The central meaning of (an ideal and complete) circular movement experiences various metonymic and metaphorical transformations in the spatial domain, and particularly in verbs with abstract meanings. The transformations of the CIRCULAR MOVEMENT schema related to the length of the circular path are illustrated in Figure 1; that length ranges from a full circle and total encirclement to a semicircular arc–like movement. Moreover, the radius of the arc can vary considerably.

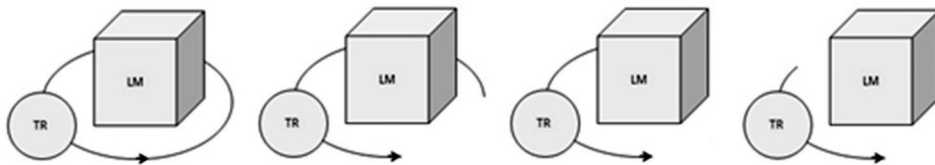Figure 1: The prototype of o(b)–: spatial image schema of circular movement and its variants/transformations

Table 1 presents the four most prominent sub–meanings of *o*(*b*)– in Bulgarian and Croatian to which the vast majority of *o*(*b*)– verbs relate. These meanings of course overlap and cannot be understood as discrete.

Table 1: Main sub–meanings of o(b)– in Croatian and Bulgarian

(1) **Circular movement: moving around (an object)** (full circle, semicircle, arc, etc.)	
Cro. *obići* 'move/go/travel (all) around', *objedriti* 'sail around'; bypass/avoid: *obići* 'avoid'; turn around: *obazreti se* 'turn around'	BG *obhodja* 'move around', *obkrača* 'pace around', *obikolja* 'go around', etc.
(2) **Surrounding** (enveloping; covering, etc.), concrete and metaphorical	
Cro. *opkoliti* 'encircle', *opisati* 'describe', *obasuti* 'shower/load with'	BG *obleja* 'pour over', *ovaljam* 'roll over', *opiša* 'describe', etc.
3) **Affecting a surface**	
Cro. *obojiti* 'paint', *obrijati* 'shave off'	BG *odera* 'scratch', *obelja* 'peel', *obrăsna* 'shave', etc.

156

4) **Imposing/acquiring a new feature** (concrete and abstract)	
Cro. *obnažiti* 'lay/strip bare', *osmjeliti* (*se*) 'become courageous', *omaciti* (*se*)[20] 'have kittens'	BG *ohladja* 'cool, chill', *oglupeja* 'become stupid', *osmelja se* 'become courageous', *okotja se* 'give birth to kittens', *obagnja se* 'give birth to a lamb', etc.

Circular movement can be concrete and/or metaphorical (both meanings coexist in some verbs; for example, Cro. *obigrati – obigravati* 'dance/go around; solicit, annoy, ply;[21] BG *omotaja – omotavam* 'wind round, entangle, involve'). Several verbs referring to circular movement in space in Croatian seem to imply Avoid (e.g., *obići*), but we do not consider Avoid to be a distinct meaning because circular movement in space inherently implies avoiding some objects inside the circle/arc, as Figure 2 illustrates:

Figure 2: Avoid: A separate meaning?

Our knowledge of the world and the context reinforce the sense Avoid with typical spatial obstacles and when one or another contextual factor indicates avoidance.

In both Croatian and Bulgarian, the central spatial meaning of circular movement directly motivates the sub–meanings Surround and Affect a surface. When moving around a LM, a TR can enclose it, contacting and affecting its boundaries, its surface, or its volume (by adding or removing). An object can be surrounded, covered, and enveloped spatially (as with, e.g., Cro. *obložiti*

---

20 *O*(*b*)– as a prefix in the semantic group of verbs meaning 'bring forth a X', where X stands for a newborn animal, is attested in relevant examples of verbs in Slavic etymological dictionaries (see, e.g., Skok 1988: 746 regarding *jăgnjiti (se)*). Existing analyses of the Slavic prefix *o*(*b*)– do not questions the existence of this prefix in verbs belonging to this semantic group (e.g., Croatian *ojanjiti* (*se*) 'give birth to a lamb', Bulgarian *okuča se* 'bring forth a puppy', etc.). An anonymous reviewer pointed to the possibility of having the ablative prefix *od*– in similar verbs in Croatian. It is true that the Croatian ablative *od*– is realized as *o*– (and *ot*–) in certain verbs, but this occurs in some specific environments only: these requirements are not met in the group of verbs represented by *omaciti (se)*. *Od*– could have been realized in this subgroup of verbs had the verbs in question originally contained the prefix *od*–. As insightfully indicated by the same anonymous reviewer, with similar verbs, a construal implying a spatial schema of ablative movement seems logical: the newborn animal "departs" from the body it was part of.

21 In some contexts, this verb also means 'avoid direct contact.' We owe this remark to an anonymous reviewer.

157

'cover, coat', BG *ogladja* 'iron, press') and metaphorically. When an activity surrounds, envelops, or covers an object, METAPHORICAL SURROUND/ENVELOP/CO-VER is realized, as illustrated by the Cro. verbs *opisati* 'describe' and *obasuti (ljubavlju)* 'shower with (love)'[22] and BG *ogledam* 'survey, examine'.

We found it difficult, if not impossible to consider 'surrounding', 'enveloping', and 'covering' as discrete senses in the spatial contexts of our verbs. Separating these senses is even more difficult in abstract contexts: this is the reason for considering surrounding, enveloping, and covering as one general category.

A consequence of covering and surrounding can be affecting a surface (or they may happen simultaneously). The sub–meaning SURROUND introduces the idea of close contact of a TR with a LM that can affect the LM's qualities or appearance. In some cases, AFFECT A SURFACE implies removal of some substance from the surface, as with Cro. *obrijati* 'shave off', *očerupati* 'pluck/pick clean', *očetkati* 'brush(up)/clean'[23] (or removal of fabrics, such as with *obnažiti* 'lay bare'), and BG *okălčam* 'cut too short/uneven', *okărša* 'break, snap', *oskubja* 'pluck/pull out'. When we cut something (e.g. *okălčam*) we remove some of its parts. When a tree or bush is trimmed (e.g. *okărša*) it loses some of its boughs and twigs. When a person's hair is plucked (e.g. *oskubja*) he or she loses a number of hairs. However, we do not regard LOSING PARTS as a distinct meaning because affecting a surface inherently implies the occurrence of some effects/marks on that surface. These effects can be a partial loss of integrity, as Figure 3 illustrates:

Figure 3: LOSING PARTS

---
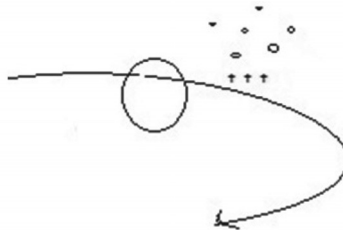
22 *Obasuti* also has a concrete meaning, as in *obasuti cvijećem* 'shower with flowers'.

23 The verbs *očerupati*, *očetkati*, and similar ones clearly illustrate overlapping components in the semantic networks of the prefixes *o(b)*– and *od*– (BG. *ot*–). The image schema that these verbs relate to is not only circular movement, but also removal. However, comprehensive etymological information (e.g., Trubačev 2001) unambiguously relates similar verbs in Slavic to the prefix *o(b)*–. A semantic overlap of parts of the semantic networks of two or more different prefixes in Slavic is not unusual, and confirms basic cognitive principles and a gradual organization of linguistic categories. Although individual prefixes often share parts of their semantic space, as a rule they do not share their central meaning. Verbs such as *očerupati* and *očetkati* do indeed suggest the ablative schema; however, Cro. *oguliti* 'peel off' and *obrijati* 'shave off' suggest the same schema, and there are no morphological reasons for not adding the prefix *od*– (i.e., nothing blocks *odguliti* and *odbrijati*) had these verbs been etymologically related to *od*–. Regarding the semantic network of the prefix *od*–/*ot*– in Croatian and Bulgarian, see Šarić and Tchizmarova (2013). A comparative analysis of *o(b)*– in Slavic that, among other things, discusses similar verbs is provided by Endresen et al. (forthcoming).

158

In other cases, affecting a surface implies APPLYING AND ATTACHING SOME-THING TO A SURFACE, as with Cro. *obojiti* 'color', *obaviti* 'envelop, wrap up', BG *obleja* 'pour over', *očapam* 'make dirty', *obvija* 'envelop, wrap up'. Similarly to LOSING PARTS, we do not consider this as a distinct meaning but just another effect when affecting a surface. Instead of having parts removed from the LM, in these cases we have parts covering the LM or added to its surface, as Figure 4 shows:

Figure 4: APPLYING AND ATTACHING SOMETHING TO A SURFACE
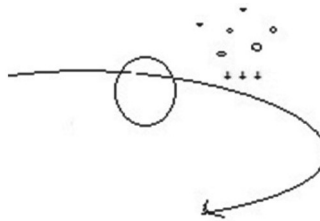
Imposing and acquiring a new feature can be concrete, as in Cro. *obnažiti* 'lay bare' and BG *oteka* 'swell', or abstract, as in Cro. *osmjeliti (se)* 'become courageous' and BG *osramja* 'put to shame'. Thus, the concept of affecting and changing applies to both the spatial and abstract domains.

Some verbs in both Croatian and Bulgarian with the meaning IMPOSE/ACQUIRE A NEW FEATURE imply changes in physical and psychological states (e.g., Cro. *opiti (se)* 'get drunk', *opametiti se* 'become wise', *okameniti se* 'become petrified, turn to stone', and BG *oslepeja* 'go blind', *osvirepeja* 'grow savage/furious', *osvestja se* 'come to one's senses, collect oneself') and are conceptually linked to the sub–meanings METAPHORICAL SURROUND/ENVELOP. In some other verbs, the meaning IMPOSE/ACQUIRE A NEW FEATURE directly relates to the meaning AFFECT A SURFACE. Being conceptually motivated by both spatial and metaphorical sub–meanings of *o(b)*–, the meaning IMPOSE/ACQUIRE A NEW FEATURE is thus incorporated well into the semantic network.

We adhere to the view that *o*– and *ob*– in spatial contexts in Croatian relate to the same spatial concept as the preposition *oko* 'around', with which they combine in their typical spatial usages. Our database supports this view. *O*– and *ob*–[24] appear in all four sub–meanings; however, *ob*– is much less frequent in the database (see Table 2). Due to the polysemy of many verbs, providing numbers for individual sub–meanings would be misleading and it is more plausible to mention (strong) tendencies: *ob*– is predominant in the concrete spatial meaning MOVE AROUND, ENCIRCLE, whereas *o*– is found in both non–spatial and spatial uses, and it is predominant in the metaphorical meaning ACQUIRE/IMPOSE A NEW FEATURE and AFFECT A SURFACE (both physically and metaphorically; very few *o(b)*– verbs belong to the last category; e.g., Cro. *obgorjeti* 'burn around/on the surface'). In these categories, *o*– and *ob*– exhibit the largest distribution differences.

---

24   References to the Croatian *ob*– in this text always include its allomorphs *oba*– and *op*–.

Following the traditional studies of prefixes (Gerov 1977: 288; *Bălgarski etimologičen rečnik* [Bulgarian Etymological Dictionary] 1995: 735), we consider *o(b)*– in Bulgarian to be a single prefix semantically related to the preposition *okolo,* which often combines with *o(b)*– verbs denoting circular movement in space. Similarly to Croatian, *ob*–[25] appears in the four sub–meanings discussed above, although it is much less frequent than *o*– in the database (see Table 2). *O*–, on the other hand, is much more frequent than *ob*– but it is not found with the meaning MOVE AROUND (AN OBJECT). There are a number of polysemous verbs that fall into more than one category. The overall tendency is that Bulgarian *ob*– prevails in the concrete spatial meaning SURROUND/ENVELOP. *O*– is prevalent in the meanings ACQUIRE/IMPOSE A NEW FEATURE and AFFECT A SURFACE.

Table 2 provides a summary of our findings related to the differences in frequency and distribution of *o*– and *ob*– in different semantic subgroups of Croatian and Bulgarian verbs.

Table 2: Cro. o– and ob– (oba–, op–), BG o– and ob– (obo–): differences in frequency and distribution (summary of findings)

Croatian	Bulgarian
Ob– (*oba*–, *op*–): much less frequent than *o*–; found in 100 out of 564 verbs (18% of all verbs)	*Ob*– (*obo*– only one verb): much less frequent than *o*–: found in 57 out of 396 verbs (14% of all the verbs)
Ob–: predominant in the concrete spatial meaning MOVE AROUND, ENCIRCLE	*Ob*–: predominant in the concrete spatial meaning SURROUND/ENVELOP
O–: predominant in the meanings ACQUIRE/IMPOSE A NEW FEATURE and AFFECT A SURFACE	*O*–: predominant in the meanings ACQUIRE/IMPOSE A NEW FEATURE and AFFECT A SURFACE

3. Semantic classification and challenges with identifying additional (peripheral) meanings

In certain cases, the semantic categorization of prefixed verbs depends on the perspective of a particular language user or researcher (who in some cases is an active user of the language analyzed). Subjectivity is a consequence of introspection and certainly an interfering factor in working out the semantic networks of prefixes.

The semantic network of *o(b)*– in Bulgarian and Croatian made on the basis of all the verbs in our databases coincides in the four central meanings (see Table 1): CIRCULAR MOVEMENT (MOVE AROUND (AN OBJECT)), SURROUNDING (ENVELOPING; covering, etc.), AFFECTING A SURFACE, and IMPOSING/ACQUIRING A NEW FEATURE. However, some verbs in our databases indicate

---

25   There is only one variant of *ob*– found in our database, namely *obo*–, and it is used in only one Bulgarian verb (*obozra* 'comprehend, encompass'); therefore we cannot talk with confidence about allomorphy in this case.

that there may exist some more sub–meanings related to a relatively limited number of verbs (see (A)–(F) below). We consider these meanings peripheral because they are observable in a few or very limited number of verbs. Some of these peripheral meanings do not coincide in Croatian and Bulgarian; that is, certain meanings are found in one language only. Compare the examples in (D)–(F):

(A) REMAIN IN PROXIMITY (Cro. *opstati*, *održati se* 'survive, hold out'; BG *oživeja* 'survive', *ostana* 'remain, stay')

(B) OVERDO (Cro. *opiti (se)* 'become/get drunk, make smb drunk'; BG *oharča se* 'spend a lot of money', *oreva* 'scream, yell blue')

(C) DO SUPERFICIALLY/HASTILY/AFFECT SLIGHTLY (Cro. *ovlažiti* 'moisten', *očešati se* 'graze, scrape, *okrznuti* 'graze, glance, scratch, sideswipe', BG *okašljam se* 'cough', *oblegna se* 'lean, rest')

(D) EQUIP (Cro. *oboružati* 'arm')

(E) VISIT MANY (Cro. *obići* 'go around, visit')

(F) AFFECT MANY (BG *oglasyja* 'make sth re–echo', *opasa* 'graze down/away')

Croatian

Very few verbs in the Croatian database are linked to (A), (B), and (D) above, and several more (although still a low number) to (C) and (E). We consider (A) and (B) "metaphorical surrounding." In the cases of the verbs in (A), the base verbs' meanings and contextual factors having to do with duration reinforce the sense REMAIN IN PROXIMITY; (B) implies acquiring a new feature, and metaphorical surrounding. Due to the evaluative stance, the sense OVERDO is related to the base verbs' meaning in contexts with a specific object (alcohol).[26] With (D), the base verb's meaning already implies EQUIP/SUPPLY. (E) arises with motion verbs, predominantly with *obići* (whose original spatial meaning is 'encircle') and its synonyms. However, a detailed examination of a sample of corpus examples in Corpus 2 (CLR) with this verb shows that the sense VISIT MANY (found in approximately 340 (40%) out of 849 examples) is dependent on the context, as illustrated by the following examples:

. . . *obišli* su sve hotele na ovdašnjem području. 'They visited all the hotels in this region.'

. . . *obišli* su mnoge europske zemlje . . . 'They visited many European countries'. (CLR)

Among the most frequent collocates of the verb *obići* in our sample are the quantifiers *sve* 'all' (the tenth collocate on the frequency list of collocates[27] for *obići*), *nekoliko* 'several; some' (fifteenth), and *više* 'some' (twenty–second). Moreover, in some contexts *obići* implies 'visit one', in some 'visit a few', and in some others 'visit

---

26 We found another verb of consumption, *jesti* 'eat' in this meaning, in a fifteenth–century source.

27 A frequency list of collocates can be obtained for each lemma included in CLR. We have examined the frequency list for the verb *obići* and the information above is based on that list.

many'. This contextually inferred meaning relates to circular movement: encircling implies looking at something from different spatial points, and seeing more than one or all sides of an object: similar settings reinforce 'examining' as an additional contextually dependent sense of 'visit one'. Circular movement also implies being at different spatial points (of the "circle" or "arc" at different times), and relates to the contextually inferred sense 'visit several/ many locations within the same broad location'.

The meaning 'visit (or, its possible consequence, affect) many' is strongly supported by other elements in the context in all our examples (see the examples with *obići* above), and so its distinct sense status is uncertain: if all the instances of *obići*, *optrčati* meaning 'visit many' have as their object plural nominals and nominals with quantifiers, such as *many* or *all,* the sense 'visit many' is clearly inferable from other contextual elements. This poses a general question as to how to isolate prefixed verbs from their context, and how to treat contextual factors, which are unavoidable when applying semantic network models.

A few Croatian verbs that imply Equip/supply illustrate the problem of separating the semantic contribution of the base verb and semantic contribution of the prefix. When a certain number of base verbs share a meaning and choose the same prefix, we tend to attribute the verbs' meaning to the prefix.[28] A prefix and a base verb can of course have overlapping meanings: SURROUND is certainly related to EQUIP/SUPPLY.

Bulgarian

Similarly to Croatian, very few Bulgarian verbs belong to the (A) Remain/ stay in proximity subcategory, which is metaphorically related to surrounding. Subcategory (B) Overdo refers to a small number of verbs in Bulgarian. They denote activities related to human bodies (eating *ojam se* 'eat too much',[29] drinking *opija se* 'get drunk', listening *oslušam se* 'listen carefully', etc.) and feelings (crying *oreva* 'scream / yell blue', shouting *ovikam* 'shout at smb rudely', whistling *osvirkam* 'catcall, boo', etc.). These verbs are united by the meaning of the prefix, which suggests 'passing a limit'. This meaning is semantically related to Acquire/impose a new feature because the TR changes its appearance or behavior as a consequence of performing the activity denoted by the base verbs.[30]

---

28   For example, several verbs in both Bulgarian and Croatian refer to giving birth to an animal (Bg. *okotja se* 'give birth to kittens', *obagnja se* 'give birth to a lamb', Croatian *ojanjiti* 'give birth to a lamb', *okoziti* 'give birth to a kid', etc., and one could tend to attribute the meaning 'bring forth . . .' to the prefix itself).

29   There is another verb in Bulgarian with the meaning 'eat too much': *prejam. Ojam se*, however, is used in cases when one ate too much of something particular in the past and he or she does not want to eat it anymore; for example, *Ojadoha se veče orlite, mrǎʧtjat se na gǎlǎbi!* 'The eagles ate too much and they are frowning upon pigeons now!' The verb has the nuance of meaning 'become choosy/particular'.

30   There are no *ob*– verbs attested in the Bulgarian database belonging to this category.

Both (F) Affect many and (C) Affect slightly are related to the category Affect a surface. They are not very productive, especially (C) Affect slightly, where the verbs once again refer to basic human activities (e.g., *okašljam (se)* 'cough', *oblegna (se)* 'lean, rest', *opusna* 'drop, blunder out'). (F) Affect many is motivated by the category Affect a surface, but in this subcategory the activity of the TR affects a number of LMs instead of just one; for example, *oglasja* 'make sth re–echo', *opasa* 'graze down/away', *obera* 'pick, rob, plunder', and so on. The meaning Affect many/all is reinforced by the context:

(Te) štjaha da pazjat dobităka njakolko časa, dokato opase oskădnata treva naokolo.' (BulNC) 'They would keep an eye on the cattle while grazing the scarce grass around'

'Grass' is usually used as an uncountable noun and it can stand for an entire field or at least a tuft of grass that the cows are eating, not just a single blade.

. . . zlovešto praštene *oglasi* celija tunel. '. . . an ominous crackle echoed throughout the tunnel' (BulNC)

The sound mentioned in the example above affects the entire space denoted by the LM.

. . . *obiram* stari ženi, fraškani s biseri. (BulNC) '. . . I mug old women that have a lot of pearls'

A thief can steal from a lot of people or a single person: . . . *obiram* veski, kogoto moga '. . . I steal from everyone I can'.

Consequently, it can be claimed that the verbs in the sub–category Affect many/all are compatible with both singular and plural LMs, depending on the context and, in the case of *oglasja* 'echo, make resound', the LM is seen as a 3D container that is affected by the verb together with all its content.

4. Corpora used: Some challenges

We now turn to our sources and the challenges related to collecting and evaluating our empirical data. These are largely related to the corpora used, but we also include dictionaries to some extent in our discussion.

4.1. Corpora: Features

We used version 2.5 of the Croatian National Corpus (Corpus 1) with about 101 million tokens. This corpus is unbalanced. [31] We noticed a predominance of newspaper texts over literary texts, with a single newspaper overshadowing a few others.[32] According to official information, this corpus is fully morphologically tagged; however, our searches do not confirm this. We noticed errors in part–of–speech tagging and in morphological tagging.

[31] The "ideal structure" that the corpus aims at, according to its internet presentation, is 74% informative texts, 23% imaginative texts, and 3% mixed texts. This information is provided by the corpus managers at http://www.hnk.ffzg.hr/struktura.html.

[32] The authors are not aware of any publicly available Croatian spoken language corpus.

It was not possible to automatically acquire a list of all the *o(b)*– verbs from Corpus 1. We obtained very erroneous lists for *ob*–, *oba*–, and *op*–,[33] but no list for *o*–.

The total size of Corpus 1 (the version we used) is comparable to the size of Corpus 2 (Croatian Language Repository).[34] Corpus 2 allows only simple searches, not lemma searches. We calculated a total number of occurrences for only a limited number of verbs that we needed for some comparisons with Corpus 1.

Corpus 1 theoretically enables lemma searches (theoretically because, in many cases examined, not all verb forms are actually included in the results; implying that some frequencies do not match reality). We checked the frequencies for our 564 verbs one by one.

The Bulgarian National Corpus incorporates several electronic corpora, namely parallel corpora, the "Brown corpus" of Bulgarian (patterned on methodology from Brown University), the Bulgarian POS annotated corpus, and so on. It reflects the state of Bulgarian (mainly in its written form) from the mid–twentieth century (1945) to the present. The kernel of the BulNC, consisting of all Bulgarian texts in the corpus, currently amounts to 979.6 million tokens, which makes it relatively large and representative. Although the compilers aimed for a relative balance of the texts, "written texts prevail significantly (91.11%), with spoken data representing only 8.89% of the tokens and being limited in variety–parliamentary proceedings, lectures, and subtitles" (Koeva et al. 2012: 89). The following styles are pointed out as included in BulNC: administrative (18.53%), popular science (4.51%), fiction (25.11%), science/administrative (1.27%), science (3.31%), journalism (38.63%), informal (0%), and informal/fiction (film subtitles, 8.61%) (Koeva et al. 2012: 91). Most of the styles are clear–cut, and two of them are complex: Informal/Fiction and Science/Administrative. The former can be defined as informal texts within fiction (subtitles), and the latter as highly specialized (scientific, e.g., medicine) texts within the administrative style. "As a result the BulNC illustrates a large number of the styles, genres and thematic domains typical for modern Bulgarian" (Koeva, Blagoeva, & Kolkovska 2010: 3680). However, according to these authors (ibid.), there are several areas for improvement in the present structure of the corpus. One problem is the small share of spoken data and that they are limited to two narrow domains. It is necessary to increase the

---

33 The lists also include parts of speech other than verbs. Different morphological forms of many verbs were listed as separate verbs. We discovered many spelling errors. Many hours of additional correction by hand and extra calculation was necessary.

34 According to information found on the internet, the CLR covers various functional domains and genres. It includes literature and other written sources from the second half of the nineteenth century onward. "Currently the corpus contains more than 100 million tokens" (Ćavar & Brozović–Rončević 2012). Official numbers as of 2014: entire corpus = 84,976,049 occurrences; 1,298,850 lemmas. Of these: (A) books = 17,628,945 occurrences; 665,828 lemmas; (B) newspapers = 67,496,544 occurrences; 916,747 lemmas. The corpus administrators (personal communication) indicated that the corpus has at least fifteen million more occurrences than indicated in the official numbers in 2014 because new material was added. Newspaper texts dominate also in this corpus (only 20.7% is books; i.e., literature and school books).

number and variety of scientific texts and other nonfiction, as well as of poetry, drama, and other fiction samples.

Having in mind such important indicators as representativeness and balance, showing the correlation between text distribution and lexis diversity in the BulNC and the actual population of Bulgarian (printed and electronic) texts for a given period (direct correlation with the population of texts) is not appropriate. Due to the large percentage of media texts, their representativeness will be stronger than good balancing (Przepiórkowski et al., 2008).

As a result of corpus data research, ninety–six new verbs were added to Pashov's list and the outdated verbs with no hits in both online resources were excluded. At present the database contains 396 *o(b)*– verbs.

Multiple prefixation is a typical feature of Slavic languages. "Up to seven prefixes can stack on a single verbal root in Bulgarian" (Istratkova 2004).[35] *O(b)*– is not very productive in combination with other prefixes. As the first prefix in a prefixed verb, *o*– takes part in an outstanding group of verbs, in which it combines with another prefix: *bez*– 'without'. This group stands out because there are twenty–eight verbs with double prefixation *o*– + *bez*– in our database (e.g., *obezvodnjavam* (*se*)) 'dehydrate'. However, only eleven of them appear with a frequency greater than fifty.

Included in the Croatian database are six verbs with the prefix combination *o–bez*– from dictionaries: *obescijeniti* 'depreciate', *obeshrabriti* 'discourage', *obeskličiti* 'disinfect' *obezglaviti (se)* 'behead; confuse', *obezličiti* 'deprive of individual traits', *obeznaniti se* 'faint': of these, two have a frequency higher than one hundred in Corpus 1, and two are not attested. The results for *o–bez*– (and *o–bes–*, *o–beš–*) in Corpus 2 yielded additional fairly frequent verbs (perhaps around twenty altogether; the number of verbs could not be detected automatically). Six verbs found in the corpus are very or relatively frequent: *obeštetiti* 'indemnify' (653 occurrences of 754 for *obeš–*), *obeščastiti* 'dishonor'; *obeshrabriti*, *obespraviti* 'deprive of rights', *obezvrijediti* 'depreciate', and *obezglaviti*. Other verbs are much less frequent (e.g., *obeznaniti:* 45 of 935 occurrences for *obez–*).[36]

Another frequent prefix combination in Croatian is *o–ne–*. Eight verbs with this combination are included in the database. We intuitively assume that this group is larger than *o–bez–*. Unfortunately, the prefix combination *o–ne–* in Corpus 2 could not be examined; the results include all words beginning in *one–*, and one would have to manually search for concrete individual verbs.

4.2. Frequencies in the corpora: Croatian

Twenty–three percent of verbs from the database (128 verbs) found in other sources were not found in Corpus 1.[37] One hundred seventy verbs have more than one hundred occurrences. Ninety–seven verbs from our initial database occur fewer than five times; of these, forty verbs occur only once.

---

35 Istratkova (2004), however, does not study *o(b)*.

36 The search results contain verbal nouns, and so the numbers are relative.

37 However, as mentioned, not all lemma searches with no hits are reliable. In some cases, lemma searches showed no hits, but some verb forms occurred in simple word form searches.

We compared the frequency results for eleven randomly chosen low–frequency verbs in both corpora to examine how reliable corpora are with regard to usage frequencies they reflect, compared to each other, and juxtaposed with the internet, which reflects everyday language and includes a variety of different sources. We found that a few verbs have similar frequencies in both corpora; for example, the verb *obići* exhibits similar frequency: 2,132 in CNC and 2,302 in CLR; compare also *oklesati, ostrviti se, obnarodovati, omaciti (se), opkopati* in Table 3. However, the majority are much more frequent in Corpus 2. The underlined verbs show the greatest frequency differences (for *omiliti*, the difference is 1:39 and, for *osiliti se*, 1:27). This comparison indicates that the frequency of some verbs in Corpus 1 does not reflect their frequency in other corpora (most probably not in everyday language either). The frequencies are corpus–specific because they depend on the text types included in a particular corpus. In this particular case, frequencies are related to the specific lexical preferences of the predominant sources (newspapers) in the corpus. One should by no means take the results as absolute.[38]

Table 3: Verbs with low frequency in CNC and CLR: comparison

	Corpus 1 CNC v2.5[39]	Corpus 2 CLR
oklesati 'dress/carve, chisel'	1	2
opojiti (se) 'intoxicate'	1	33
ostrviti se 'grow wild'	1	1
ozepsti 'become very cold'	1	37
omrknuti 'grow dark'	1	19
obigrati 'dance around, fool'	2	15
obnarodovati 'make known'	2	2
okvasiti 'moisten'	2	20
opsjesti 'besiege'	2	44
onjušiti 'sniff all over'	2	24
ogranuti 'shine/break through'	3	26
omaciti (se) 'have kittens'	3	2
omiliti 'make dear, sympathize'	3	116
optrčati 'run around'	3	26
osiliti se 'become despotic'	3	82
opkopati 'dig a trench around'	3	5

---

38   We express some reservations about presenting any exact statistics (see Section 4.4.).
39   At present, there are two versions of CNC. We checked frequencies in version 2.5. Version 3.0 is supposed to contain additional material (216.8 million tokens). However, we noticed that the older version, 2.5, contains tokens that the newer version does not contain. For example, *ogranuti* (see Table 3) is not found in version 3.0.

166

Are the verbs with low frequency in Corpus 1 (and Corpus 2) really infrequent in everyday language usage? The answer for some of them is probably yes, but some other verbs intuitively seem fairly frequent. Without having a reliable corpus of spoken language and a possibility of spoken corpus analysis, we can only make assumptions. Research indicates that there are considerable differences in the frequency of certain words in written and spoken corpora (see, e.g., Allwood 1998).

The internet is a good source of informal language, and so we wanted to compare corpus frequencies with some internet frequency indicators: we checked frequencies for some verbs using the Google search engine (being aware of all the problems related to Google searches, such as multiple occurrences of the same example, the fact that Google does not list more than 1,000 results for any query,[40] etc.) Table 4 shows some results (August 2014). We searched for only one morphological form: the singular *l*–participle for a few verbs (feminine for the verb *omaciti,* masculine for all other verbs). The results in the CNC column apply to lemma searches.

Table 4: Some infrequent verbs in Corpus 1 and Google hits

Verb forms	CNC (lemma search)	Google (approx.)
obigrao 'danced around'	2	92
opkopao 'dug a trench around'	2	31
omrknuo 'grew dark'	2	1,360
omacila 'had kittens'	3	15,800
optrčao 'ran around'	3	5,150

Table 5 presents results for several examples that show no occurrences in CNC.

Table 5: Some verbs with no occurrences in Corpus 1 (CNC) vs. Google

Verb forms	CNC	Google
okolčati (+ *okolčao, okolčala*) 'stake around'	0	15
okrastati (+ *okrastao, okrastala*) 'become covered with scabs'	0	32
okljaštrio [*okljaštriti]* 'prune'	0	9,850
okozila [okoziti] 'have a kid'	0	1,220

As can be seen, the results vary. Some verbs with a very specific meaning do not have many Google hits, whereas some verbs occur relatively frequently

---

40   http://www.googleguide.com/last_results_page.html

on the internet. The results for *omacila*, for example, reflect numerous discussions about kittens in online forums.

4.3. Frequencies: Bulgarian

The initial unabridged database amounted to 396 *o(b)–* verbs, 299 from Pashov's book and ninety–seven from dictionary sources. Sixty–six of them (16%) do not occur in BulNC. Thirty–eight verbs in the final database have less than five occurrences; of these, eight verbs occur only once. One hundred fifty–eight verbs have more than one hundred occurrences. However, these numbers can be considered relative because we found that the lemma search of the Bulgarian National corpus is inoperative. Although the compilers of the corpus claim that it is equipped with various types of monolingual annotation,[41] such as tokenization, sentence splitting, lemmatization, word sense annotation, and so on, a manual check showed that the frequency results correspond only to the particular token in the search field. This prompted us to check the frequency of the different forms of a single verb in order to identify the forms most frequently used. The results we arrived at distinguished the forms for first person singular and third person singular. Present tense was the most widely used. These were the tokens we checked manually for each verb and then we summed up the results. The final numbers we use in the corpus show the results only for first person singular and third person singular present for each *o(b)–* prefixed verb in Bulgarian. They exclude the forms for the second person singular and first, second, and third persons plural present tense, as well as the forms for the past, which makes the results approximate but representative because the first person singular and third person singular, present tense, exhibit the highest frequency of use.

Another deficiency of the corpus is homonymy. The search engine cannot differentiate among homonyms when they belong to the same part of speech; for example, the verbs *ozdraveja* 'become healthier' and *ozdravja* 'make healthy'. The past tense form of *ozdraveja* in the third person singular is *ozdravja*, which is a homograph of the second verb. These could only be sorted manually.[42] The same refers to another pair of verbs, *oprosteja* 'become stupid' and *oprostja* 'make simpler', with the same homographic problem.

A detailed study of the examples of *o(b)–* verbs revealed a number of repeated sentences in the search results. This is another proof of our assumption that the numbers in the search results cannot be regarded as absolute and final.

The corpus cannot be of much help when we try to set apart concrete (spatial) from metaphorical uses. For instance, *omotaja* means to wind a thread around a person or thing and, accordingly, it is classified in the subcategory Surround/ envelop. In its metaphorical use, *omotaja* means 'entangle/involve'

---

41   http://dcl.bas.bg/en/corpora_en.html.

42   The same applies to Croatian homographs (e.g., *òbaviti* 'accomplish' and *obàviti* 'wrap' are not distinguished). The corpus does not differentiate words with different accents that are otherwise identical. The results for these two verbs were sorted manually.

168

and the two uses, concrete and metaphorical, can only be discerned in context. Similar cases of verbs impossible to discriminate using the corpus search engine are the following (Table 6):

Table 6: Polysemous verbs with both spatial and metaphorical meanings

Verb	Gloss	Sub category	Verb	Gloss	Sub category
opetnja	'smear, soil'	AFFECT A SURFACE	*opetnja*	'throw mud at smb'	SURROUND, METAPH.
oslušam se	'listen for, give an ear'	OVERDO	*oslušam se*	'drag one's feet, disobey'	ACQUIRE A QUALITY
očernja	'make dirty'	AFFECT A SURFACE	*očernja*	'blacken, smear, defame'	ACQUIRE A QUALITY

As mentioned previously, thirty–eight verbs in the database occur fewer than five times in BulNC. Some of these thirty–eight verbs are not found in *Eurodict* and *Rečnik na bălgarskija ezik*, the dictionaries we used for reference. We decided to use Google to check those low–frequency verbs that are not present in the dictionaries; we took thirteen low–frequency verbs from the corpus at random (see Table 7).

Table 7: Some verbs with low frequency in the BG National Corpus and Google hits

Verb forms	BulNC	Google (approx.)
obkiča 'adorn, decorate'	4	2,360
obintovam 'bandage all over'	3	98
ogreba 'scrape up, scoop up'	3	5,300
oklepja 'make dirty'	3	189
obrimča 'stitch'	2	234
ogruham 'pound, mortar'	2	773
okărpja 'patch up, repair all'	2	39
obkopaja 'dig, hoe'	2	400
obriša 'wipe out'	1	5,320
odrănkam (se) 'take smb's money, drink up'	1	813
olepja 'stick all over'	1	1,880
odremja se 'become drowsy'	0	582
okalpazanja se 'spoil, go bad'	0	190
okirlivja 'turn dirty'	0	105

169

Table 7 shows that these verbs are much more frequent on the internet (the results belong only to the tokens listed in the table). This comparison displays a significant difference between the corpus and Google frequencies (February 2015). Similarly to Croatian, the low–frequency verbs in the table belong mainly to the everyday language and informal register that are predominant in the majority of internet sources. The national corpus, on the other hand, contains only 8.61% spoken texts (film subtitles).

4.4. Corpora: what do they make possible?

We now turn to the advantages of using corpora. Our analysis suggests that the numbers we obtain when analyzing corpora are not absolute but only approximate and this fact should be considered when making conclusions. The most frequent verbs in Corpus 1 and BulNC are in Table 8. Some of the most frequent verbs in CNC are obviously related to newspaper language (e.g., the second Croatian verb, *objaviti* 'make known').

Table 8: The twenty most frequent verbs in Corpus 1 (CNC) and in BulNC

Verbs	Frequency in CNC (v2.5)	Verbs	Frequency in BulNC
1. *ostati* 'stay behind'	56,157 (551.2 per million)	*ostana* 'remain, stay'	79,454
2. *objaviti* 'make known'	27,449 (269.4 per million)	*objasnja* 'clarify, make clear'	42,246
3. *održati (se)* 'hold out'	22,543 (221.3 per million)	*ogledam* 'survey, examine'	33,827
4. *ostvariti (se)* 'realize'	21,704 (213.0 per million)	*okaža (se)* 'turn out'	25,717
5. *osigurati (se)* 'insure'	21,186 (207.9 per million)	*opredelja* 'define, determine'	20,701
6. *omogućiti* 'enable'	13,541 (132.9 per million)	*osvobodja* 'liberate'	14,069
7. *osvojiti* 'conquer'	12,378 (121.5 per million)	*opravja (se)* 'put in order, set right'	12,449
8. *ocijeniti* 'evaluate'	10,658 (104.6 per million)	*objavja* 'announce'	11,553
9. *ostaviti (se)* 'leave; abandon'	10,657 (104.6 per million)	*osăznaja (se)* 'come to one's senses, collect oneself'	11,307
10. *obaviti* 'get sth done'	10,628*	*obljagam se* 'lean, rest'	9,172

170

11. *objasniti* 'explain'	10,018 (98.3 per million)	*opitam* 'try'	7,503
12. *osnovati* 'establish'	9,098 (89.3 per million)	*oženja (se)* 'get married'	7,187
13. *obavijestiti* 'inform'	7,024 (68.9 per million)	*ozova se* 'find o.s., end up'	7,033
14. *okupiti (se)* 'bring together'	6,607 (64.8 per million)	*obleka* 'put clothes on'	4,838
15. *osuditi* 'condemn'	6,354 (62.4 per million)	*obărkam* 'make a mistake'	4,153
16. *odobriti* 'approve'	6,168 (60.5 per million)	*osăštestvja* 'bring about, realize; come true'	3,877
17. *obećati* 'promise'	6,106 (59.9 per million)	*obhvaštam* 'embrace, envelop'	3,169
18. *optužiti* 'accuse'	6,012 (59.9 per million)	*ovladeja* 'master, gain command of'	3,049
19. *osjetiti* 'feel'	5,565 (54.6 per million)	*osmelja se* 'dare'	2,909
20. *ozlijediti (se)* 'injure'	5,532 (54.3 per million)	*očenja* 'value, evaluate, rate'	2,789

*The frequency per million could not be obtained for *obaviti*.

There is an interesting observation resulting from the comparison of the two databases. Four verbs equivalent in their meaning are among the most frequent in both databases: *ostati – ostana* (first in Cro. and in BG); *objaviti – objavja* (second in Cro., eighth in BG), *objasniti – objasnja* (eleventh in Cro, second in BG), and *ocijeniti – očenja* (eighth in Cro., twentieth in BG). This fact reveals certain common contemporary tendencies (of the predominant sources included in the corpora) in the two languages.

Regarding Croatian, as indicated, a few newspapers are predominant in Corpus 1, and so the most frequent *o(b)*– verbs reflect the preferences of these newspapers, not the salience of these verbs in general language use or in cognition. Balanced corpora can better help us determine how cognitively prominent a certain verb is, whereas unbalanced corpora will provide a "false" image. In our case, it appears more promising to work with different corpora (if available) in combination with dictionaries that together provide better insight into the actual use of prefixed verbs, and their most prominent and rarely occurring meanings. However, the "accuracy" of the results we obtain is highly dependent on the texts that are included in the corpora and the text sources used by dictionary compilers.

171

Regarding Bulgarian, the Bulgarian National Corpus was initially a compilation of the Written Bulgarian Archive, which represents 55.95% of the entire corpus. Later, two domain–specific corpora were included; namely, a corpus of medical administrative texts and film subtitles amounting respectively to 1.27% and 8.61% of BulNC. A large amount of news data in the Written Bulgarian Archive were provided by the publishers of various Bulgarian newspapers. Due to a lack of details about the corpus structure, the exact amount of newspaper texts included could not be determined.

As indicated, we have reservations about classifying both Cro. and BG *o(b)*– verbs into several meaning groups and providing statistical evidence for how many verbs are attested for these individual sub–meanings because of the verbs' polysemy and overlapping meanings (e.g., Affect a surface and Surround overlap, as well as the meanings Affect a surface and Acquire a new feature).

We believe, however, that it does make sense to examine how often a single sub–meaning occurs: some sub–meanings are certainly related to a great number of verbs, and some to a few verbs only. Salient meanings of words are the meanings we encounter most often, and these meanings stand out as most prominent and accessible in our minds. For instance, the verb *opetnja* 'smear, soil' has a metaphorical meaning: *opetnja* 'throw mud on smb'. As BulNC shows, the metaphorical meaning is more salient because it is the one more widely used (see Table 9):

Table 9: Opetnja: spatial and metaphorical meaning with frequencies in BulNC

Verb	Gloss	Frequency	Subcategory	Context: concrete vs. abstract LMs
opetnja	'smear, soil'	5	affect a surface	hands with blood, house, white fur, the floor, bedding, environment
opetnja	'throw mud on smb'	115	surround, metaphorically	one's name, honor, reputation, memory, innocence, soul, holiness, democracy, flag, coat of arms, national anthem, peace

The salience of the metaphorical use is augmented by the limited number of spatial LMs the verb appears with. The examples in the corpus show that the verb participates in specific contexts and these are remembered as chunks. These high–frequency multiword collocations are stored as conventionalized form–meaning pairs, according to construction grammar approaches (Goldberg 2003), which makes them better processed and better remembered. These metaphorical uses override the spatial meaning of the verb and prove

to be more salient. The same pattern can be followed, for instance, with the Croatian verb *ocrniti*. HJP lists 'blacken, paint black' as the verb's first meaning, and 'slander, calumniate; denigrate, asperse, vilify' as the second.[43] The examples with this verb in Corpus 2 (206)[44] are all metaphorical except for one, implying a metaphorical surrounding. In the only example suggesting the meaning 'blacken, paint black', the active participle is used as an adjective (*ocrnio od prašine* 'blackened from dust').

Large, well–balanced corpora theoretically contain more verbs than listed in dictionaries. However, the Croatian corpora used for this research are not well balanced. The number of verbs initially found in the dictionary sources is much larger than the number attested in Corpus 1. Yet the corpora revealed some o(b)– verbs not listed in dictionaries: in incidental searches we found new verbs in both corpora, and we assume that the corpora contain many more. For example, we found the verb *obgorjeti* in Corpus 2 meaning 'burn around; burn on the surface'.[45] Unfortunately, in their publicly available form, the Croatian corpora do not allow extraction of all the verbs with the prefix o(b)–, and so we do not know how many verbs the dictionaries do not register (which we consequently now lack in our database).

In the Bulgarian National Corpus, we found fifty verbs that were not listed in the dictionaries. Similarly to Croatian, the total number of *o(b)*– verbs cannot be extracted from the corpus because searching for the prefix will yield not only verbs, but also nouns, adjectives, and adverbs as results.

Corpora, especially large and balanced ones, enable identification of verbs that are frequent and cognitively most prominent from a language user's perspective. They show which meanings are dominant (most frequently realized), which are outdated, and which new meanings (not attested in dictionaries) emerge. Some of that information is also provided by smaller and less balanced corpora. For example, a detailed examination of examples in Corpus 2 (Cro.) has shown that the verb *obići* means 'outdo' in older sources. Contemporary dictionaries do not list this meaning, nor can it be found in contemporary usage examples. A corresponding example is found in BulNC. The verb *ojam se*, which also has a doublet with *u–*, *ujam se*, is represented with two meanings in the corpus 'overdo, eat too much' and 'become choosy/particular'. Only the second meaning is mentioned in the *Eurodict* online dictionary.

The verb *ogruham* 'hull, mortar' is used in older texts with its spatial meaning 'to hull wheat'. Later on it developed a metaphorical meaning 'to beat, pound', which is attested in BulNC, but the spatial one is not. Neither of the two meanings is found in the dictionaries.

Croatian corpora examples also show that *obigrati* 'dance around' is used as an expressive variant of *obići* 'go/travel around' in contexts suggesting the

---

43   See HJP: http://hjp.novi–liber.hr/index.php?show=search_by_id&id=eFtlWBY%3D&keyword =ocrniti.

44   Search of March 3rd, 2015. http://riznica.ihjj.hr/philocgi–bin/showrest_?kwic.6.1.9890.0.205. Cijelihr.

45   . . . nastalo je 612 šumskih požara, obgorjela površina iznosi 45.657,37 hektara 'in 612 forest fires, 45,657.37 hectares were partially burned' (CLR, contemporary example, newspapers).

scenario Visit many. HJP does not provide this meaning at all; *obigrati* is cross–referenced to *obigravati*, meaning *salijetati koga* 'entreat, pester'.[46]

The corpora used also show that:

(A) Some verbs prefixed with *o(b)*– do not exist in contemporary usage, but do in older sources; for example, the factitive Croatian verb *olijepiti* 'become beautiful'.[47] Similarly, BG *obruša* 'ruin, destroy' is found in older texts but not in contemporary ones.[48]

(B) The verbs that dictionaries (and our intuition) relate to a concrete spatial meaning only also have abstract, metaphorical meanings and occur with abstract trajectors and/or landmarks. Corpora examples thus show that individual verbs have more meanings than dictionaries indicate, but in our case this cannot be determined automatically. This finding is a result of detailed examination of examples obtained for some lemmas. For example, we found that *obgrliti*[49] 'embrace' is not related solely to a concrete spatial meaning, as dictionaries suggest. It appears with abstract TRs and/or abstract LMs (*polutami . . . koja posjetitelja obgrli pri ulasku* 'semidarkness . . . that embraces a visitor upon entrance' (CNC)). The same applies to the verbs *oblijepiti* 'glue all over' and *obastrijeti* 'drape over'. Likewise, the Bulgarian verb *obvărža* 'bind, tie down' is not only related to inanimate LMs as the dictionaries suggest but, according to the corpus examples, it can also be accompanied by animate LMs and can be used as a synonym of the verb 'marry'.

(C) The corpora also show that some verbs with *o(b)*– appear in a concrete meaning, although dictionaries define only an abstract meaning: in our dictionary source, we found the verb *ophoditi se* (Cro.) in the meaning 'behave',[50] but not *ophoditi* in the meaning 'go around'. However, this meaning is realized in contemporary usage and attested in the corpora.[51]

(D) Some common verbs prefixed with *o*– have (less frequent or archaic) synonyms in verbs prefixed with *ob*–: Cro. *obgraditi* 'fence in; enclose' is attested in Corpus 2 (six occurrences (e. g., . . . *nije svijet bedemima **obgrađen*** '. . . the world is not fenced in by walls' (CLR)). The much more frequent contemporary synonym is *ograditi*. The pair *ograždam* – *obgraždam* 'enclose, fence off' is present in the Bulgarian National Corpus. Both of them are widely used (with 1,123 and 238 examples, respectively).

———————

46 http://hjp.novi–liber.hr/index.php?show=search_by_id&id=eFxhXBk%3D&keyword=obigrava ti.

47 Example from CLR: . . . *u milosti oblijepi se, dokle **olijepi**. (maže se dok ne postane* lijepa 'She applies cosmetics until she becomes beautiful').

48 A study of historical changes in the use of prefixes requires a separate analysis, and so does a potential revival of archaic forms in contemporary language usage. Specially designed corpora are needed for such an analysis. The ones available that we used cannot provide systematic information on these matters.

49 HJP: http://hjp.novi–liber.hr/index.php?show=search_by_id&id=eFxhXRc%3D&keyword=ob grliti..

50 See HJP: http://hjp.novi–liber.hr/index.php?show=search_by_id&id=eFliXhk%3D&keyword= ophoditi+se.

51 This meaning is attested in the example *U zaštitarskoj ulozi robot će **ophoditi** kućom* . . . 'In his role of a security guard, the robot will walk around the house . . .' (CLR, newspaper sub–corpus).

174

CLR indicates that the frequent Croatian verb *okružiti* 'encircle' has an archaic synonym in *obkružiti* (three occurrences; e.g., *žarkiem suncem* **obkružena**, CLR). Our Google search resulted in 213 hits for the infinitive *opkružiti*. Some internet examples indicate that there is a slight difference in the spatial scenarios of *opkružiti* and *okružiti*, but this has to be examined in greater detail.[52] The same pair of prefixed verbs exists in Bulgarian: *okrăža – obkrăža* 'surround, encircle'. Their frequency in BulNC, however, appears to be opposite to that of the Croatian verbs: *obkrăža* is much more frequent (268 examples) than *okrăža* (sixteen examples).

(E) Corpora show that some meanings listed in dictionaries are very infrequent: this is the case with, for instance, Cro. *obnositi* and its concrete spatial meaning 'encircle something carrying something else around'.[53] This meaning is observable in only one of four occurrences in CLR (*Vjetar* **obno-si** *pjesme okolo i na pregršti baca u kavanu* 'The wind carries songs around and tosses a handful of them into the café'). The three other examples with *obnositi* contain the words *čast* 'honor' and *dužnost* 'duty', as is usual in contexts with the imperfective *obnašati* ("carry" honor/duty).[54] Accordingly, the Bulgarian verbs *okirlivja* 'turn dirty' and *oskoteja* 'become like cattle in habits, intellect, behavior' are listed in the dictionaries but no examples with such verbs were attested in the corpus, and some other verbs that are not listed in dictionaries exist in BulNC (e.g., *obkrača* 'pace out/off').

Corpora and dictionaries as a rule complement each other because large well–balanced corpora of contemporary language represent the language in use, whereas dictionaries focused more on preserving the richness of the language, including passive vocabulary and obsolete forms.

(F) A detailed examination of the corpora examples could show possible fine–grained semantic differences between verbs that seem to be and/or are synonymous according to dictionaries; for example, between Cro. *oblizati* and *olizati* 'lick all over'; *objužiti* and *ojužiti* (synonyms according to HJP) 'thaw'; and *objesiti* and *ovjesiti* (synonyms according to HJP) 'hang'. Similar synonymous pairs are attested by Bulgarian dictionaries as well: *okiča – obkiča* 'adorn, decorate'; *olepja – oblepja* 'stick all over'; *oreža – obreža* 'trim, cut off'. Detailed research on these verb pairs, however, requires further work with additional sources of naturally occurring examples, which will be a subject of a separate analysis.

---

52 For example, *Trebaju olovkom* **opkružiti** *one zemlje za koje smatraju da će . . . 2020. g. biti članica Evropske unije*. 'They should circle with a pen those countries that they believe will . . . be members of the European Union in 2020' (INT).

53 See HJP: http://hjp.novi–liber.hr/index.php?show=search_by_id&id=eFxvXhY%3D&keyword=obnositi.

54 See HJP: http://hjp.novi–liber.hr/index.php?show=search_by_id&id=eFxvXxg%3D&keyword=obna%C5%A1ati.

5. Conclusion

Our study followed one of the central tenets of cognitive linguistics; namely, its fundamentally usage–based orientation. We used dictionaries because they complement our corpora, as shown in this analysis, and as a rule they also rely on sources that illustrate at least some language registers. Language is regarded as a stock of dynamic constructions that are constantly updated by and in turn, adapting to language use. Corpora are repositories of actual language usage. Large corpora should contain more verbs than dictionaries, but this was not always the case with our corpora. Large and well–balanced corpora help us see how cognitively prominent a certain verb is, but not all of the corpora used here can be claimed to do so.

We have used available corpora in our study to point to the advantages of such an approach, but also its limitations. Large and well–balanced corpora show which meanings are the most outstanding, and which occur rarely or are outdated. Our corpora also indicated outdated meanings (e.g., *obići*, 'outdo' in Cro.) and new emerging meanings by displaying prefixed verbs in context. What is more, our corpora show that some verbs have meanings (abstract or concrete) that dictionaries do not register (e.g., Cro. *obgrliti* 'embrace' and BG *obvărža* 'tie all over').

In our research, the corpora additionally revealed some infrequent synonyms (with *ob*–) of some frequent verbs (with *o*–) (e.g., Cro. *obgraditi* and *ograditi* 'fence in; enclose'; BG *okiča* and *obkiča* 'adorn, decorate') or, vice versa, frequent verbs with *ob*– corresponding to infrequent verbs with *o*– attached to the same verb base (e.g., BG *okrăža* and *obkrăža* 'surround, encircle'; *olepja* and *oblepja* 'stick all over').

Furthermore, our analysis has shown that some meanings listed in dictionaries turn out to be very infrequent in the corpora (e.g., Cro. *obnositi* 'encircle carrying sth around'; BG *okirlivja* 'turn dirty', *oskoteja* 'become like cattle in habits, intellect, behavior'). Some other verbs, which are not listed in dictionaries, exist in the corpora (e.g., Cro. *ophoditi* 'go around'; BG *obkrača* 'pace out/off'). This indicates that dictionaries and corpora complement and enrich each other, and that a combination of both is advisable, especially when analyzing languages without large, well–balanced corpora.

Some specific details we concentrated on in this study are partially related to the imperfect nature of our corpora, and we aimed to show that, due to their structure, one should not aim to present the numbers they indicate as definite. Frequencies in a particular corpus reflect lexical preferences of the specific sources included, and not absolute frequencies in language in general. We are aware of the fact that there are no "perfect" corpora, but there are certainly differences in the volume and types of texts they include (balanced and less–balanced structures), which affects their usefulness. Although far from perfect, our corpora proved useful for semantic analyses. For example, they provided us with examples of all the semantic categories relevant for the semantic network of *o(b)*– and we do not expect to come across any other categories. We addressed other advantages in preceding sections. However, we

176

believe that one should be cautious when providing statistical information–and when basing semantic analyses in their entirety on such information–especially if that information is based on small and unbalanced corpora.

References

Allwood J. (1998). Some Frequency based Differences between Spoken and Written Swedish. *Proceedings of the 16th Scandinavian Conference of Linguistics*, Department of Linguistics, University of Turku. http://www.ling.gu.se/projekt/old_tal/Nordicfreq.html

Andrejčin L. (1944). *Osnovna bălgarska gramatika.* Sofia: Hemus.

Babić S. (1986). *Tvorba riječi u hrvatskom književnom jeziku: nacrt za gramatiku.* Jugoslavenska akademija znanosti i umjetnosti.

Barlow M., S. Kemmer. (2000). Introduction. Barlow M., S. Kemmer, eds. *Usage–Based Models of Language*. Stanford: Center for the Study of Language and Information Press.

Baydimirova A. (2010). *Russian aspectual prefixes O, OB and OBO: A case study of allomorphy.* Master's thesis. Universitetet i Tromsø. http://www.ub.uit.no/munin/handle/10037/2767

Będkowska–Kopczyk A., W. Lewandowski. (2012). OB– in Slovene and Polish. A Comparative Cognitive Analysis, Presentation at the SCLC 2012, Zagreb.

Belaj B. (2008). *Jezik, prostor i konceptualizacija: Shematična značenja hrvatskih glagolskih prefiksa.* Osijek: Sveučilište Josipa Jurja Strossmayera, Filozofski fakultet.

Ćavar D., D. Brozović Rončević. (2012). Riznica: The Croatian language corpus. http://cavar.me/damir/resources/CLC–Slavicorp.pdf.

Dobrušina E., D. Paillard. (2001). *Russkie pristavki: mnogoznačnost' i semantičeskoe edinstvo.* Moscow: Russkie slovari.

Endresen A. (2013). Samostojatel'nie morfemy ili pozicionnie varianty? Morfologičeskij status ruskix pristavok o– i ob– v svete novyx dannyx: korpus i eksperiment. *Voprosy jazykoznanija. Moscow.* 5, 89–125.

Endresen A., L. Šarić, S. Nedelcheva. (forthcoming) Cognitive Linguistics and Typology: Comparing Polysemous Prefixes O(B)– Across Slavic.

Geeraerts D. (1997). *Diachronic Prototype Semantics: A Contribution to Historical Lexicology.* Oxford: Clarendon Press.

Glynn D., K. Fischer, eds. (2010). *Quantitative Methods in Cognitive Semantics: Corpus–driven Approaches*. Berlin/New York: Mouton de Gryuter, 43–61.

Glynn D. (2014). The many uses of run: Corpus methods and Socio–Cognitive Semantics. Glynn, D., J. A. Robinson, eds. *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy.* John Benjamins Publishing Company, viii, 117–144.

Goldberg A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences* 7: 219–224.

Istratkova V. (2004). On multiple prefixation in Bulgarian. *Nordlyd* 32.2, Special Issue on Slavic Prefixes: 301–321.

Ivanova K. (1966). *Desemantizaciya na glagolnite predstavki v săvremenniya bălgarski knižoven ezik.* Sofia: Bulgarian Academy of Sciences.

Janda L. A. (1986). *A Semantic Analysis of the Russian Verbal Prefixes za–, pere–, do–, and ot–.* Munich: Verlag Otto Sagner.

Janda L. A. (2013). *Cognitive Linguistics – The Quantitative Turn. The Essential Reader.* Mouton.

Janda, L. A., A. Endresen, J. Kuznetsova, O. Lyashevskaya, A. Makarova, T. Nesset, S. Sokolova. (2013). Why Russian Aspectual Prefixes Aren't Empty. Prefixes as Verb Classifiers. *Slavica.*

Janda, L. A. (2007). Aspectual clusters of Russian verbs. *Studies in Language* 31.3: 607–648.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason.* Chicago: University of Chicago Press.

177

Koeva Sv., D. Blagoeva, S. Kolkovska. (2010). Bulgarian National Corpus Project. Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Calzolari N., K. Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, M. Rosner, D. Tapias eds., Valletta, European Language Resources Association (ELRA), 3678–3684.

Koeva Sv., I. Stoyanova, S. Leseva, Ts. Dimitrova, R. Dekova, E. Tarpomanova. (2012). The Bulgarian National Corpus: Theory and Practice in Corpus Design. *Journal of Language Modelling* Vol 0, No 1: 65–110.

Konstantinova V. (1982). *Predlozite v bălgarskata gramatična literatura*. Bulgarian Academy of Sciences, Sofia.

Kostov N. (1939). *Bălgarska gramatika*. Sofia.

Lakoff G. (1987). *Women, fire, and dangerous things: what categories reveal about the mind*. University of Chicago Press.

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

Langacker R. W. (1999). *Grammar and Conceptualization*. Berlin: Mouton de Gruyter.

Nedelcheva S. (2010). On the semantics of the preposition ZA and the prefix ZA– in Bulgarian. *Text Processing and Cognitive Technologies* N19, Proceedings of the XIIth International Conference Cognitive Modeling in Linguistics, Dubrovnik, Croatia. Ed. V. Solovyev, V. Polyakov. 144–148.

Nedelcheva S. (2012). Bulgarian Ingressive Verbs: The Case of Za– and Do–. Yearbook, Konstantin Preslavsky Publishing House. Shumen, 72–89.

Pashov P. (1966). *Bălgarskijat glagol*. Sofia: Nauka i izkustvo.

Przepiórkowski A., R. L. Górski, B. Lewandowska–Tomaszczyk, M. Łaziński. (2008) Towards the National Corpus of Polish. *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, LREC, Marrakech. ELRA.

Przybylska R. (2006). *Schematy wyobrażeniowe a semantyka polskich prefiksów czasownikowych do–, od–, prze–, roz–, u–*. Kraków: Universitas.

Šarić L., P. Mikolić (forthcoming) A semantic analysis of the verbal prefix o(b)– in Croatian. *Croatica et Slavica Iadertina*.

Šarić L. (2014). *Prostor u jeziku i metafora*. Naklada Jesenski i Turk.

Šarić L. (2008). *Spatial Concepts in Slavic. A Cognitive Linguistic Study of Prepositions and Cases*. Wiesbaden: Harrassowitz.

Šarić L., I. Tchizmarova. (2013). Space and metaphor in verbs prefixed with od–/ot– 'from' in Bosnian/Croatian/Serbian and Bulgarian. *Oslo Studies in Language* (OSLa) 5(1): 1–27.

Tabakowska E. (2003). Space and Time in Polish: The Preposition *za* and the Verbal Prefix *za*. H. Cuyckens, T. Berg, R. Dirven & K. U. Panther, eds. *Motivation in Language*. Amsterdam: Benjamins,153–177.

Tchizmarova I. (2005). *Verbal Prefixes in Bulgarian and Their Correspondences in American English: A Cognitive Linguistic Analysis*. Ball State University.

Tchizmarova I. (2006). A Cognitive Linguistic Analysis of the Bulgarian Verbal Prefix Pre–'across, through, over'. *Glossos*.

Tilkov D., S. Stoyanov, K. Popov. (1993). *Gramatika na săvremennija bălgarski knižoven ezik*. Sofia: Bulgarian Academy of Sciences.

Trubačev, O. N. ed. (2001). *Etimologičeskij slovar' slavjanskih jazykov. Praslavjanskij leksičeskij fond. Vypusk 28.*oblězti – obpovĕdan'je*. Moscow: Nauka.

Twardzisz P. (1994). The semantics of the Polish verbal prefix *ob(e)*–. A prototype approach. Gussmann, Edmund, Henryk Kardela, eds. *Focus on language: Papers from the 2nd Conference of the Polish Association for the Study of English*, Kazimiez '93. Lublin: Maria Curie–Skłodowska University Press, 217–232.

Tyler A., V. Evans. (2003). The *Semantics of English Prepositions: Spatial Scenes, Embodied Meaning and Cognition*. Cambridge: Cambridge University Press.

Corpora used

Bulgarian National Corpus. Available at: http://search.dcl.bas.bg/

Hrvatska jezična riznica (Croatian Language Repository). Available at: http://riznica.ihjj.hr/

Hrvatski nacionalni korpus (Croatian National Corpus). Available at: http://filip.ffzg.hr/bonito2/run.cgi/first_form

Dictionaries used

Anić, V. (2004). Rječnik hrvatskoga jezika. Zagreb: Novi Liber.

Bujas, Ž. (2001). Veliki hrvatsko–engleski rječnik. Zagreb: Nakladni zavod Globus.

Bălgarski etimologičen rečnik. (1995). Vol. IV. Sofia: Akademično izdatelstvo "Prof. Marin Drinov".

Gerov, N. (1977). Rečnik na bălgarskija ezik. Sofia: Bulgarski pisatel.

Rečnik na bălgarskija ezik. Available at: http://ibl.bas.bg/rbe/

Eurodict – free online dictionary. Available at: http://www.eurodict.com/

Hrvatski jezični portal. http://hjp.srce.hr/index.php?show=search.

Rečnik na bălgarskija ezik. (2002). Volume 11. Sofia: Akademično izdatelstvo "Prof. Marin Drinov".

Sažetak

U analizi se iz komparativne perspektive razmatra glagolski prefiks *o*(*b*)– u hrvatskome i bugarskome jeziku. Teorijski je okvir kognitivna lingvistika. Prva je tema na koju se osvrćemo značenjska mreža ovoga prefiksa u svjetlu polisemije. U tom sklopu posebno razmatramo sljedeća pitanja: 1) kako odvojiti značenje prefiksa od značenja osnovnih glagola, 2) kako identificirati središnje značenje i osnovna podznačenja i kako ih imenovati. Analiza se temelji na opsežnom inventaru prefigiranih glagola prikupljenom u rječnicima i korpusima. U radu s korpusima bilo je nekih izazova, pa se analiza stoga (uz spomenutu problematiku povezanu s razradom semantičke mreže) osvrće i na pitanje kako doći do kvantitativno relevantnih rezultata na temelju korpusa koji su ili ograničena opsega ili imaju druge vrste ograničenja.

Key words: verbal prefixes *o*(*b*)–, Croatian language, Bulgarian language, cognitive linguistics
Ključne riječi: glagolski prefiksi *o(b)–,* hrvatski jezik, bugarski jezik, kognitivna lingvistika